



**Evaluation von digitalen Arbeitsblättern mit Differenzierungsangebot unter Betrachtung der Motivation, der Usability und der Nutzung von Lernhilfen im Chemieunterricht**

Vom Fachbereich Chemie der Rheinland-Pfälzischen Technischen Universität Kaiserslautern-Landau zur Verleihung des akademischen Grades „Doktor der Naturwissenschaften“ genehmigte Dissertation

vorgelegt von

**Nils Fitting, M.Ed.**

Datum der wissenschaftlichen Aussprache: 20.12.2023

Betreuerin: apl. Prof. Dr. Gabriele Hornung

DE-386





Die vorliegende Arbeit wurde im Zeitraum von Oktober 2019 bis Oktober 2023 im Fachbereich Chemie der ehemaligen TU Kaiserslautern beziehungsweise jetzt RPTU in Kaiserslautern in der Arbeitsgruppe Fachdidaktik Chemie unter der Leitung von Frau apl. Prof. Dr. Gabriele Hornung angefertigt.

### **Prüfungskommission**

Vorsitz: Prof. Dr. Georg Manolikakes

1. Berichterstatter: apl. Prof. Dr. Gabriele Hornung

2. Berichterstatter: Jun.-Prof. Dr. Johann-Nikolaus Seibert



# Inhaltsverzeichnis

INHALTSVERZEICHNIS .....	I
EIDESSTATTLICHE ERKLÄRUNG.....	III
LEBENS LAUF .....	IV
DANKSAGUNG.....	V
KURZZUSAMMENFASSUNG.....	VI
ABSTRACT .....	VII
ABBILDUNGSVERZEICHNIS .....	VIII
TABELLENVERZEICHNIS.....	XIV
ABKÜRZUNGSVERZEICHNIS .....	XXIV
<b>1. EINLEITUNG .....</b>	<b>1</b>
1.1 AUSGANGSSITUATION UND ZIEL DER DISSERTATION .....	1
1.2 STUDIENÜBERBLICK.....	4
1.3 VORGEHENSWEISE.....	7
1.4 AUFBAU DER ARBEIT .....	8
<b>2. THEORIE.....</b>	<b>11</b>
2.1 MOTIVATION UND INTERESSE .....	11
2.2 USABILITY .....	22
2.3 HETEROGENITÄT UND DIFFERENZIERUNG.....	34
2.4 HYPERDOCS UND HYPERDOCSYSTEMS.....	42
<b>3. FORSCHUNGSSTAND UND FORSCHUNGSLÜCKEN .....</b>	<b>56</b>
3.1 MOTIVATIONS- UND INTERESSENFÖRDERNDES POTENTIAL DIGITALER MEDIEN UND MULTIMEDIALER LERNUMGEBUNGEN.....	56
3.2 USABILITY .....	65
3.3 EINSATZ VON LERNHILFEN UND LERNUNTERSTÜTZUNGEN .....	75
3.4 ZUSAMMENFASSENDE DARSTELLUNG DER FORSCHUNGSLÜCKEN.....	86
3.5 FORSCHUNGSFRAGEN.....	91
<b>4. FORSCHUNGSDESIGN .....</b>	<b>97</b>
4.1 METHODEN .....	97
4.2 DATENBASIS .....	107
4.3 HILFSMITTEL .....	113
4.4 PILOTIERUNG.....	131
4.5 DURCHFÜHRUNG DER HAUPTSTUDIE .....	145

<b>5.</b>	<b>ERGEBNISSE .....</b>	<b>174</b>
5.1	BESCHREIBUNG DER STICHPROBE.....	175
5.2	INTRINSISCHE MOTIVATION UND INTERESSE.....	181
5.3	USABILITY .....	223
5.4	LERNHILFEN UND VERTIEFUNGSAUFGABEN .....	248
5.5	VARIABLEN IN ZUSAMMENHANG MIT DIGITALEM LERNEN .....	277
<b>6.</b>	<b>DISKUSSION .....</b>	<b>315</b>
6.1	DISKUSSION UND EINORDNUNG DER ERGEBNISSE.....	315
6.2	FACHDIDAKTISCHES FAZIT ZUR UNTERRICHTSREIHE.....	352
6.3	METHODENKRITIK UND LIMITATIONEN .....	354
6.4	EMPFEHLUNGEN UND IMPLIKATIONEN.....	362
6.5	FORSCHUNGSBEDARF .....	368
6.6	KRITISCHE WÜRDIGUNG .....	372
<b>7.</b>	<b>FAZIT.....</b>	<b>374</b>
7.1	ZUSAMMENFASSUNG.....	374
7.2	AUSBLICK .....	378
<b>8.</b>	<b>LITERATURVERZEICHNIS .....</b>	<b>381</b>
<b>9.</b>	<b>ANHANG.....</b>	<b>420</b>
9.1	METHODEN .....	428
9.2	PILOTIERUNG .....	475
9.3	ERGEBNISSE .....	495

## Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Dissertation mit dem Titel „Evaluation von digitalen Arbeitsblättern mit Differenzierungsangebot unter Betrachtung der Motivation, der Usability und der Nutzung von Lernhilfen im Chemieunterricht“ eigenständig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben. Die Ergebnisse beteiligter Mitarbeiter:innen und anderer Autor:innen sind klar gekennzeichnet. Die Dissertation oder Teile der Dissertation als Prüfungsarbeit wurden bei keinem anderen Fachbereich eingereicht. Ein früheres Promotionsverfahren an einer anderen Hochschule wurde nicht beantragt. Die für mich geltende Promotionsordnung des Fachbereichs Chemie der ehemaligen TU Kaiserslautern, jetzt RPTU in Kaiserslautern, ist mir bekannt.

Kaiserslautern, den .....

.....

(Unterschrift)

## Lebenslauf

### Persönliche Daten

---

Name: Nils Fitting

### Schulbildung und Studium

---

08/2004 – 03/2013 Gymnasium: Leibniz-Gymnasium Pirmasens (Abschluss: Abitur)

10/2013 – 09/2016 Studium an der TU Kaiserslautern mit Abschluss:  
Bachelor of Education: Biologie und Chemie

10/2016 – 10/2018 Studium an der TU Kaiserslautern mit Abschluss:  
Master of Education – Gymnasium: Biologie und Chemie

10/2017 – 09/2019 Studium an der TU Kaiserslautern mit Abschluss:  
Erweiterungsprüfung (3.Fach): Informatik

### Beruf

---

10/2015 – 10/2016 Studentische Hilfskraft in der Fachdidaktik Chemie der TU Kaiserslautern

11/2016 – 12/2018 Wissenschaftliche Hilfskraft in der Fachdidaktik Chemie der TU Kaiserslautern

01/2019 – 07/2021 Wissenschaftlicher Mitarbeiter in der Fachdidaktik Chemie und Bioverfahrenstechnik der TU Kaiserslautern

08/2021 – 01/2023 Studienreferendar am Staatlichen Studienseminar für das Lehramt an Gymnasien in Kaiserslautern (2. Staatsexamen)

02/2023 - heute Studienrat am Helmholtz-Gymnasium Zweibrücken

## Danksagung

Für die erfolgreiche Fertigstellung dieser Arbeit möchte ich im Folgenden einigen Personen meinen besonderen Dank ausdrücken. Die Nennungen sind nicht als abgeschlossene Liste zu verstehen. Zuallererst möchte ich meiner Doktormutter, Frau apl. Prof. Dr. Gabriele Hornung, aus vollem Herzen danken, dass sie diese Arbeit überhaupt erst ermöglicht hat. Seien es die vielen Gespräche rund um die Dissertation oder auch einfach das Gefühl der vollen Unterstützung bei allen anfallenden Tätigkeiten und Problemen. Ich möchte mich außerdem bei Herrn Jun.-Prof. Dr. Johann-Nikolaus Seibert für die Übernahme des 2. Berichterstatters und bei Herrn Prof. Dr. Georg Manolikakes für die Übernahme des Prüfungsvorsitzes bedanken.

Mein Dank gilt zudem Herrn Dr. Harald Hemm für die technische Mitentwicklung von HyperDocSystems und für die vielen interessanten Gespräche auch außerhalb der Promotion und ebenso Frau Ingrid Hemm für die organisatorische Unterstützung in vielen Belangen. Ich möchte mich bei Carola bedanken, die ich seit dem Anfangsstudium kenne und die immer wieder, auch in den letzten empirischen Erhebungen, helfend zur Seite stand. Mein weiterer Dank gilt Herrn Prof. Dr. Roland Ulber, in dessen Fachbereich ich herzlich aufgenommen wurde. Die Bioverfahrenstechnik stand mir immer helfend zur Seite. Außerdem möchte ich mich bei der Deutschen Telekom Stiftung und dem Bundesministerium für Bildung und Forschung für die Finanzierung meiner Tätigkeit als wissenschaftlicher Mitarbeiter bedanken.

Ich danke zudem allen Lehrkräften, Schulleitungen und nicht zuletzt den Schüler:innen, die die empirischen Erhebungen dieser Arbeit ermöglicht haben. Mein besonderer Dank gilt in diesem Zusammenhang auch Herrn Dr. Lars Czubatinski, der mir bei didaktischen und methodischen Fragen beratend zur Seite stand und der bei der Vermittlung von Untersuchungsgruppen mit-half.

Zum Schluss möchte ich mich bei meinem engsten Personenkreis, meiner Familie, bedanken. Mein volles Dankeschön gilt meiner Mutter und meinem Vater für die gesamte Unterstützung in dieser Zeit und auf dem Weg dorthin. Ich habe mir zu keinem Zeitpunkt Sorgen um fehlende Unterstützung machen müssen und konnte mich bei Problemen immer an sie wenden. An dieser Stelle möchte ich mich ganz besonders bei meiner wundervollen Freundin Nadja bedanken, die zu jeder Zeit Verständnis aufbrachte und mich immer bei meinem Vorhaben unterstützte. Sie litt an manchen Tagen sicherlich genauso viel wie ich. Vielen Dank für alles!

## Kurzzusammenfassung

Mit der vorliegenden Dissertation wurde ein Werkzeug für die Erstellung volldigitaler binnendifferenzierter Arbeitsblätter im Regelunterricht Chemie evaluiert und weiterentwickelt, das ein motivations- und interessensförderndes Potential aufweist. Es konnten Zusammenhänge zur Benutzbarkeit der Anwendung und zum Cognitive Load hergestellt werden. Die Ergebnisse stützen damit die Erkenntnisse im Bereich des Lernens mit digitalen Medien. Die Integration von digitalen Werkzeugen in den Lernprozess ist berechtigt. Sie zeigen einerseits für Schüler:innen ein motivationsförderndes Potential und andererseits für Lehrende praktische Vorteile, indem auf vielfältige Weise Informationen dargeboten werden können – zum Beispiel im Bereich der Differenzierung. Mit HyperDocSystems können binnendifferenzierte digitale Arbeitsblätter erstellt und bearbeitet werden. Diese so genannten HyperDocs können von Lehrenden mit Lernhilfen in verschiedenen Darstellungsformen angereichert und von Lernenden voll-digital im Browser mit einem Stylus oder der Tastatur bearbeitet werden.

Im Rahmen einer quasi-experimentellen Feldstudie wurde der Einsatz dieser neuartigen HyperDocs erstmals unter Betrachtung der intrinsischen Motivation und des Interesses, der Usability sowie der Nutzung des multimedialen Differenzierungsangebots analysiert. Die Studie fand über vier Schulstunden im Regelunterricht Chemie der Mittelstufe (Gymnasium / Gesamtschule) und Oberstufe (Gymnasium) statt. Dabei wurden auch der Cognitive Load und die tabletbezogenen Kompetenzen der Lernenden berücksichtigt. Die Ergebnisse lassen auf ein motivationsförderndes Potential der HyperDocs gegenüber analogen Arbeitsblättern schließen. Dabei zeigen sich Unterschiede zwischen den Geschlechtern, die zum Teil auf den Cognitive Load zurückzuführen sind und abhängig vom Alter der Lernenden (Mittel- und Oberstufe) auftreten. Die Lernhilfen werden in diesem Zusammenhang häufig aus Interesse und Neugier verwendet. Schüler:innen nutzen insbesondere Lernhilfen in Form von Text und Bild. Die Nutzungshäufigkeit des Differenzierungsangebots gibt jedoch nicht unmittelbar Aufschlüsse über die Motivation oder den Cognitive Load der Lernenden. Bei der Usability handelt es sich um ein wichtiges Kriterium beim Einsatz von digitalen Lernprogrammen, da sich unter anderem ein Zusammenhang zu den Variablen der intrinsischen Motivation und zum Cognitive Load beim Lernen mit HyperDocs herstellen lässt. Die Usability ist dabei jedoch abhängig vom Messzeitpunkt. HyperDocs weisen eine hohe Usability auf und können daher uneingeschränkt in der Mittel- und Oberstufe eingesetzt werden.



## Abstract

In this dissertation, HyperDocSystems, a tool for creating digital differentiated worksheets with a motivation and interest-promoting potential, was evaluated and developed in regular chemistry school lessons. Correlations could be found between the aforementioned variables, the usability and the cognitive load. The results support findings in the field of learning with digital media. The integration of digital tools in the learning process is thus justifiable. On the one hand, these tools show motivation-promoting potential for students, and on the other hand, the tools show practical benefits for teachers by presenting information in many ways – for example in the field of internal differentiation. HyperDocSystems can be used to create and edit differentiated digital worksheets. Teachers can equip these so-called HyperDocs with learning aids in different representations which can be processed by learners with a keyboard or stylus.

The use of these new HyperDocs was tested under consideration of intrinsic motivation, interest, usability and the use of multimedia learning aids within an experimental field study. The study was conducted over the course of four chemistry school lessons in middle school (“Gymnasium” / “Gesamtschule”) and high school (“Gymnasium”). Cognitive load and tablet-related competencies were considered. The findings suggest a motivation promotion potential of HyperDocs towards analog worksheets. Thereby differences between genders in relation to cognitive load and age of students (middle and high school) were found. In this context, learning aids were often used because of interest and curiosity. In particular, students use learning aids in the form of text and pictures. The frequency of use gives no clear information on the motivation or the cognitive load of learners. Usability is an important aspect of the usage of digital learning programs since there are correlations between usability, variables of intrinsic motivation and cognitive load while learning with HyperDocs. However, Usability is dependent on the time of measurement. HyperDocs show a high usability and middle and high schools may use them without any restrictions.

## Abbildungsverzeichnis

<i>Abbildung 1: Vereinigtes Modell von Heckhausen &amp; Heckhausen</i> .....	12
<i>Abbildung 2: Das Modell der Selbstbestimmungstheorie im Überblick</i> .....	17
<i>Abbildung 3: Interessensmodell nach Krapp</i> .....	20
<i>Abbildung 4: Ganzheitliche Betrachtung der Usability</i> .....	23
<i>Abbildung 5: Modell der Systemakzeptanz, der die Usability untergeordnet ist</i> .....	24
<i>Abbildung 6: Usabilityinspektions- und Testmethoden</i> .....	32
<i>Abbildung 7: Übersicht über Strategien im Umgang mit Heterogenität</i> .....	38
<i>Abbildung 8: Ausschnitt eines HyperDoc mit den meisten Funktionen im Überblick</i> .....	44
<i>Abbildung 9: Kennzeichnung der Informationskärtchen auf einem HyperDoc</i> .....	45
<i>Abbildung 10: Beide Varianten der Textfeldbeschriftung</i> .....	46
<i>Abbildung 11: Ausschnitt aus dem Editor zum Erstellen und Bearbeiten eines HyperDoc</i> .....	48
<i>Abbildung 12: Detaillierte Übersicht der Hilfenutzung in HyperDocSystems</i> .....	50
<i>Abbildung 13: Die Übersicht zeigt die am häufigsten genutzten Hilfen</i> .....	50
<i>Abbildung 14: Die Übersicht zeigt die Nutzung der Hilfen, aufgeschlüsselt nach Account</i> ....	51
<i>Abbildung 15: Kognitive Theorie des multimedialen Lernens</i> .....	52
<i>Abbildung 16: Mittelwerte mit Standardabweichung der Befragung zur aktualisierten Version von HyperDocSystems</i> .....	142
<i>Abbildung 17: Rotierte Komponentenmatrix der Hauptkomponentenmatrix des Instruments für den Cognitive Load</i> .....	144
<i>Abbildung 18: Ablauf der Studie mit Erhebungsinstrumenten in der Interventionsgruppe der Mittel- und Oberstufe</i> .....	147
<i>Abbildung 19: Vorder- und Rückseite einer Hilfe mit zwei verschiedenen Darstellungsformen</i> .....	170
<i>Abbildung 20: Mittelwerte mit Standardabweichung des Interesses (SDT) beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe</i> .....	183
<i>Abbildung 21: Mittelwerte mit Standardabweichung des Fachinteresses Chemie, aufgeteilt nach Geschlecht und Gruppe im Pre- und Posttest in der Mittelstufe</i> .....	187

---

<i>Abbildung 22: Mittelwert mit Standardabweichung des Sachinteresses Naturwissenschaften in der Mittelstufe, aufgeteilt nach Geschlecht und Gruppe im Pre- und Posttest in der Mittelstufe.....</i>	<i>189</i>
<i>Abbildung 23: Mittelwerte mit Standardabweichung der wahrgenommenen Wahlfreiheit beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe .....</i>	<i>191</i>
<i>Abbildung 24: Mittelwerte mit Standardabweichung des Kompetenzerlebens beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>193</i>
<i>Abbildung 25: Mittelwerte mit Standardabweichung des Drucks beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>195</i>
<i>Abbildung 26: Mittelwerte mit Standardabweichung des Interesses (SDT) beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe .....</i>	<i>198</i>
<i>Abbildung 27: Mittelwerte mit Standardabweichung des Fachinteresses Chemie, aufgeteilt nach Geschlecht und Gruppe im Pre- und Posttest in der Oberstufe.....</i>	<i>203</i>
<i>Abbildung 28: Mittelwerte mit Standardabweichung des Sachinteresses Naturwissenschaften, aufgeteilt nach Geschlecht und Gruppe im Pre- und Posttest in der Oberstufe .....</i>	<i>204</i>
<i>Abbildung 29: Mittelwerte mit Standardabweichung der wahrgenommenen Wahlfreiheit beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe.....</i>	<i>206</i>
<i>Abbildung 30: Mittelwerte mit Standardabweichung des Kompetenzerlebens beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe .....</i>	<i>209</i>
<i>Abbildung 31: Mittelwerte mit Standardabweichung des Drucks beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe. Hohe Werte entsprechen einem geringeren Druck.....</i>	<i>211</i>
<i>Abbildung 32: Mittelwerte mit Standardabweichung des Interesses (SDT) beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>213</i>
<i>Abbildung 33: Mittelwerte mit Standardabweichung des Fachinteresses Chemie, aufgeteilt nach Geschlecht und Schulform im Pre- und Posttest in der Mittelstufe.....</i>	<i>215</i>
<i>Abbildung 34: Mittelwerte mit Standardabweichung des Sachinteresses Naturwissenschaften, aufgeteilt nach Geschlecht und Schulform im Pre- und Posttest in der Mittelstufe .....</i>	<i>217</i>
<i>Abbildung 35: Mittelwerte mit Standardabweichung der wahrgenommenen Wahlfreiheit beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>219</i>

<i>Abbildung 36: Mittelwerte mit Standardabweichung des Kompetenzerlebens beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>221</i>
<i>Abbildung 37: Mittelwerte mit Standardabweichung des Drucks beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe .....</i>	<i>223</i>
<i>Abbildung 38: Mittelwerte mit Standardabweichung der Nützlichkeit (Usability) der HyperDocs, aufgeteilt nach Geschlecht und Schulform nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Mittelstufe.....</i>	<i>225</i>
<i>Abbildung 39: Mittelwerte mit Standardabweichung der Zufriedenheit (Usability) mit HyperDocs, aufgeteilt nach Geschlecht und Schulform nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Mittelstufe.....</i>	<i>227</i>
<i>Abbildung 40: Mittelwerte mit Standardabweichung der Einfachheit der Nutzung (Usability) der HyperDocs, aufgeteilt nach Geschlecht und Schulform nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Mittelstufe .....</i>	<i>229</i>
<i>Abbildung 41: Mittelwerte mit Standardabweichung der Einfachheit des Lernens der Nutzung (Usability) der HyperDocs, aufgeteilt nach Geschlecht und Schulform nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Mittelstufe.....</i>	<i>231</i>
<i>Abbildung 42: Mittelwerte mit Standardabweichung der Nützlichkeit (Usability) der HyperDocs, aufgeteilt nach Geschlecht nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Oberstufe .....</i>	<i>233</i>
<i>Abbildung 43: Mittelwerte mit Standardabweichung der Zufriedenheit (Usability) mit HyperDocs, aufgeteilt nach Geschlecht nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Oberstufe .....</i>	<i>234</i>
<i>Abbildung 44: Mittelwerte mit Standardabweichung der Einfachheit der Nutzung (Usability) der HyperDocs, aufgeteilt nach Geschlecht nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Oberstufe .....</i>	<i>236</i>
<i>Abbildung 45: Mittelwerte mit Standardabweichung der Einfachheit des Lernens der Nutzung (Usability) der HyperDocs, aufgeteilt nach Geschlecht nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Oberstufe .....</i>	<i>237</i>
<i>Abbildung 46: Mittelwerte mit Standardabweichung der Nützlichkeit der Lernhilfen, aufgeteilt nach Gruppe in den einzelnen Stunden in der Mittelstufe.....</i>	<i>249</i>
<i>Abbildung 47: Mittelwerte mit Standardabweichung der Nützlichkeit der Lernhilfen, aufgeteilt nach Gruppe in den einzelnen Stunden in der Oberstufe .....</i>	<i>249</i>
<i>Abbildung 48: Mittelwerte mit Standardabweichung der präferierten Angebotsform (digital) der Lernhilfen in der Mittel- und Oberstufe im Pre- und Posttest (MZP 1 und MZP 2).....</i>	<i>250</i>
<i>Abbildung 49: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Lernhilfen in der Interventionsgruppe der Mittelstufe nach Messzeitpunkt (Stunde).....</i>	<i>251</i>

---

<i>Abbildung 50: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Vertiefungsaufgaben in der Interventionsgruppe der Mittelstufe nach Messzeitpunkt (Stunde) .....</i>	<i>252</i>
<i>Abbildung 51: Mittelwerte mit Standardabweichung der Beurteilung der zeitlichen Komponente bei der Bearbeitung der HyperDocs in der Mittelstufe (Interventionsgruppe) nach Messzeitpunkt (Stunde) .....</i>	<i>252</i>
<i>Abbildung 52: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Lernhilfen in der Vergleichsgruppe der Mittelstufe nach Messzeitpunkt (Stunde). .....</i>	<i>253</i>
<i>Abbildung 53: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Vertiefungsaufgaben in der Vergleichsgruppe der Mittelstufe nach Messzeitpunkt (Stunde) .....</i>	<i>254</i>
<i>Abbildung 54: Mittelwert mit Standardabweichung der Beurteilung der zeitlichen Komponente bei der Bearbeitung der Arbeitsblätter in der Mittelstufe (Vergleichsgruppe) nach Messzeitpunkt (Stunde) .....</i>	<i>254</i>
<i>Abbildung 55: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Lernhilfen in der Interventionsgruppe in der Oberstufe nach Messzeitpunkt (Stunde). .....</i>	<i>255</i>
<i>Abbildung 56: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Vertiefungsaufgaben in der Interventionsgruppe der Oberstufe nach Messzeitpunkt (Stunde). .....</i>	<i>256</i>
<i>Abbildung 57: Mittelwert mit Standardabweichung der Beurteilung der zeitlichen Komponente bei der Bearbeitung der HyperDocs in der Oberstufe (Interventionsgruppe) nach Messzeitpunkt (Stunde). .....</i>	<i>256</i>
<i>Abbildung 58: Mittelwerte mit Standardabweichung der Gründe für die Nutzung der Lernhilfen in der Vergleichsgruppe in der Oberstufe nach Messzeitpunkt (Stunde). .....</i>	<i>257</i>
<i>Abbildung 59: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Vertiefungsaufgaben in der Vergleichsgruppe der Oberstufe nach Messzeitpunkt (Stunde). .....</i>	<i>258</i>
<i>Abbildung 60: Mittelwert mit Standardabweichung der Beurteilung der zeitlichen Komponente bei der Bearbeitung der Arbeitsblätter in der Oberstufe (Vergleichsgruppe) nach Messzeitpunkt (Stunde) .....</i>	<i>258</i>
<i>Abbildung 61: Prozentuale Verteilung der Hilfenutzung nach Typ der Hilfe (gelesen) in der Mittelstufe .....</i>	<i>260</i>
<i>Abbildung 62: Prozentuale Verteilung der Hilfenutzung nach Typ der Hilfe (Rohdaten) in der Mittelstufe .....</i>	<i>260</i>
<i>Abbildung 63: Prozentuale Verteilung der Hilfenutzung nach Typ der Hilfe (gelesen) in der Oberstufe .....</i>	<i>261</i>

<i>Abbildung 64: Prozentuale Verteilung der Hilfenutzung nach Typ der Hilfe (Rohdaten) in der Oberstufe .....</i>	<i>262</i>
<i>Abbildung 65: Genormte Mittelwerte für die Hilfenutzung, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Interventionsgruppe in der Mittelstufe .....</i>	<i>265</i>
<i>Abbildung 66: Genormte Mittelwerte für die Hilfenutzung, aufgeteilt nach Geschlecht im zeitlichen Verlauf in der Vergleichsgruppe in der Mittelstufe.....</i>	<i>267</i>
<i>Abbildung 67: Genormte Mittelwerte für die Vertiefungsnutzung, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Interventionsgruppe in der Mittelstufe .....</i>	<i>269</i>
<i>Abbildung 68: Genormte Mittelwerte für die Hilfenutzung, aufgeteilt nach Geschlecht im zeitlichen Verlauf in der Interventionsgruppe in der Oberstufe.....</i>	<i>271</i>
<i>Abbildung 69: Genormte Mittelwerte für die Hilfenutzung, aufgeteilt nach Geschlecht im zeitlichen Verlauf in der Vergleichsgruppe in der Oberstufe.....</i>	<i>273</i>
<i>Abbildung 70: Genormte Mittelwerte für die Vertiefungsnutzung, aufgeteilt nach Geschlecht im zeitlichen Verlauf in der Interventionsgruppe in der Oberstufe.....</i>	<i>274</i>
<i>Abbildung 71: Mittelwerte mit Standardabweichung des Intrinsic Load beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>279</i>
<i>Abbildung 72: Mittelwerte mit Standardabweichung des Intrinsic Load beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe .....</i>	<i>282</i>
<i>Abbildung 73: Mittelwerte mit Standardabweichung des Extraneous Load beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>284</i>
<i>Abbildung 74: Mittelwerte mit Standardabweichung des Extraneous Load beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe .....</i>	<i>287</i>
<i>Abbildung 75: Mittelwerte mit Standardabweichung des Germane Load beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>289</i>
<i>Abbildung 76: Mittelwerte mit Standardabweichung des Germane Load beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe .....</i>	<i>291</i>
<i>Abbildung 77: Mittelwerte mit Standardabweichung des Mental Effort beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>293</i>

---

<i>Abbildung 78: Mittelwerte mit Standardabweichung des Mental Effort beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe .....</i>	<i>295</i>
<i>Abbildung 79: Mittelwerte mit Standardabweichung des Intrinsic Load beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>298</i>
<i>Abbildung 80: Mittelwerte mit Standardabweichung des Extraneous Load beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>301</i>
<i>Abbildung 81: Mittelwerte mit Standardabweichung des Germane Load beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>304</i>
<i>Abbildung 82: Mittelwerte mit Standardabweichung des Mental Effort beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.....</i>	<i>306</i>
<i>Abbildung 83: Mittelwerte mit Standardabweichung der Variablen TAB1 und TAB2 über die vier Messzeitpunkte (Stunden) in der Mittelstufe.....</i>	<i>310</i>
<i>Abbildung 84: Mittelwert mit Standardabweichung der Variablen TAB1 und TAB2 über die vier Messzeitpunkte (Stunden) in der Oberstufe.....</i>	<i>310</i>
<i>Abbildung 85: Mittelwerte mit Standardabweichung der Variablen TAB3 nach der ersten Stunde (MZP 1) und nach der vierten Stunde (MZP 2) in der Mittel- und Oberstufe .....</i>	<i>311</i>
<i>Abbildung 86: Mittelwerte mit Standardabweichung der tabletbezogenen Selbstwirksamkeit, aufgeteilt nach Schulform und Geschlecht im Pre- und Posttest in der Mittelstufe .....</i>	<i>312</i>
<i>Abbildung 87: Mittelwerte mit Standardabweichung der Variable „Tablets beim Lernen“, aufgeteilt nach Schulform und Geschlecht im Pre- und Posttest in der Mittelstufe .....</i>	<i>313</i>
<i>Abbildung 88: Mittelwerte mit Standardabweichung der tabletbezogenen Selbstwirksamkeit, aufgeteilt nach Geschlecht im Pre- und Posttest in der Oberstufe.....</i>	<i>314</i>
<i>Abbildung 89: Mittelwerte mit Standardabweichung der Variable „Tablets beim Lernen“, aufgeteilt nach Geschlecht im Pre- und Posttest in der Oberstufe.....</i>	<i>314</i>
<i>Abbildung 90: Übersicht über die Ergebnisse und Trends der Studie in der Mittelstufe.....</i>	<i>315</i>
<i>Abbildung 91: Übersicht über die Ergebnisse und Trends der Studie in der Oberstufe.....</i>	<i>316</i>

## Tabellenverzeichnis

<i>Tabelle 1: Verwendete R-Pakete in der Datenauswertung.....</i>	<i>100</i>
<i>Tabelle 2: Anzahl der fehlenden Werte pro Fragebogen in Prozent für die Mittel- und Oberstufe der Interventions- und Vergleichsgruppe.....</i>	<i>106</i>
<i>Tabelle 3: Anzahl der Klassen und Kurse, aufgeteilt nach Jahrgangsstufen und Schulform in der Interventionsgruppe.....</i>	<i>108</i>
<i>Tabelle 4: Anzahl der Klassen und Kurse, aufgeteilt nach Jahrgangsstufen und Schulform in der Vergleichsgruppe.....</i>	<i>109</i>
<i>Tabelle 5: Reliabilitäten der Skalen des Cognitive Load (n = 629).....</i>	<i>116</i>
<i>Tabelle 6: Kennzahlen der verschiedenen Faktorenmodelle (n = 629).....</i>	<i>117</i>
<i>Tabelle 7: Reliabilitäten der Skalen der Tabletbezogenen Selbstwirksamkeit mit dem angepassten Erhebungsinstrument (n = 377).....</i>	<i>119</i>
<i>Tabelle 8: Kennzahlen der verschiedenen Faktorenmodelle (n = 377).....</i>	<i>120</i>
<i>Tabelle 9: Reliabilitäten der Skalen der Usability mit dem angepassten Erhebungsinstrument (n = 407).....</i>	<i>121</i>
<i>Tabelle 10: Kennzahlen der verschiedenen Faktorenmodelle (n = 407).....</i>	<i>122</i>
<i>Tabelle 11: Reliabilitäten der Subskalen der intrinsischen Motivation mit dem angepassten Erhebungsinstrument (n = 634).....</i>	<i>125</i>
<i>Tabelle 12: Verwendete Items zur Erfassung der intrinsischen Motivation über das Intrinsic Motivation Inventory.....</i>	<i>125</i>
<i>Tabelle 13: Kennzahlen der verschiedenen Faktorenmodelle (n = 634).....</i>	<i>126</i>
<i>Tabelle 14: Reliabilitäten der Skala Fachinteresse Chemie und Sachinteresse Naturwissenschaften (n = 588).....</i>	<i>126</i>
<i>Tabelle 15: Stundeninhalte der ersten Pilotierung in einer Gesamtschule.....</i>	<i>132</i>
<i>Tabelle 16: Reliabilitäten der einzelnen Skalen in der ersten Pilotierung.....</i>	<i>134</i>
<i>Tabelle 17: Stundeninhalte der zweiten Pilotierung im Gymnasium (Mittelstufe).....</i>	<i>136</i>
<i>Tabelle 18: Stundeninhalte der zweiten Pilotierung in der Oberstufe.....</i>	<i>137</i>
<i>Tabelle 19: Reliabilitäten der einzelnen Skalen in der zweiten Pilotierung.....</i>	<i>138</i>
<i>Tabelle 20: Stundeninhalte der dritten Pilotierung an der Gesamtschule (Mittelstufe).....</i>	<i>141</i>
<i>Tabelle 21: Stundeninhalte der dritten Pilotierung an dem Gymnasium (Oberstufe).....</i>	<i>141</i>
<i>Tabelle 22: Reliabilitäten der einzelnen Subskalen in der dritten Pilotierung.....</i>	<i>143</i>



---

<i>Tabelle 23: Geschlecht und Alter (in Jahren) der Proband:innen in der Mittelstufe der Interventionsgruppe, aufgeteilt nach Schulform.....</i>	<i>175</i>
<i>Tabelle 24: Mittelwerte mit Standardabweichung (in Klammern) Einschätzungen der Fähigkeiten im Fach Chemie, den Naturwissenschaften (Nawi) und in der Schule allgemein in der Mittelstufe der Interventionsgruppe.....</i>	<i>175</i>
<i>Tabelle 25: Mittelwerte mit Standardabweichung (in Klammern) der häuslichen Nutzung digitaler Endgeräte in der Mittelstufe.....</i>	<i>176</i>
<i>Tabelle 26: Mittelwerte mit Standardabweichung (in Klammern) der schulischen Nutzung digitaler Endgeräte in der Mittelstufe.....</i>	<i>177</i>
<i>Tabelle 27: Geschlecht und Alter (in Jahren) der Proband:innen in der Oberstufe der Interventionsgruppe, unterteilt nach Schulform. ....</i>	<i>177</i>
<i>Tabelle 28: Mittelwerte mit Standardabweichung (in Klammern) der Einschätzungen der Fähigkeiten im Fach Chemie, den Naturwissenschaften (Nawi) und in der Schule allgemein in der Oberstufe der Interventionsgruppe.....</i>	<i>178</i>
<i>Tabelle 29: Mittelwerte mit Standardabweichung (in Klammern) der häuslichen Nutzung digitaler Endgeräte in der Oberstufe.....</i>	<i>178</i>
<i>Tabelle 30: Mittelwerte mit Standardabweichung (in Klammern) der schulischen Nutzung digitaler Endgeräte in der Oberstufe.....</i>	<i>179</i>
<i>Tabelle 31: Geschlecht und Alter (in Jahren) der Proband:innen in der Mittelstufe der Vergleichsgruppe.....</i>	<i>179</i>
<i>Tabelle 32: Mittelwerte mit Standardabweichung (in Klammern) der Einschätzungen der Fähigkeiten in Chemie, den Naturwissenschaften (Nawi) und in der Schule allgemein in der Mittelstufe der Vergleichsgruppe.....</i>	<i>180</i>
<i>Tabelle 33: Geschlecht und Alter (in Jahren) der Proband:innen in der Oberstufe der Vergleichsgruppe, unterteilt nach Schulform. ....</i>	<i>180</i>
<i>Tabelle 34: Mittelwerte mit Standardabweichung (in Klammern) der Einschätzungen der Fähigkeiten in Chemie, den Naturwissenschaften (Nawi) und in der Schule allgemein in der Oberstufe der Vergleichsgruppe.....</i>	<i>181</i>
<i>Tabelle 35: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf das Interesse (SDT) beim Lernen mit Arbeitsblättern in der Mittelstufe. ....</i>	<i>182</i>
<i>Tabelle 36: Ergebnisse der robusten Varianzanalyse zum Interesse (SDT) beim Lernen mit Arbeitsblättern in der Mittelstufe.....</i>	<i>183</i>
<i>Tabelle 37: Ergebnisse der Untersuchung der einfachen Haupteffekte der Interaktion „Geschlecht*Gruppe“ in der Mittelstufe.....</i>	<i>184</i>
<i>Tabelle 38: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Gruppe“ der Interaktion „Gruppe*Messzeitpunkt“ (Interesse) in der Mittelstufe.....</i>	<i>185</i>

<i>Tabelle 39: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf Fachinteresse Chemie in der Mittelstufe.</i>	186
<i>Tabelle 40: Ergebnisse der robusten Varianzanalyse zum Fachinteresse Chemie in der Mittelstufe.</i>	186
<i>Tabelle 41: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf das Sachinteresse an Naturwissenschaften in der Mittelstufe.</i>	188
<i>Tabelle 42: Ergebnisse der robusten Varianzanalyse zum Sachinteresse Naturwissenschaften in der Mittelstufe.</i>	188
<i>Tabelle 43: Ergebnisse der Untersuchung der einfachen Haupteffekte der Interaktion „Geschlecht*Gruppe“ in der Mittelstufe für das Sachinteresse.</i>	189
<i>Tabelle 44: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf die wahrgenommene Wahlfreiheit beim Lernen mit Arbeitsblättern in der Mittelstufe.</i>	190
<i>Tabelle 45: Ergebnisse der robusten Varianzanalyse zur wahrgenommenen Wahlfreiheit beim Lernen mit Arbeitsblättern in der Mittelstufe.</i>	191
<i>Tabelle 46: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf das Kompetenzerleben beim Lernen mit Arbeitsblättern in der Mittelstufe.</i>	192
<i>Tabelle 47: Ergebnisse der robusten Varianzanalyse zum Kompetenzerleben beim Lernen mit Arbeitsblättern in der Mittelstufe.</i>	193
<i>Tabelle 48: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Geschlecht“ der Interaktion „Geschlecht*Messzeitpunkt“ (Kompetenzerleben) in der Mittelstufe.</i>	194
<i>Tabelle 49: Untersuchung des Zwischensubjektfaktors „Gruppe“ der Interaktion „Gruppe*Messzeitpunkt“ (Kompetenzerleben) in der Mittelstufe.</i>	194
<i>Tabelle 50: Deskriptive Statistik der Interventions- und Vergleichsgruppe zum Druck beim Lernen mit Arbeitsblättern in den einzelnen Unterrichtsstunden, aufgeteilt nach Gruppe und Geschlecht in der Mittelstufe.</i>	196
<i>Tabelle 51: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf das Interesse (SDT) beim Lernen mit Arbeitsblättern in der Oberstufe.</i>	197
<i>Tabelle 52: Ergebnisse der robusten Varianzanalyse zum Interesse (SDT) beim Lernen mit Arbeitsblättern in der Oberstufe.</i>	198
<i>Tabelle 53: Untersuchung des Zwischensubjektfaktors „Geschlecht“ der Interaktion „Geschlecht*Messzeitpunkt“ (Interesse) in der Oberstufe.</i>	199
<i>Tabelle 54: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Gruppe“ der Interaktion „Gruppe*Messzeitpunkt“ (Interesse) in der Oberstufe.</i>	199
<i>Tabelle 55: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf das Fachinteresse Chemie in der Oberstufe.</i>	201

---

<i>Tabelle 56: Ergebnisse der robusten Varianzanalyse zum Fachinteresse Chemie in der Oberstufe.....</i>	<i>202</i>
<i>Tabelle 57: Teststatistik der jeweiligen Untersuchung der Interaktion „Geschlecht*Gruppe“ (Fachinteresse Chemie) im Pre- und Posttest in der Oberstufe. ....</i>	<i>202</i>
<i>Tabelle 58: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf das Sachinteresse an Naturwissenschaften in der Oberstufe. ....</i>	<i>203</i>
<i>Tabelle 59: Ergebnisse der robusten Varianzanalyse zum Sachinteresse Naturwissenschaften in der Oberstufe. ....</i>	<i>204</i>
<i>Tabelle 60: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf die wahrgenommene Wahlfreiheit beim Lernen mit Arbeitsblättern in der Oberstufe. ....</i>	<i>205</i>
<i>Tabelle 61: Ergebnisse der robusten Varianzanalyse zur wahrgenommenen Wahlfreiheit beim Lernen mit Arbeitsblättern in der Oberstufe.....</i>	<i>206</i>
<i>Tabelle 62: Deskriptive Statistik zum Kompetenzerleben beim Lernen mit Arbeitsblättern in der Oberstufe.....</i>	<i>207</i>
<i>Tabelle 63: Ergebnisse der robusten Varianzanalyse zum Kompetenzerleben beim Lernen mit Arbeitsblättern in der Oberstufe .....</i>	<i>208</i>
<i>Tabelle 64: Teststatistik der jeweiligen Untersuchung der Interaktion „Gruppe*Geschlecht“ (Kompetenzerleben) zu den einzelnen Messzeitpunkten. ....</i>	<i>208</i>
<i>Tabelle 65: Ergebnisse der Untersuchung der einfachen Haupteffekte der Interaktion „Geschlecht*Gruppe“ (Kompetenzerleben) in der Oberstufe.....</i>	<i>209</i>
<i>Tabelle 66: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Druck beim Lernen mit Arbeitsblättern in der Oberstufe. ....</i>	<i>210</i>
<i>Tabelle 67: Deskriptive Statistik des Interesses (SDT) beim Lernen mit HyperDocs im Schulformvergleich in der Mittelstufe.....</i>	<i>212</i>
<i>Tabelle 68: Ergebnisse der robusten Varianzanalyse des Interesses (SDT) beim Lernen mit HyperDocs im Schulformvergleich in der Mittelstufe.....</i>	<i>212</i>
<i>Tabelle 69: Deskriptive Statistik des Fachinteresses Chemie im Schulformvergleich in der Mittelstufe.....</i>	<i>214</i>
<i>Tabelle 70: Ergebnisse der robusten Varianzanalyse des Fachinteresses Chemie im Schulformvergleich in der Mittelstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.....</i>	<i>215</i>
<i>Tabelle 71: Deskriptive Statistik des Sachinteresses Naturwissenschaften im Schulformvergleich in der Mittelstufe.....</i>	<i>216</i>

<i>Tabelle 72: Ergebnisse der robusten Varianzanalyse des Sachinteresses an Naturwissenschaften im Schulformvergleich in der Mittelstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.</i>	217
<i>Tabelle 73: Deskriptive Statistik der wahrgenommenen Wahlfreiheit beim Lernen mit HyperDocs im Schulformvergleich in der Mittelstufe.</i>	218
<i>Tabelle 74: Ergebnisse der robusten Varianzanalyse zur wahrgenommenen Wahlfreiheit beim Lernen mit Arbeitsblättern in der Mittelstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.</i>	219
<i>Tabelle 75: Deskriptive Statistik des Kompetenzerlebens beim Lernen mit HyperDocs im Schulformvergleich in der Mittelstufe.</i>	220
<i>Tabelle 76: Ergebnisse der robusten Varianzanalyse des Kompetenzerlebens beim Lernen mit HyperDocs im Schulformvergleich in der Mittelstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.</i>	221
<i>Tabelle 77: Deskriptive Statistik des Drucks beim Lernen mit HyperDocs im Schulformvergleich in der Mittelstufe.</i>	222
<i>Tabelle 78: Deskriptive Statistik der Nützlichkeit (Usability) im Schulformvergleich in der Mittelstufe.</i>	224
<i>Tabelle 79: Ergebnisse der robusten Varianzanalyse der Nützlichkeit im Schulformvergleich in der Mittelstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.</i>	224
<i>Tabelle 80: Deskriptive Statistik der Zufriedenheit im Schulformvergleich in der Mittelstufe.</i>	226
<i>Tabelle 81: Ergebnisse der robusten Varianzanalyse der Zufriedenheit im Schulformvergleich in der Mittelstufe.</i>	227
<i>Tabelle 82: Deskriptive Statistik der Einfachheit der Nutzung im Schulformvergleich in der Mittelstufe.</i>	228
<i>Tabelle 83: Ergebnisse der robusten Varianzanalyse der Einfachheit der Nutzung im Schulformvergleich in der Mittelstufe.</i>	229
<i>Tabelle 84: Deskriptive Statistik der Einfachheit des Lernens der Nutzung im Schulformvergleich in der Mittelstufe.</i>	230
<i>Tabelle 85: Ergebnisse der robusten Varianzanalyse Einfachheit des Lernens der Nutzung im Schulformvergleich in der Mittelstufe.</i>	231
<i>Tabelle 86: Deskriptive Statistik der Nützlichkeit in der Oberstufe.</i>	232
<i>Tabelle 87: Teststatistik der robusten gemischten Varianzanalyse bezüglich der Nützlichkeit in der Oberstufe.</i>	232
<i>Tabelle 88: Deskriptive Statistik der Zufriedenheit in der Oberstufe.</i>	233

<i>Tabelle 89: Teststatistik der robusten gemischten Varianzanalyse bezüglich der Zufriedenheit in der Oberstufe.</i>	234
<i>Tabelle 90: Deskriptive Statistik der Einfachheit der Nutzung in der Oberstufe.</i>	235
<i>Tabelle 91: Teststatistik der robusten gemischten Varianzanalyse bezüglich der Einfachheit der Nutzung in der Oberstufe.</i>	235
<i>Tabelle 92: Deskriptive Statistik der Einfachheit des Lernens der Nutzung in der Oberstufe.</i>	236
<i>Tabelle 93: Teststatistik der robusten gemischten Varianzanalyse bezüglich der Einfachheit des Lernens der Nutzung in der Oberstufe.</i>	237
<i>Tabelle 94: Korrelationen (Spearman's rho) zwischen der Nutzungshäufigkeit von Tablets im häuslichen und schulischen Umfeld sowie den Faktoren der Usability in der ersten Stunde der Mittelstufe.</i>	238
<i>Tabelle 95: Korrelationen (Spearman's rho) zwischen der Nutzungshäufigkeit von Tablets im häuslichen und schulischen Umfeld sowie den Faktoren der Usability in der ersten Stunde der Oberstufe.</i>	238
<i>Tabelle 96: Korrelationen zwischen den Faktoren der Usability und den verschiedenen Cognitive Loadtypen sowie dem Mental Effort in der Mittelstufe in der ersten und vierten Stunde (n = 300).</i>	239
<i>Tabelle 97: Korrelationen zwischen den Faktoren der Usability und den verschiedenen Cognitive Loadtypen sowie dem Mental Effort in der Oberstufe in der ersten und vierten Stunde (n = 100).</i>	241
<i>Tabelle 98: Korrelationen zwischen den Faktoren der Usability und den Faktoren der intrinsischen Motivation (SDT) in der Mittelstufe in der ersten und vierten Stunde (n = 306).</i>	242
<i>Tabelle 99: Korrelationen zwischen den Faktoren der Usability und Faktoren der intrinsischen Motivation (SDT) in der Oberstufe in der ersten und vierten Stunde (n = 101).</i>	243
<i>Tabelle 100: Korrelationen zwischen den Faktoren der Usability und der tabletbezogenen Selbstwirksamkeit sowie der Variable „Tablets beim Lernen“ in der Mittelstufe in der ersten und vierten Stunde (n = 287).</i>	244
<i>Tabelle 101: Korrelationen zwischen den Faktoren der Usability und der tabletbezogenen Selbstwirksamkeit sowie Tablets beim Lernen in der Oberstufe in der ersten und vierten Stunde (n = 90).</i>	245
<i>Tabelle 102: Deskriptive Statistik der Usability der Eingabeform in der Mittelstufe.</i>	246
<i>Tabelle 103: Deskriptive Statistik der Usability der Eingabeform in der Oberstufe.</i>	246
<i>Tabelle 104: Mittelwerte mit Standardabweichung (in Klammern) der genormten Nutzung der Lernhilfen nach Typ, bezogen auf gelesene Hilfen in der Mittelstufe (n = 331).</i>	259

<i>Tabelle 105: Mittelwerte und Standardabweichung (in Klammern) der genormten Nutzung der Lernhilfen nach Typ, bezogen auf gelesene Hilfen in der Oberstufe (n = 110, Ausnahme Stunde 3: 109).</i>	261
<i>Tabelle 106: Mittelwerte mit Standardabweichung (in Klammern) bezüglich der Art der Hilfenutzung in der Mittelstufe (n = 151 Schüler:innen).</i>	262
<i>Tabelle 107: Mittelwerte mit Standardabweichung (in Klammern) bezüglich der Art der Hilfenutzung in der Oberstufe (n = 53 Schüler:innen).</i>	263
<i>Tabelle 108: Deskriptive Statistik der Hilfenutzung im Schulformvergleich in der Mittelstufe.</i>	264
<i>Tabelle 109: Ergebnisse der robusten Varianzanalyse der Hilfenutzung im Schulformvergleich in der Mittelstufe.</i>	265
<i>Tabelle 110: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Geschlecht“ der Interaktion „Geschlecht*Messzeitpunkt“ (Lernhilfen) in der Mittelstufe.</i>	266
<i>Tabelle 111: Deskriptive Statistik der Hilfenutzung in der Vergleichsgruppe in der Mittelstufe.</i>	267
<i>Tabelle 112: Ergebnisse der robusten Varianzanalyse der Lernhilfenutzung in der Vergleichsgruppe der Mittelstufe.</i>	267
<i>Tabelle 113: Deskriptive Statistik der Vertiefungsnutzung in der Interventionsgruppe der Mittelstufe.</i>	269
<i>Tabelle 114: Deskriptive Statistik der Hilfenutzung in der Interventionsgruppe der Oberstufe.</i>	270
<i>Tabelle 115: Ergebnisse der robusten Varianzanalyse der Hilfenutzung in der Interventionsgruppe der Oberstufe.</i>	271
<i>Tabelle 116: Deskriptive Statistik der Hilfenutzung in der Vergleichsgruppe der Oberstufe.</i>	272
<i>Tabelle 117: Ergebnisse der robusten Varianzanalyse in der Vergleichsgruppe der Oberstufe.</i>	272
<i>Tabelle 118: Deskriptive Statistik der Vertiefungsnutzung in der Oberstufe.</i>	273
<i>Tabelle 119: Korrelationen (Spearman's rho) zwischen den Faktoren der intrinsischen Motivation und der Nutzung der Lernhilfen in der Mittelstufe (n = 316) in der ersten und vierten Stunde.</i>	275
<i>Tabelle 120: Korrelationen (Spearman's rho) zwischen den Faktoren der intrinsischen Motivation und der Nutzung der Hilfen in der Oberstufe (n = 109) in der ersten und vierten Stunde.</i>	275
<i>Tabelle 121: Korrelationen (Spearman's rho) zwischen den Faktoren des Cognitive Load und der Nutzung der Hilfen in der Mittelstufe (n = 312) in der ersten und vierten Stunde.</i>	276

<i>Tabelle 122: Korrelationen (Spearman's rho) zwischen den Faktoren des Cognitive Load und der Nutzung der Hilfen in der Oberstufe (n = 109) in der ersten und vierten Stunde.....</i>	<i>276</i>
<i>Tabelle 123: Korrelationen (Spearman's rho) zwischen der Selbsteinschätzung der Schulleistungen und der Nutzung der Hilfen in der Mittelstufe (n = 320) in der ersten und vierten Stunde.....</i>	<i>277</i>
<i>Tabelle 124: Korrelationen (Spearman's rho) zwischen der Selbsteinschätzung der Schulleistungen und der Nutzung der Hilfen in der Oberstufe (n = 110) in der ersten und vierten Stunde.....</i>	<i>277</i>
<i>Tabelle 125: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Intrinsic Load in der Mittelstufe.....</i>	<i>278</i>
<i>Tabelle 126: Ergebnisse der robusten Varianzanalyse des Intrinsic Load in der Mittelstufe..</i>	<i>279</i>
<i>Tabelle 127: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Gruppe“ der Interaktion „Gruppe*Messzeitpunkt“ (Intrinsic Load) in der Mittelstufe.....</i>	<i>280</i>
<i>Tabelle 128: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Geschlecht“ der Interaktion „Geschlecht*Messzeitpunkt“ (Intrinsic Load) in der Mittelstufe..</i>	<i>280</i>
<i>Tabelle 129: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Intrinsic Load in der Oberstufe.....</i>	<i>281</i>
<i>Tabelle 130: Ergebnisse der robusten Varianzanalyse des Intrinsic Load in der Oberstufe. .</i>	<i>282</i>
<i>Tabelle 131: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Extraneous Load in der Mittelstufe.....</i>	<i>283</i>
<i>Tabelle 132: Ergebnisse der robusten Varianzanalyse des Extraneous Load in der Mittelstufe.</i>	<i>284</i>
<i>Tabelle 133: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Geschlecht“ der Interaktion „Geschlecht*Messzeitpunkt“ (Extraneous Load) in der Mittelstufe.</i>	<i>285</i>
<i>Tabelle 134: Untersuchung des einfachen Haupteffektes des Zwischensubjektfaktors bezüglich der Interaktion „Gruppe*Messzeitpunkt“ (Extraneous Load) in der Mittelstufe mittels robusten t-Tests.....</i>	<i>285</i>
<i>Tabelle 135: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Extraneous Load in der Oberstufe.....</i>	<i>286</i>
<i>Tabelle 136: Ergebnisse der robusten Varianzanalyse des Extraneous Load in der Oberstufe ..</i>	<i>287</i>
<i>Tabelle 137: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Germane Load in der Mittelstufe.....</i>	<i>288</i>

<i>Tabelle 138: Ergebnisse der robusten Varianzanalyse des Germane Load in der Mittelstufe. ..</i>	289
<i>Tabelle 139: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Germane Load in der Oberstufe. ....</i>	290
<i>Tabelle 140: Ergebnisse der robusten Varianzanalyse des Germane Load in der Oberstufe. ..</i>	291
<i>Tabelle 141: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Mental Effort in der Mittelstufe. ....</i>	292
<i>Tabelle 142: Ergebnisse der robusten Varianzanalyse des Mental Effort in der Mittelstufe. ..</i>	293
<i>Tabelle 143: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Mental Effort in der Oberstufe. ....</i>	294
<i>Tabelle 144: Ergebnisse der robusten Varianzanalyse des Mental Effort in der Oberstufe. ..</i>	295
<i>Tabelle 145: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Gruppe“ der Interaktion „Gruppe*Messzeitpunkt“ (Mental Effort) in der Oberstufe. ....</i>	296
<i>Tabelle 146: Deskriptive Statistik des Intrinsic Load im Schulformvergleich in der Mittelstufe. ....</i>	297
<i>Tabelle 147: Ergebnisse der robusten Varianzanalyse des Intrinsic Load im Schulformvergleich in der Mittelstufe. ....</i>	298
<i>Tabelle 148: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Schulform“ der Interaktion „Schulform*Messzeitpunkt“ in der Mittelstufe. ....</i>	299
<i>Tabelle 149: Deskriptive Statistik des Extraneous Load im Schulformvergleich in der Mittelstufe. ....</i>	300
<i>Tabelle 150: Ergebnisse der robusten Varianzanalyse des Extraneous Load im Schulformvergleich in der Mittelstufe. ....</i>	301
<i>Tabelle 151: Teststatistik der jeweiligen Untersuchung der Interaktion „Geschlecht*Schulform“ zu den einzelnen Messzeitpunkten in der Mittelstufe. ....</i>	302
<i>Tabelle 152: Deskriptive Statistik des Germane Load im Schulformvergleich in der Mittelstufe. ....</i>	303
<i>Tabelle 153: Ergebnisse der robusten Varianzanalyse des Germane Load im Schulformvergleich in der Mittelstufe. ....</i>	303
<i>Tabelle 154: Deskriptive Statistik des Mental Effort im Schulformvergleich in der Mittelstufe. ....</i>	305



---

<i>Tabelle 155: Ergebnisse der robusten Varianzanalyse des Mental Effort im Schulformvergleich in der Mittelstufe. ....</i>	<i>305</i>
<i>Tabelle 156: Korrelationen zwischen den Faktoren der intrinsischen Motivation und des Cognitive Load in der Mittelstufe (n = 466) in der ersten Stunde.....</i>	<i>307</i>
<i>Tabelle 157: Korrelationen zwischen den Faktoren der intrinsischen Motivation und des Cognitive Load in der Mittelstufe (n = 466) in der vierten Stunde.....</i>	<i>307</i>
<i>Tabelle 158: Korrelationen zwischen den Faktoren der intrinsischen Motivation und des Cognitive Load in der Oberstufe (n = 146) in der ersten Stunde. ....</i>	<i>308</i>
<i>Tabelle 159: Korrelationen zwischen den Faktoren der intrinsischen Motivation und des Cognitive Load in der Oberstufe (n = 146) in der vierten Stunde.....</i>	<i>309</i>
<i>Tabelle 160: Deskriptive Statistik der tabletbezogenen Selbstwirksamkeit im Schulformvergleich in der Mittelstufe.....</i>	<i>311</i>
<i>Tabelle 161: Deskriptive Statistik von Tablets beim Lernen im Schulformvergleich in der Mittelstufe.....</i>	<i>312</i>
<i>Tabelle 162: Deskriptive Statistik der tabletbezogenen Selbstwirksamkeit in der Oberstufe.</i>	<i>313</i>
<i>Tabelle 163: Deskriptive Statistik der Variable „Tablets beim Lernen“ in der Oberstufe....</i>	<i>314</i>

## Abkürzungsverzeichnis

ANOVA .....	Analysis of Variance
CL.....	Cognitive Load
CUSE.....	Computer User Self-Efficacy Scale
EM.....	Expectation-Maximization
IMI.....	Intrinsic Motivation Inventory
MZP.....	Messzeitpunkt
SD.....	Standard Deviation
SDT .....	Self-Determination-Theory
SUS.....	System Usability Scale
USE .....	Usefulness, Satisfaction and Ease of Use
UX.....	User Experience

# 1. Einleitung

Die Einleitung dieser Arbeit umfasst neben der Ausgangssituation, in der unter anderem die Beweggründe zur Verfassung dieser Dissertation offengelegt werden, weitere Kapitel, die dem Überblick dienen. Dazu zählt beispielsweise der Studienüberblick, der in die Schwerpunkte und Forschungsgegenstände dieser Arbeit aus praktischer Sicht einführt. Aus diesem Überblick leiten sich die Vorgehensweise der Promotion und die damit verbundene Studie ab. Daher wird dort eine grobe Übersicht über das Studiendesign und die Methodik gegeben. Ein detaillierter Aufbau dieser Dissertation wird im letzten Teilkapitel vorgestellt. Damit soll eine bessere Übersicht dieser umfassenden Arbeit ermöglicht werden.

## 1.1 Ausgangssituation und Ziel der Dissertation

Die Digitalisierung ermöglicht neue Wege des Lehrens und Lernens in verschiedenen Situationen und Lebensabschnitten. Das zeigte sich insbesondere während der Corona-Pandemie, in der digitale Medien eine große Rolle spielten. Lehrkräfte unterrichteten während der Corona-Pandemie Schüler:innen oftmals dezentral über Videokonferenzen und Arbeitsaufträge. Durch die Unterstützung digitaler Tools konnte eine abwechslungsreiche Interaktion mit den Lernenden ermöglicht werden. Gleichzeitig eröffneten sich durch den Fernunterricht jedoch bislang unbekannte Probleme, insbesondere dadurch, dass Lehrkräfte den Lernenden nicht mehr unmittelbar vor Ort Unterstützung geben konnten. Gleichzeitig wurde nicht der gesamte Unterricht über Videokonferenzen abgedeckt, sondern Schüler:innen mussten eigenständig Arbeitsaufträge beziehungsweise Wochenpläne erfüllen (Wößmann et al., 2020). In diesen Situationen fehlte die Lehrkraft als Ansprechperson gänzlich. Gerade schwächere Lernende könnten damit schnell überfordert sein.

Diese Ausführung zeigt, dass eine Auflösung der analogen zentralen Unterrichtsform erhebliche Probleme mit sich bringt. Aus diesem Grund müssen Alternativen geschaffen und analoge Methoden des Regelunterrichts neu gedacht werden, um diese Methoden flexibel nutzen zu können. Ein weiterer Aspekt, der aus der Beschreibung hervorgeht, ist eine erwartbare Zunahme der Heterogenität im Bildungssystem durch den ausgefallenen Unterricht (Helbig et al., 2022). Die Expert:innenkommission der Friedrich-Ebert-Stiftung fordert daher unter anderem einen Fortschritt in der Digitalisierung und eine Stärkung von Diagnose und Förderung (Maaz et al., 2021).

Heterogene Lerngruppen sind in allen Fächern und Schulformen anzutreffen. Häufig variiert eine Lerngruppe natürlicherweise durch die jeweiligen Geschlechter oder auch durch die Interessen der Lernenden. Es gibt Jugendliche, die eine im Vergleich zum Durchschnitt erhöhte oder verminderte Intelligenz aufweisen (Vock & Gronostaj, 2017, S. 37 - 40). Zusätzlich können verschiedene Lernbeeinträchtigungen vorliegen (Vock & Gronostaj, 2017, S. 32 - 34). Damit sieht sich die Lehrkraft mit einer Vielzahl verschiedener Lernvoraussetzungen innerhalb einer Lerngruppe konfrontiert. Dieser Sachverhalt erfordert verschiedene Maßnahmen, um der Heterogenität zu begegnen beziehungsweise für alle Schüler:innen ein gleichermaßen reichhaltiges Lernangebot bereitzustellen. Ein Beispiel dafür ist die Binnendifferenzierung durch verschieden aufbereitete Lernmaterialien. Konkret können Lehrkräfte Lernhilfen für leistungsschwächere Lernende und Vertiefungsaufgaben für leistungsstärkere und schnellere Schüler:innen bereitstellen (Letzel & Otto, 2019).

Die zur Binnendifferenzierung notwendigen Lernmaterialien können neben dem analogen Format mit *HyperDocs* zum Beispiel auch digital gestaltet und den Lernenden zur Verfügung gestellt werden. Es handelt bei HyperDocs um digitale Arbeitsblätter, die unter anderem Hilfen und Vertiefungsaufgaben beinhalten. Während die digitalen Arbeitsblätter anfänglich mit PDF-Dokumenten und Hilfen sowie Vertiefungsaufgaben über den Browser umgesetzt wurden, konnte seit Ende 2019 eine vollständig browserbasierte Version entwickelt werden. Dabei können Lehrkräfte HyperDocs im Browser erstellen und Schüler:innen diese bearbeiten. Im Vergleich zu analogen Lernmaterialien ergibt sich eine Reihe von Vorteilen. So können vielfältige multimediale Inhalte eingebunden werden. Die Darstellung der Inhalte beschränkt sich dabei nicht nur auf geschriebene Texte und Abbildungen, sondern es können auch gesprochene Texte (Audio) und bewegte Bildelemente (Videos) eingebunden werden. Gleichzeitig entfällt eine Bereitstellung der Materialien (Arbeitsblätter und Hilfen) in Klassengröße. Die digitale Umsetzung dieser Methode ermöglicht weiterhin die Überwachung des Nutzungsverhaltens der Hilfen und Vertiefungsaufgaben. Es kann damit festgestellt werden, welches Format wie oft und wie lange geöffnet wurde. Einige Autor:innen empfehlen in diesem Zusammenhang den Einsatz von Software, um Differenzierungsmaßnahmen besser bewerten zu können (Deunk et al., 2018).

Einerseits sollte in der vorliegenden Studie mit HyperDocs ein digitales Lernwerkzeug evaluiert werden, das die Binnendifferenzierung über digitale Lernhilfen ermöglicht.

Andererseits bestehen in der Literatur verschiedene offene Forschungsfragen, die mithilfe von HyperDocs beantwortet werden sollten. Ausgehend von der eben beschriebenen Darstellung wurden für diese wissenschaftliche Arbeit drei wesentliche Forschungsschwerpunkte gewählt: Die intrinsische Motivation und damit verbunden das Interesse, die Benutzbarkeit (*engl.: Usability*) und der digitale Einsatz von Lernhilfen.

Die intrinsische Motivation und das Interesse bilden den ersten Forschungsschwerpunkt. Der Einfluss von digitalen Unterrichtswerkzeugen auf die Motivation und das Interesse ist bereits belegt (Higgins et al., 2019). Das Potential der digitalen Medien ist dabei jedoch von deren Beschaffenheit abhängig (Hillmayr et al., 2020), weshalb das motivationsfördernde Potential für HyperDocs geprüft wird. In diesem Zusammenhang spielt beispielsweise auch die Usability eine große Rolle (Karapanos et al., 2018).

Den zweiten Schwerpunkt bildet daher die Usability, auch Benutzbarkeit genannt. Sie ist wichtig, um ein zuverlässiges und zufriedenstellendes Benutzer:innenerlebnis zu ermöglichen. Die Benutzbarkeit sollte vor und während des Einsatzes einer Software evaluiert werden (Van Nuland et al., 2017). Zusätzlich liefert die vorliegende Studie Antworten auf Forschungsfragen im Bereich der Usabilityforschung, die insbesondere für den Schulkontext von hoher Relevanz sind, da ein Einfluss der Usability auf die Lernwirksamkeit bekannt ist (Meiselwitz & Sadera, 2008).

Den dritten Schwerpunkt bildet die Erforschung des Einsatzes von digitalen Lernhilfen durch HyperDocSystems. In diesem Bereich existieren bisher kaum Studien. Im Gegensatz dazu besteht für den analogen Einsatz von Lernhilfen bereits eine Vielzahl an Studien, die jedoch das genaue Nutzungsverhalten der Hilfen nur begrenzt nachvollziehen konnten (z. B. Arnold et al., 2016; Kleinert et al., 2021). Aus diesem Grund werden verschiedene Forschungsfragen für den digitalen Einsatz von Lernhilfen aufgeworfen und explorativ untersucht.

In Zusammenhang mit den eben aufgeführten Forschungsschwerpunkten der intrinsischen Motivation sowie dem Interesse, der Usability und den Lernhilfen stehen weitere Variablen wie die tabletbezogenen Kompetenzen, der Cognitive Load und die Nutzungshäufigkeit von digitalen Medien (vgl. Kapitel 4.2.2). Sie wurden daher ebenfalls zur Beantwortung der Forschungsfragen (vgl. Kapitel 3.5) herangezogen. Die Forschungsschwerpunkte wurden unter Berücksichtigung der Mittel- und Oberstufe sowie des Gymnasiums und der Gesamtschule in einer quasi-experimentellen Feldstudie untersucht. Dazu wurden die HyperDocs in einer vierstündigen Unterrichtsreihe zum Thema

„Farbstoffe in Lebensmitteln“ eingesetzt (vgl. Kapitel 4.5.3). Zu den oben definierten Forschungsschwerpunkten existieren bereits verschiedene Studien. Einige davon werden nun im folgenden Teilkapitel dargestellt.

## **1.2 Studienüberblick**

Mit HyperDocSystems wurde eine Webanwendung evaluiert, die im Unterricht in verschiedenen Kontexten eingesetzt und damit als Lernmanagementsystem oder E-Learningssystem angesehen werden kann. Im gleichen Zuge ermöglicht das digitale Werkzeug die Binnendifferenzierung über Lernhilfen. Das gesamte Lernmaterial kann multimedial, durch verschiedene Darstellungsformen, angereichert werden. Hieraus ergeben sich die bereits aufgeführten Forschungsschwerpunkte. Im Folgenden wird daher ein grober Studienüberblick über die genannten Themen gegeben, der im Rahmen des Forschungsstandes in Kapitel 3 vertieft wird.

### **1.2.1 Motivation, Interesse und Lernwirksamkeit digitaler Medien**

Der Einfluss digitaler Medien auf die Motivation von Lernenden ist belegt (Hillmayr et al., 2020). Dabei ist die Beschaffenheit des verwendeten digitalen Werkzeugs von hoher Relevanz (Hillmayr et al., 2020). Zudem ist nach Higgins et al. (2019) die Motivationsförderung von der Einsatzdauer abhängig. Mit zunehmender Dauer sinkt das Niveau der Motivation (Higgins et al., 2019). Diese Beobachtung lässt sich vermutlich auf den Neuhheitseffekt zurückführen (Hillmayr et al., 2020). Ein Beispiel für digitale Medien, die einen positiven Einfluss auf die Motivation haben, sind Tutoringsysteme (Hillmayr et al., 2020). In ihrer Metaanalyse mit 92 Studien untersuchten Hillmayr und Kolleg:innen den Effekt von digitalen Medien auf das Lernen. Insbesondere Tools, die die Autonomie, das Kompetenzerleben oder die soziale Eingebundenheit ermöglichen beziehungsweise fördern, entfalten ein motivationsförderndes Potential (Hillmayr et al., 2020; Jenö et al., 2019b). Diese Beobachtung steht in Einklang mit der Selbstbestimmungstheorie (engl. *Self-Determination Theory*) von Deci und Ryan (1985). Im Bereich des Mathematikunterrichts fanden Higgins und Kolleg:innen (2019) in ihrer Metaanalyse mit 56 Publikationen schwache bis mittelstarke Effekte auf die Motivation beim Lernen mit digitalen Werkzeugen. Die Effektstärken unterscheiden sich zum Teil je nach Forschungssetting. Andere Autor:innen fanden ebenfalls positive Auswirkungen auf die Motivation von Lernenden beim Einsatz von digitalen Werkzeugen (Bruckermann et al., 2016). Ran und

Kolleg:innen fanden in ihren Metaanalysen (2021, 2022) positive Effekte von digitalen Medien auf die Lernwirksamkeit im Mathematikunterricht. Hierbei liegen, je nach Einsatz der Technologie, unterschiedliche Effektstärken vor. Zudem profitieren insbesondere schwächere Schüler:innen vom Einsatz digitaler Medien im Unterricht. Die Effekte könnten ebenfalls unter Einbezug der Self-Determination-Theory erklärt werden, indem zum Beispiel die Autonomie oder das Kompetenzerleben gefördert werden. Insgesamt besteht nach einer Metaanalyse von Lazowski und Hulleman (2016) ein mittlerer Einfluss der Motivation auf die Lernwirksamkeit, womit es sich um eine wichtige Komponente der schulischen Wissensvermittlung handelt. Das gezeigte Interesse am Fach Chemie, das in Verbindung mit der intrinsischen Motivation steht, unterscheidet sich zudem zwischen den Geschlechtern. Demnach finden Schülerinnen chemische Inhalte weniger interessant als Schüler (z. B. Jansen et al., 2013). Zusätzlich führt der Einsatz von digitalen Medien bei Schülerinnen nicht zwangsläufig zu einem höheren Lernzuwachs (z. B. Bruckermann et al., 2016). Für alle Schüler:innen, die mit digitalen Werkzeugen lernen, spielt allerdings die Benutzbarkeit einer Anwendung eine große Rolle, da unter anderem ein Zusammenhang zum Interesse als ein wichtiger Bestandteil der intrinsischen Motivation nachgewiesen ist (Karapanos et al., 2018).

### **1.2.2 Usability**

Die Usability wird häufig als ein notwendiges Kriterium für den erfolgreichen Einsatz einer Anwendung gesehen. Unter der Benutzbarkeit einer Anwendung wird im allgemeinen Sprachgebrauch, und früher unter Softwareentwickler:innen, die Benutzerfreundlichkeit verstanden (Nielsen, 1993, S. 23). Dabei können sowohl Gegenstände (z. B. Wallace & Yu, 2009) als auch Softwareanwendungen (z. B. Hendra et al., 2018) eine bestimmte Gebrauchstauglichkeit aufweisen. Usabilitystudien sind jedoch eher bei Letzteren anzutreffen. In Bezug auf das Lernen mit verschiedenen Softwaresystemen zeigen sich in Studien wichtige Zusammenhänge. Karapanos und Kolleg:innen (2018) konnten in einer quasi Feldstudie eine positive Korrelation zwischen dem Abschneiden in einem Leistungstest und der Usability des eingesetzten Programms sowie zwischen dem Interesse und der Usability feststellen. In einer 90-minütigen Unterrichtseinheit arbeiteten 31 Schüler:innen zweier Biologieleistungskurse in einem virtuellen Labor. Die Autor:innen fanden außerdem Unterschiede zwischen den Geschlechtern. Männliche Teilnehmer bewerteten die Usability und das Interesse am virtuellen Labor höher (Karapanos et al., 2018). Tselios et al. (2001) evaluierten zwei unterschiedliche Lernmanagementsysteme mit 88

Erstsemesterstudierenden, und diejenigen Student:innen, die mit dem besser bewerteten System arbeiteten, zeigten auch einen größeren Lernzuwachs. Die Forscher:innen stellen daher die Wichtigkeit der Usability für Lernmanagementsoftwares heraus (Tselios et al. 2001). Auch Meiselwitz & Sadera (2008) konnten eine positive Korrelation zwischen der erhobenen Usability und dem Lernzuwachs bei Studierenden feststellen. Sie fanden allerdings keine Unterschiede zwischen den Geschlechtern in Bezug auf die Usability. Hornbæk (2006) sichtet 180 Studien im Bereich der Mensch-Computer-Interaktion, die Usabilitymessungen beinhalteten. Er formulierte daraus verschiedene Herausforderungen und offene Fragen in Bezug auf die Beforschung der Usability. Er leitet aus seinem Review unter anderem ab, dass mehr Faktoren und Zusammenhänge zur Usability untersucht werden müssen. Weiterhin stellt er heraus, dass Usabilitymessungen stärker auf makroskopischer Ebene, das umfasst einen längeren Zeitraum der Nutzung sowie kompliziertere Programme und Aufgaben, durchgeführt werden müssten (Hornbæk, 2006).

### **1.2.3 HyperDocSystems und Binnendifferenzierung**

Bisher existieren nur wenige, meist im Rahmen dieser und einer weiteren laufenden Dissertation publizierte, Ergebnisse zu HyperDocSystems (Fitting et al., 2021; Fitting & Hornung, 2022; Fitting et al., 2023; Loth & Döhrmann, 2022; Loth, 2023). Der Grund hierfür liegt darin, dass HyperDocSystems zunächst in dieser Promotion weiterentwickelt und evaluiert wurde. Im Rahmen der wissenschaftlichen Tätigkeit wurde das Werkzeug allerdings öffentlich vorgestellt, sodass gegebenenfalls weitere Publikationen folgen, die allerdings in der vorliegenden Arbeit noch keine Berücksichtigung finden. Auch zum früheren Einsatz der ersten Version von HyperDocSystems existieren lediglich die Abschlussarbeit von Fitting (2018) sowie ein Konferenzbeitrag (Fitting et al., 2018). Dort konnte gezeigt werden, dass Schüler:innen verschiedene Darstellungsformen von Hilfen gegenüber anderen bevorzugen. Beispielsweise werden text- und bildbasierte Hilfen häufiger genutzt als Video- oder Audiohilfen. Weiterhin zeigte sich eine tendenziell abnehmende Nutzungshäufigkeit der Hilfen über die Einsatzdauer von mehreren Wochen. Die untersuchte Stichprobe war jedoch klein. Eine weitere qualitative Studie untersuchte HyperDocSystems in inklusiven Lernsettings, um die Datenkompetenz von Lernenden zu fördern und Gestaltungsmerkmale der Lernumgebung zu identifizieren, die die Teilhabe von Schüler:innen fördern oder hindern (Loth & Döhrmann, 2022; Loth, 2023). Bei der genannten Studie steht die Beforschung von HyperDocSystems selbst jedoch nicht im Vordergrund.



Zum Teil existieren jedoch ähnliche Konzepte und Softwaresysteme wie HyperDocSystems. Ein Beispiel hierfür sind Apple-iBooks, die jedoch auf den Einsatz von Applegeräten beschränkt sind (Seibert et al., 2020). Ein Beispiel für die Verknüpfung von digitalen Lernhilfen mit analogen Arbeitsblättern sind sogenannte qr-Lernhilfen (Hamers et al., 2022). Zu traditionellen analogen Lernhilfen existieren verschiedene Veröffentlichungen. Arnold und Kolleg:innen (2016) konnten die Lernwirksamkeit von analogen gestuften Lernhilfen in Bezug auf das wissenschaftliche Denken nachweisen. Zudem konnte der Cognitive Load verringert werden. Großmann und Woest (2014, 2015) setzten unstrukturierte Lernhilfen in einer Längsschnittstudie ein und stellten Unterschiede in der Nutzung der Lernhilfen bei leistungsstarken und leistungsschwachen sowie unmotivierten Schüler:innen fest. Lernende mit einem positiven Selbstbild nutzten weniger häufig eine Lernhilfe. Zudem muss die Methode über einen längeren Zeitraum eingeführt werden, bis die Lernenden die Hilfen effektiv nutzen. Eine Methode, die von Wolff und Wolff (2016) vorgestellt wurde, beinhaltet gestufte multimediale Hilfekarten. Das Konzept wurde allerdings nur in einem kleinen Rahmen erprobt. Auch wenn die bisher aufgeführten Studien einen positiven Einfluss der Binnendifferenzierung durch Lernhilfen zeigen, merken Pozas und Schneider (2019) an, dass bisher wenig Studien zum Einfluss von nonverbalen Lernhilfen auf das Lernen bestehen.

### **1.3 Vorgehensweise**

Zur Beantwortung der Forschungsfragen wurde eine quasi-experimentelle Feldstudie (Rost, 2013, S. 80 - 84) durchgeführt, um HyperDocSystems im Regelunterricht des Fachs Chemie zu evaluieren und mit der analogen Bearbeitung von Arbeitsblättern zu vergleichen. Die Feldstudie bestand aus zwei vierstündigen Unterrichtsreihen mit dem Thema „Farbstoffe in Lebensmitteln“, die für die Mittelstufe (Klasse 9-10) und Oberstufe (Stufe 11-12), ausgehend von Czubatinski et al. (2020) und Nieß et al. (2020), auf das entsprechende Setting zur Durchführung der Studie angepasst wurden (vgl. Kapitel 4.5.3). Dazu wurde im Vorfeld der Erhebung eine neue Version von HyperDocSystems, ausgehend von Fitting (2018) entwickelt. An der Entwicklung waren verschiedene Mitarbeiter:innen der Fachdidaktik Chemie der RPTU in Kaiserslautern beteiligt. Die aktualisierte Version ermöglicht unter anderem das volldigitale Bearbeiten der Arbeitsblätter direkt auf dem Tablet. Dabei wurde auf die Entwicklung einer einfachen und zielführenden Bedienung geachtet. Im Gegensatz zu früheren Versionen muss nicht mehr zwischen

verschiedenen Programmen, zum Beispiel Browser und PDF-Reader, gewechselt werden. Ausgehend von dieser Version wurden zunächst drei Pilotierungen durchgeführt, um Fehler im Programm zu finden und die gesamte Unterrichtsreihe zu optimieren (vgl. Kapitel 4.4)

Die eigentliche Erhebung zur Gewinnung der relevanten Daten fand ab Sommer 2020 (Interventionsgruppe) statt. Die Erhebung der Vergleichsgruppe erfolgte ab Frühjahr 2021. Dazu wurden bestehende Kontakte zu Lehrkräften und Schulen genutzt, um Klassen für die Durchführung der Studie zu gewinnen. Die Auswahl der Klassen erfolgte ohne spezielle Auswahlkriterien. Es nahmen damit auch Lehrkräfte teil, zu denen im Vorfeld der Erhebung noch kein Kontakt bestand. Für die Studie wurden Apple iPads mit Apple Pencils durch die Fachdidaktik Chemie zur Verfügung gestellt. Alle Erhebungen wurden vom Verfasser dieser Arbeit durchgeführt. Die Studie fand im Umfeld der Corona-Pandemie statt, was jedoch keine Auswirkungen auf die Stundenstruktur des Regelunterrichts hatte. Auswirkungen gab es lediglich durch den im Dezember 2020 beschlossenen Lockdown, wodurch es zu einer Verschiebung der Erhebung der Vergleichsgruppe kam. Bei den Pilotierungen fanden die Stunden teilweise mit geteilten Lerngruppen im wöchentlichen Wechsel statt. Während der Studie wurden Daten in Form einer Befragung der Schüler:innen und in Form von Protokolldaten aus HyperDocSystems gewonnen. Die Daten aus der Befragung wurden zunächst digitalisiert und anschließend aufbereitet. Unter anderem wurden fehlende Werte imputiert und stark unvollständige Datensätze ausgeschlossen. Die Protokolldaten wurden mittels eines Python-Scripts aufbereitet und ausgewertet. Die Nutzungsdaten der Schüler:innen der Vergleichsgruppe mussten händisch ausgewertet werden. Die eigentlichen Analysen erfolgten in *SPSS* (Version 26) und *RStudio* (verschiedene fortlaufende Versionen). Dazu wurden parametrische und nicht-parametrische Verfahren eingesetzt. Neben der Prüfung der Gütekriterien, unter anderem durch konfirmatorische Faktorenanalysen und Reliabilitätsanalysen, fanden Korrelations- und Vergleichsanalysen statt. Die gewonnenen Ergebnisse werden im Diskussionsteil mit der bestehenden Literatur verglichen und Schlussfolgerungen gezogen.

## **1.4 Aufbau der Arbeit**

Das erste Kapitel der vorliegenden Dissertation behandelte die Einführung und die Zielsetzung. Damit wurde ein kurzer Studienüberblick über die drei Schwerpunkte verbunden. Diese Schwerpunkte werden nun im zweiten Kapitel theoretisch hergeleitet und

wichtige Begriffe definiert. Das betrifft das vorliegende Motivationsverständnis nach der Selbstbestimmungstheorie von Deci und Ryan (1985) und den Interessensbegriff nach Krapp (1992b), auch bekannt als die Person-Gegenstand-Beziehung. Im Rahmen des zweiten Untersuchungsschwerpunktes wird eine für diese Arbeit gültige Definition des Usabilitybegriffs erarbeitet. In diesem Zusammenhang werden Methoden vorgestellt, um die Benutzbarkeit eines Systems zu beurteilen. Zum dritten Schwerpunkt erfolgt die Herleitung des vorliegenden Heterogenitätsverständnisses. Dabei werden Möglichkeiten der Binnendifferenzierung aufgezeigt. Als eine Möglichkeit der Binnendifferenzierung werden zum Schluss des zweiten Kapitels HyperDocSystems und HyperDocs vorgestellt. Dabei ist zu beachten, dass es sich bei dem System einerseits um ein eigenständiges Lernwerkzeug handelt, das zunächst aus technischer und unterrichtspraktischer Sicht beschrieben wird und andererseits um ein Methodenwerkzeug, um digitale binnendifferenzierte Arbeitsblätter im Unterricht zu beforschen. Eng verbunden mit der Konzeption von HyperDocs ist die Theorie des multimedialen Lernens und der Informationsverarbeitung, die mit der Theorie des Cognitive Load in Zusammenhang steht.

Ausgehend von dem Studienüberblick in Kapitel eins und der dargelegten Theorie in Kapitel zwei, erfolgt im dritten Kapitel eine umfassende Betrachtung des aktuellen Forschungsstandes zu den drei gewählten Schwerpunkten. Dieser Forschungsstand führt zur Ausschärfung der Forschungslücken und schließlich zur Formulierung der Forschungsfragen.

Um die Forschungsfragen adäquat beantworten zu können, wurde eine quasi-experimentelle Feldstudie geplant und durchgeführt. Dies bedurfte einer gewissen Methodik, die im vierten Teil der Arbeit detailliert ausgeführt wird. Begonnen wird dabei mit den eingesetzten Methoden und der vorliegenden Datenbasis (Stichprobe und erhobene Variablen). Bei der Erhebung kamen außerdem verschiedene Hilfsmittel zum Einsatz (z. B. Erhebungsinstrumente). Daran schließt sich ein Teilkapitel an, in dem die drei durchgeführten Pilotierungen vorgestellt werden. Die Pilotierungen münden schließlich in die Konzeption einer Hauptstudie mit einer vierstündigen Unterrichtsreihe, deren Gestaltung und Ablauf im Kapitel 4.5 beschrieben werden.

Aus der Hauptstudie konnte eine Vielzahl von Ergebnissen gesichert werden. Diese Ergebnisse werden nun in Kapitel fünf offengelegt. Die Darstellung der Daten erfolgt nach den drei Schwerpunkten. Zentrale Analysemethoden für dieses Kapitel bilden robuste

gemischte Varianzanalysen und Zusammenhangsanalysen. Die Ergebnisse werden in Form der deskriptiven Statistik, Inferenzstatistik und Abbildungen präsentiert.

Nach der Darstellung der Daten im fünften Kapitel erfolgt in Kapitel sechs eine Beantwortung der Forschungsfragen. Dazu werden die Ergebnisse diskutiert und mit dem aktuellen Forschungsstand abgeglichen. Natürlicherweise ergaben sich bei der Gestaltung der Studie gewisse Kompromisse und Erkenntnisse, die im Rahmen der Methodenkritik geäußert werden. Diese mündet schließlich in verschiedene Limitationen. Damit die Ergebnisse dieser Arbeit in der Praxis Relevanz finden, werden außerdem Implikationen und Empfehlungen für die Forschungs- und Schulpraxis gegeben. Für die Forschungspraxis werden außerdem der weitere Forschungsbedarf und Desiderata formuliert.

Das Ende der Dissertation bildet das siebte Kapitel, das die zentralen Erkenntnisse der wissenschaftlichen Arbeit zusammenfasst. Der Ausblick beschreibt abschließend eine Zukunftsperspektive für das Lernen mit digitalen Medien und HyperDocSystems.

## 2. Theorie

In dem nachfolgenden Kapitel werden die theoretischen Grundlagen und Begriffsverständnisse dieser Arbeit dargelegt. Zunächst erfolgt die Definition der Motivation und des Interesses. Damit verbunden ist eine Darlegung des hier verwendeten Motivationsbegriffs sowie des Interessenskonzepts nach Krapp (1992b) und der Self-Determination-Theory nach Deci und Ryan (1985). Daran schließen sich die Herleitung des Usabilitybegriffs und damit verbundene Themen an, wie die Messung der Benutzerfreundlichkeit oder die Abgrenzung zur User Experience. Danach folgt der Theorieteil zur Heterogenität und der Binnendifferenzierung. Zum Schluss erfolgt eine Darstellung von HyperDocSystems. In diesem Zusammenhang wird kurz auf die Theorie des multimedialen Lernens nach Mayer (2014) und die Cognitive Load-Theorie (Sweller, 1988) eingegangen.

### 2.1 Motivation und Interesse

Der Begriff der Motivation wird in unserem alltäglichen Sprachgebrauch vielfältig und oft verwendet. Handlungen, die ohne fremde Anreize getätigt werden, sind intrinsisch motiviert. Teilweise werden die Anreize zum Ausüben der Tätigkeit durch eine externe Quelle bestimmt. In diesem Zusammenhang wird von extrinsischer Motivation gesprochen. In der Motivationsforschung geht es dabei speziell darum, das „Wozu“ und das „Wie“ zu klären. Dabei lässt sich die Motivation zu einer bestimmten Handlung durch zwei Faktoren erklären: Das Streben nach Wirksamkeit sowie die Organisation des Zielengagements und der Zieldistanzierung (Heckhausen & Heckhausen, 2018). Das Streben nach Wirksamkeit ist grundlegend durch die Verhaltensevolution bestimmt. Sie dient dazu, die soziale und physische Umgebung eines Menschen zu kontrollieren (White, 1959). Dieser Anreiz ist nicht allein auf den Menschen beschränkt, sondern er findet sich bei allen Säugetieren wieder (Heckhausen, 2000). Bei dem Zielengagement und der Zieldistanzierung handelt es sich um zwei gegenteilige Prozesse. Es geht bei beiden Modi darum, Ziele zu erreichen oder Ziele zu verwerfen, die zum Beispiel unerreichbar scheinen. Das Zielengagement beschreibt, wie „Unwichtiges ausgeblendet, Wichtiges hervorgehoben [und] Teilhandlungen [...] bereitgestellt werden“ (Heckhausen & Heckhausen, 2018). Dagegen wird bei der Zieldistanzierung in einem aktiven Prozess ein bestimmtes Handlungsstreben verworfen. Die Zieldistanzierung kann das generelle Empfinden von Personen, die sich mit unerreichbaren Zielen konfrontiert sehen, verbessern (Wrosch et al., 2003).

Zielgerichtetes Verhalten wird durch eine Vielzahl von Faktoren beeinflusst. Dazu zählen personenbezogene und situationsbezogene Faktoren, wie folgendes Modell verdeutlicht (vgl. Abbildung 1).

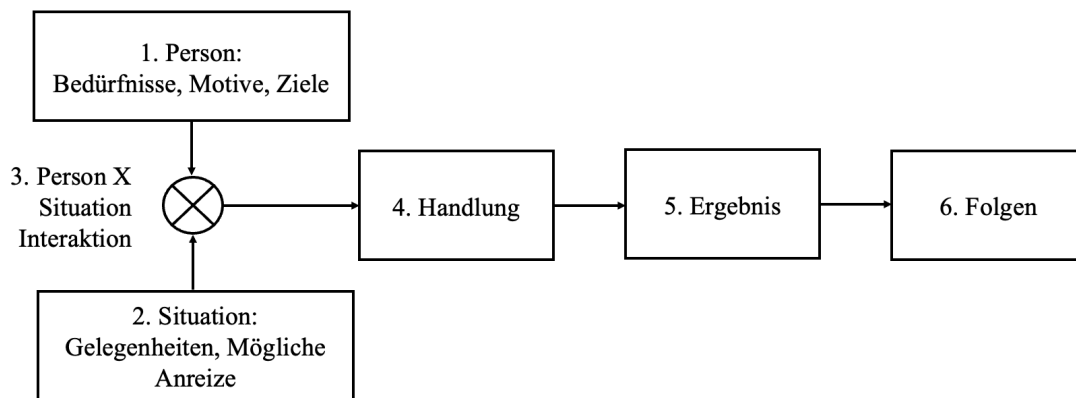


Abbildung 1: Vereinigtes Modell von Heckhausen & Heckhausen (verändert, 2018), bestehend aus dem erweiterten kognitiven Motivationsmodell (Heckhausen, 1977a, b) und dem Modell der klassischen Motivationspsychologie (Heckhausen & Rheinberg, 1980).

Es interagieren dabei sowohl personelle als auch situative Komponenten im Rahmen einer motivationalen Handlung. Zu den Personenfaktoren zählen unter anderem physiologische Bedürfnisse wie Hunger und Durst oder das Streben nach Wirksamkeit. Zu den impliziten Motiven zählen bereits in der Kindheit erlernte Eigenschaften, auf bestimmte Anreize zu reagieren (McClelland et al., 1989). Darin begründet liegt unter anderem auch das unterschiedliche Handeln verschiedener Personen. Als Beispiel führt Bak (2019, S. 129) folgendes an: Verabreden sich zwei Personen zum Tennisspielen, so kann eine Person das Spiel als Leistungsvergleich ansehen, während die andere Person darin eine soziale Handlung sieht. Generell sind Motive Vorlieben für „bestimmte Klassen von thematisch ähnlichen Anreizen“ (Rothermund & Eder, 2011, S. 91). Motive sind dafür verantwortlich, bestimmte Situationen oder Aspekte davon als „Chance, Gefahr oder Handlungsaufforderung“ (Rothermund & Eder, 2011, S. 91) zu sehen. Dabei werden die verschiedenen Motive in Gruppen differenziert: Das Leistungsmotiv (Herausforderungen bewältigen), das Anschlussmotiv (Findung von sozialer Nähe und Bindung) - das obige Beispiel verdeutlicht diese beiden Motive - und das Machtmotiv (soziale Wirksamkeit) (Heckhausen & Heckhausen, 2018; Bak, 2019, S. 132 - 135.). Zu den expliziten Motiven zählen Attribute, die Personen ihrem Handeln selbst zuschreiben. Dabei werden die Anreize dafür oft durch äußere („explicit“; McClelland et al., 1989, S. 693) Faktoren wie Belohnungen oder Erwartungen bestimmt oder selbst auferlegt (McClelland et al., 1989). Das zeigen auch Ryan und Deci (2000) in ihrer Selbstbestimmungstheorie (Bak, 2019, S.

129). Die impliziten und expliziten Motive stimmen nicht immer überein. Unterscheiden sich beide Konstrukte stark voneinander, können Nachteile für die Handlungseffizienz, das Wohlbefinden oder auch die Psyche entstehen (Heckhausen & Heckhausen, 2018; McClelland et al., 1989). Während Motivation als ein genereller Drang nach einem bestimmten Verhalten angesehen werden kann, bestimmen Motive die genaue Richtung und Qualität des Verhaltens (Bak, 2019, S. 128). Die situative Komponente des Modells (vgl. Abbildung 1) unterscheidet zwischen intrinsischen und extrinsischen Anreizen. Dabei treiben Anreize eine gewisse Handlung voran. Intrinsische Anreize entstehen aus einer Handlung oder dem Ergebnis dieser. Dagegen sind extrinsische Anreize geprägt durch eine externe Bewertung oder gegenständliche Belohnung. Die Wechselwirkung der personellen und situativen Komponente führt schließlich zu einer Handlung. Dazwischen liegen verschiedene Teilprozesse, wie beispielsweise die Bildung und Initiierung der Intention einer Handlung. Nach Durchführung der Handlung wird die Intention deaktiviert und das Ergebnis sowie dessen Folgen bewertet (Heckhausen & Heckhausen, 2018).

### **2.1.1 Vergleich zwischen intrinsischer und extrinsischer Motivation**

In der Motivationsforschung wird zwischen intrinsischer und extrinsischer Motivation unterschieden. Die Begriffe gehen vermutlich auf Woodworth (1918, S. 67 - 70) zurück und sind zum Teil unscharf definiert (Rheinberg & Engeser, 2018). Beide Arten der Motivation haben Auswirkung auf unser Handeln. Intrinsische Motivation hat eine hohe Relevanz für das Lernen (Ryan & Deci, 2020). Die intrinsische Motivation richtet sich nach einem für uns positiven Zielstand aus, der auch darin bestehen kann, bestimmte negative Ereignisse zu umgehen (Rheinberger & Engeser, 2018). Der Anreiz, eine bestimmte Handlung auszuführen, liegt im erwarteten Zielzustand. Dem gegenüber steht jedoch ein Anreiz, der aus der Tätigkeitsausführung selbst stammt. In manchen Fällen wird der tätigkeitsbezogene Anreiz als intrinsisch definiert und ein bestimmter Zielzustand als extrinsisch. Dieser extrinsische Endzustand kann auch tätigkeitshemmend wirken. Teilweise werden Tätigkeiten auch durch extrinsische Motivation begonnen und gehen im Laufe der Durchführung in eine intrinsische Motivation über (Rheinberger & Engeser, 2018; Woodworth, 1918, S. 67 - 70). Der Zusammenhang zwischen intrinsischer und extrinsischer Motivation wurde zum Beispiel von Vallerand (1997) in einem mehrstufigen Modell, ausgehend von den Arbeiten von Deci und Ryan (1985), festgehalten. Nach seinem Modell existieren drei Ebenen, die jeweils für sich eine eigene Ausprägung der Motivation (intrinsisch, extrinsisch und Amotivation) besitzen und mit den anderen

Motivationshierarchien in Verbindung stehen: die situationale Ebene, die kontextuale Ebene und die globale Ebene (Vallerand, 1997). Dieser Zusammenhang verdeutlicht die Komplexität, in der sich die verschiedenen Betrachtungen der Motivation bewegen.

Nach Deci und Ryan (1985, S. 32) ist intrinsische Motivation bestimmt durch das Erleben von Kompetenz und Autonomie beziehungsweise Selbstbestimmung. Ergänzt werden diese Bedürfnisse durch die soziale Eingebundenheit, woraus sich schließlich die Self-Determination-Theory begründete (Deci & Ryan, 1985; Rheinberger & Engeser, 2018). Eine spezielle Form der intrinsischen Motivation wird als Interesse bezeichnet. Dabei steht die Motivation gegenüber einem bestimmten Gegenstand oder Themengebiet im Vordergrund. Der pädagogische Interessensbegriff nach Krapp (1992b) ist angelehnt an die Selbstbestimmungstheorie. Demnach agieren Personen intrinsisch motiviert, wenn die Lernaktivität selbstbestimmt erfolgt und die Auseinandersetzung mit dem Lerngegenstand als wertvoll angesehen wird (Rheinberger & Engeser, 2018). Zudem bestehen noch weitere Definitionen von intrinsischer Motivation, die auf die eben genannten Erklärungen zurückgehen beziehungsweise diese teilweise anders akzentuieren (Rheinberger & Engeser, 2018; Vallerand, 1997).

Extrinsische Motivation findet eine besondere Rolle in der Schule, in der Noten und vorgegebene Lernsituationen anzutreffen sind. Auf der einen Seite werden Schüler:innen durch ihre eigenen Interessen, das Erleben von Autonomie, Kompetenz und soziale Eingebundenheit intrinsisch motiviert. Auf der anderen Seite wirken externe Faktoren, wie zum Beispiel die bevorstehende Klassenarbeit, auf das Verhalten und die Intentionen eines Lernenden (Vallerand, 1997). Gerade ab dem Jugendalter ist das Verhalten zunehmend geprägt von extrinsischer Motivation, indem an den Menschen bestimmte soziale Erwartungen gestellt werden und eine bestimmte soziale Rolle eingenommen wird (Ryan & Deci, 2020). Zunächst wurde angenommen, dass extrinsische Motivation nur durch außenstehende Faktoren, zum Beispiel Lehrkräfte oder bestimmte Ereignisse, hervorgerufen wird. Extrinsische Motivation ist jedoch ebenfalls mehrperspektivisch zu betrachten, indem eine schrittweise Internalisierung des extrinsischen Einflusses stattfinden kann (Deci & Ryan, 1985, S. 129 – 132, 261). Dazu spielt die Art des externen Faktors jedoch eine wichtige Rolle (Deci, 1971). Durchaus kann eine Person ein bestimmtes Verhalten zeigen, das nicht aus Vergnügen, aber dennoch durch eine bewusste eigene Entscheidung bestimmt ist. Beispielsweise kann eine Schülerin oder ein Schüler lernen, um den Normen der Eltern beziehungsweise Lehrkräfte zu entsprechen (nicht sehr selbstbestimmt) oder



da sie oder er einen bestimmten Sinn oder eine Relevanz im Lernen sieht (stärker selbstbestimmt). Damit spielt die Art der extrinsischen Motivation gerade für Lehrkräfte eine große Rolle, da der einwirkende Faktor aufgrund des Grades der Autonomie verschieden wahrgenommen wird (Ryan & Deci, 2020; Vallerand, 1997). Extrinsische Motivation kann schließlich in intrinsische Motivation münden. Zwischen den beiden Arten der Motivation herrscht ein gewisses Kontinuum. Letztlich scheint jedoch die intrinsische Motivation einen größeren Einfluss auf das gezeigte Verhalten und das Wohlbefinden einer Person zu haben als die extrinsische Motivation (Vallerand, 1997). Insgesamt haben sowohl die extrinsische als auch die intrinsische Motivation Einfluss auf das Lernen (Deci & Ryan, 2000).

### **2.1.2 Intrinsische Motivation nach der Self-Determination-Theory**

Die Self-Determination-Theory (Selbstbestimmungstheorie, SDT) geht auf Deci & Ryan (1985) zurück und wurde über mehrere Jahrzehnte entwickelt, ausgeschärft und in verschiedenen Lebensbereichen wie der Arbeit, Bildung oder in Familien erprobt (Ryan & Deci, 2017, S. 17 -19). Die Autoren beschreiben die Selbstbestimmungstheorie als eine „organismische und dialektische Theorie der menschlichen Motivation“ (Deci & Ryan, 1993, S. 366). Damit ist zum einen das natürliche Streben des Menschen nach Neugier und Interesse und zum anderen die stetige Wechselwirkung zwischen der sozialen Umwelt und der Integration von motivationalen Bedürfnissen gemeint (Deci & Ryan, 1993; Ryan & Deci, 2017, S. 4). Im Mittelpunkt der Betrachtung steht dabei das menschliche Bedürfnis nach Autonomie, Kompetenz und sozialer Eingebundenheit (Ryan & Deci, 2017, S. 5).

Die Theorie erklärt auf Ebene der Psychologie das menschliche Handeln als Kontinuum zwischen extrinsischer und intrinsischer Motivation. Dabei wird zwischen Amotivation, extrinsischer Motivation und intrinsischer Motivation unterschieden. Als Amotivation werden dabei Verhaltensweisen beschrieben, die keinem genauen Ziel folgen oder aus einem Impuls heraus entstehen (Ryan & Deci, 2017, S. 16). Ursache dafür ist zum Beispiel die fehlende Wertschätzung gegenüber der Tätigkeit, mangelnde Kompetenz oder fehlende Unterstützung (Ryan, 1995). Auch fehlendes Interesse kann zu Amotivation führen (Ryan & Deci, 2017, S. 16).

Handlungen werden als selbstbestimmt oder fremdbestimmt wahrgenommen. Die Rolle der intrinsischen und extrinsischen Motivation wurde in diesem Zusammenhang bereits

beschrieben. Im Sinne der Selbstbestimmungstheorie werden diese Unterscheidungen der Motivation nachfolgend definiert. Als intrinsisch motiviert gelten Personen, bei denen die Handlung interessensbestimmt und nicht von außerhalb beeinflusst ist. Zudem zeigen diese Menschen Neugier und Spontanität. Als weniger selbstbestimmt gelten extrinsisch motivierte Handlungen, da diese nicht aus Spontanität oder Interesse heraus entstehen, sondern aus dem Erlangen einer bestimmten Konsequenz resultieren. Diese Konsequenz ist meistens Teil eines äußeren Einflusses. Intrinsische und extrinsische Motivation sind nicht als Gegensätze zu verstehen (Deci & Ryan, 1993). Dennoch ist die Art der Motivation von großer Bedeutung. In der so genannten *Goal Contents Theory* beschreiben Ryan und Deci (2017, S. 272 - 276) die Auswirkung von intrinsischer und extrinsischer Motivation auf das Wohlbefinden, beispielsweise das Empfinden von Glück. Menschen mit extrinsisch motivierten Handlungen beziehungsweise Zielen weisen ein geringeres Wohlbefinden auf als solche, die intrinsisch motivierte Ziele anstreben. Damit sind nicht alle Ziele gleichbedeutend.

Insgesamt wird zwischen vier Arten von extrinsischer Motivation unterschieden, die in ihrer wahrgenommenen Selbstbestimmtheit zunehmen (vgl. Abbildung 2). Die verschiedenen Typen der Motivation und deren Überführung ineinander werden auch in einer untergeordneten Theorie, der *Organismic Integration Theory* beschrieben (Deci & Ryan, 1985, S. 129 - 140; Ryan & Deci, 2000). Externale Regulation bezeichnet Verhaltensweisen, die durch eine externe Belohnung oder eine bevorstehende Strafe ausgeführt werden. Diesen Handlungen liegen keine Freiwilligkeit und keine Autonomie zugrunde. Introjierte Regulation ist nicht mehr vollständig external reguliert. Handlungen resultieren aus einem inneren Druck heraus, der zum Beispiel aus sozialen Normen heraus resultiert. Das Ziel kann dabei sein, die Selbstachtung zu wahren. Zunehmend selbstbestimmt ist die identifizierte Regulation, indem eine Handlung als für einen selbst relevant angesehen wird. Die dahinterliegenden Normen und Ziele wurden in das Selbstkonzept integriert. Zwar besteht immer noch ein gewisser Einfluss von außerhalb, jedoch führt die Person eine Aktion auch aus einer inneren Bestrebung heraus aus. Der höchste Grad der Selbstbestimmung liegt bei der integrierten Regulation vor. Hierbei hat die betroffene Person Werte, Normen und Ziele in das Selbstbild integriert. Im Gegensatz zur intrinsischen Motivation ist die Regulation jedoch kein vollwertiger Teil des Selbstbildes, da die Handlung nicht internal, aus dem Interesse heraus, hervorgerufen wird (Deci & Ryan, 1993).

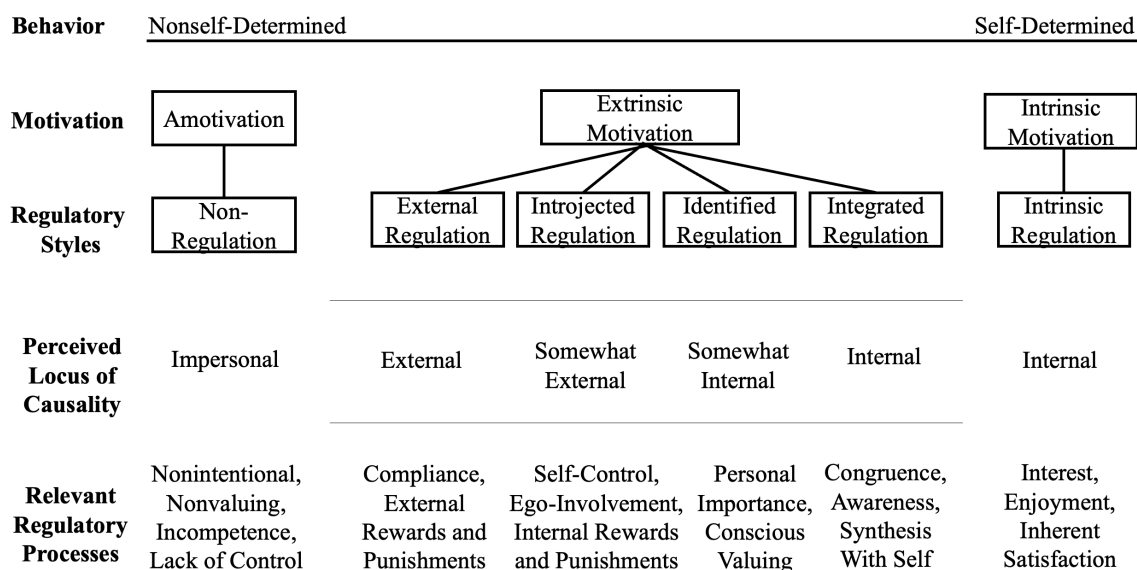


Abbildung 2: Das Modell der Selbstbestimmungstheorie im Überblick (verändert nach Ryan & Deci, 2000).

Die beiden Autoren der Self-Determination Theory verdeutlichen die vier Arten der extrinsischen Motivation anhand des nachfolgenden Beispiels (Deci & Ryan, 1993): Bei der Abiturprüfung lernt ein:e Schüler:in rein external reguliert, wenn sie oder er die Strafe ihrer oder seiner Eltern fürchtet. Introjiert lernt sie oder er, weil bei Nichtbestehen gegebenenfalls eine Bloßstellung gegenüber Freund:innen besteht. Als identifiziert reguliert gilt die oder der Lernende, wenn sie oder er ein gutes Abitur anstrebt, um einen gewünschten Studiengang studieren zu können. Integriert reguliert handelt die oder der Schüler:in, indem sie oder er verschiedene, zum Teil gegensätzliche Ziele beziehungsweise Rollen, zum Beispiel die einer guten Sportlerin oder eines guten Sportlers und Abiturientin oder Abiturient, vereint und beide Bestrebungen in Einklang miteinander bringt, ohne dabei Stress zu empfinden.

Zentraler Bestandteil der Selbstbestimmungstheorie sind die, im Sinne einer psychologischen Interpretation, menschlichen Grundbedürfnisse nach Autonomie, Kompetenz und sozialer Eingebundenheit. Das Erfüllen dieser Bedürfnisse ist essenziell für die gesunde menschliche Entwicklung, das menschliche Handeln und Wohlbefinden (Ryan, 1995). Das Erfüllen oder der Entzug der psychologischen Bedürfnisse ist objektiv messbar (Ryan & Deci, 2017, S. 10). Das Fehlen eines dieser Grundbedürfnisse führt zu einer Verminderung des Wohlbefindens beziehungsweise der intrinsischen Motivation (Ryan & Deci, 2000; Ryan & Deci, 2017, S. 11). Das Autonomiebedürfnis und das Bedürfnis von Kompetenzerleben wurden in der *Cognitive Evaluation Theory* (Ryan & Deci, 2017, S. 124) spezifiziert, um die unterschiedlichen Ausprägungen intrinsischer Motivation

erklären zu können. Gleichzeitig wird in der Theorie bereits der positive Einfluss der sozialen Eingebundenheit auf die Motivation beschrieben (Deci & Ryan, 1993). Die *Cognitive Evaluation Theory* erklärt ebenfalls den negativen Einfluss externer Faktoren, wie beispielsweise Druck, auf die intrinsische Motivation. Der Einfluss von Autonomie- und Kompetenzerleben lässt sich allerdings nur auf intrinsisch motivierte Tätigkeiten und nicht auf external indizierte Handlungen anwenden. Gleichzeitig können diese beiden Bedürfnisse allerdings die Internalisierung und Integration extrinsisch motivierter Handlungen fördern (Ryan & Deci, 2000).

Autonomie ist im Sinne der Freiwilligkeit von Handlungen definiert, nicht im Sinne der Unabhängigkeit. Das eigene Handeln wird als selbstbestimmt wahrgenommen und stimmt mit den eigenen Werten und Interessen überein. Dem gegenüber stehen Verhaltensweisen, die als nicht autonom wahrgenommen werden und damit in Konflikt mit einem selbst stehen. Nur wenige Tätigkeiten sind wirklich selbstreguliert im Sinne der Autonomie (Ryan & Deci, 2017, S. 10). Die wahrgenommene Autonomie einer Person nimmt außerdem Einfluss auf eine Vielzahl weiterer Attribute, die im Kontext des Lernens als wichtig erachtet werden. Dazu zählen unter anderem, neben der intrinsischen Motivation und dem Erleben von Kompetenz, die Aufgabenperformanz und die aufgebrauchte Anstrengung (Patall et al., 2008). Dabei gilt die wahrgenommene Wahlfreiheit als ein wichtiger Einflussfaktor auf das Autonomieerleben (Patall et al., 2008, Schneider et al., 2018). Kompetenz im Sinne der Selbstbestimmungstheorie erfahren Personen, wenn sie ihre Handlungen als effektiv wahrnehmen und sie wirklich beherrschen. Ryan und Deci (2017, S. 11) verwenden in diesem Zusammenhang auch den Ausdruck „mastery“. Das Kompetenzerleben kann in jedem Lebenskontext, zum Beispiel beim Spielen und Lernen, auftreten. Gleichzeitig wird es durch Überforderung, ständige negative Rückmeldung und zwischenmenschliche Interaktionen mit negativem Ausgang (Kritik oder soziale Vergleiche) unterbunden. Soziale Eingebundenheit erfahren Menschen, wenn sie den Eindruck haben, dass andere sich um sie sorgen, beziehungsweise kümmern. Damit einher geht ebenso ein Gefühl nach Zugehörigkeit und Wichtigkeit (Ryan & Deci, 2017, S. 11).

### **2.1.3 Interesse nach der Person-Gegenstands-Beziehung**

Die Bedeutung von Interesse beim Lernen hat sich bereits in der Selbstbestimmungstheorie gezeigt. Hier gilt unter anderem das Interesse als ein Hauptprädiktor der intrinsischen Motivation. Interesse kann demnach auch als eine besondere Form von

intrinsischer Motivation definiert werden (Rheinberger & Engeser, 2018). Eine pädagogische Einordnung von Interesse wurde von Krapp (z. B. Krapp 1992b) vorgenommen. Das Interessensverständnis von Krapp baut auf den Grundsätzen der Self-Determination Theory auf. Demnach besteht die Persönlichkeitsstruktur aus einem kognitiven Selbst, das im Laufe der Zeit Wissen und Fähigkeiten anhäuft, sowie einer emotionsgeprägten Motivationskompetente (Krapp, 1992b). Im Laufe der Entwicklung erlebt das Individuum Veränderungen, da es zum einen seine Grundbedürfnisse nach Autonomie, Kompetenzorientierung und soziale Eingebundenheit befriedigen muss und zum anderen die soziale und gegenständliche Umwelt neue Anforderungen daran stellt. Eine teilweise Übereinstimmung dieser beiden, zum Teil gegensätzlichen Ansprüche, erreicht eine Person durch die Internalisierung der extrinsischen Motive. Eine vollständige Übereinkunft der intrinsischen und extrinsischen Motive ist nicht möglich (Krapp, 1992b).

Interesse ergibt sich aus der Wechselwirkung einer Person mit seiner Umwelt (Prenzel et al., 1986). Es ist inhalts- beziehungsweise gegenstandsspezifisch (Krapp, 1998). Damit ist das Interesse einer Person als Konstrukt der „Person-Gegenstands-Beziehung“ (Krapp, 1992b, S. 300) zu verstehen. Der Begriff „Gegenstand“ umfasst hierbei einen bestimmten Ausschnitt und eingegrenzten Bereich der Umwelt, der sich von anderen Bereichen abgrenzen lässt. Sogenannte „Referenzobjekte“ (Prenzel et al., 1986, S. 166) sind Teil des Gegenstandes. Als Beispiel führen Prenzel und Kollegen (1986) Noten und Instrumente für den Gegenstand Musik auf. Als Gegenstände werden damit verschiedene „Sachverhalte in der Lebenswelt eines Menschen“ (Krapp, 1992b, S. 304) definiert. Insgesamt wird in drei verschiedene Arten von Gegenstandsbereichen differenziert: reale Objekte, Tätigkeiten und Themen (Krapp, 1992b). Je nach betrachteter Dauer der Relation zwischen Gegenstand und Person wird zwischen einer Beziehung und einem Bezug gesprochen. Beziehung definiert eine aktuelle Gegenstandsauseinandersetzung, Bezug ein über längere Zeit andauerndes Interesse, das sich jedoch nur zeitweise zeigt. Die Auseinandersetzung kann ungerichtet oder zielgerichtet erfolgen. Letzteres wird als „Interessenshandlung“ (Krapp, 1992b, S. 306) bezeichnet. Dabei variiert die Intensität der Auseinandersetzung. Die Relation zwischen Interesse und Person ist durch verschiedene Bereiche geprägt. Auf der kognitiven Ebene ist Interesse hoch komplex und vielfältig, da eine Person ein umfangreiches Wissen über den Gegenstandsbereich aufweist. Auf der emotionalen Ebene ist die Auseinandersetzung mit Interesse durch positive Gefühle geprägt, wie es für das Flow-Erleben typisch ist. Der dritte Bereich ist der Wertebereich. Der

Interessensgegenstand ist für einen selbst wertvoll. Die Person ordnet ihm einen hohen persönlichen Stellenwert zu (Prenzel et al., 1986).

Menschen handeln und lernen aus Interesse. Das Interesse gegenüber einem Gegenstand kann temporär oder dauerhaft in das Selbstbild einer Person integriert werden. Im Bereich des Schulwesens variiert das Interesse an Themen im Laufe der Schulbildung. Zum Teil bestehen Unterschiede zwischen den Geschlechtern und Schulfächern (z. B. Jansen et al., 2013). Die Entwicklung von Interesse lässt sich nach dem Modell von Krapp (1998) genauer erklären (vgl. Abbildung 3).

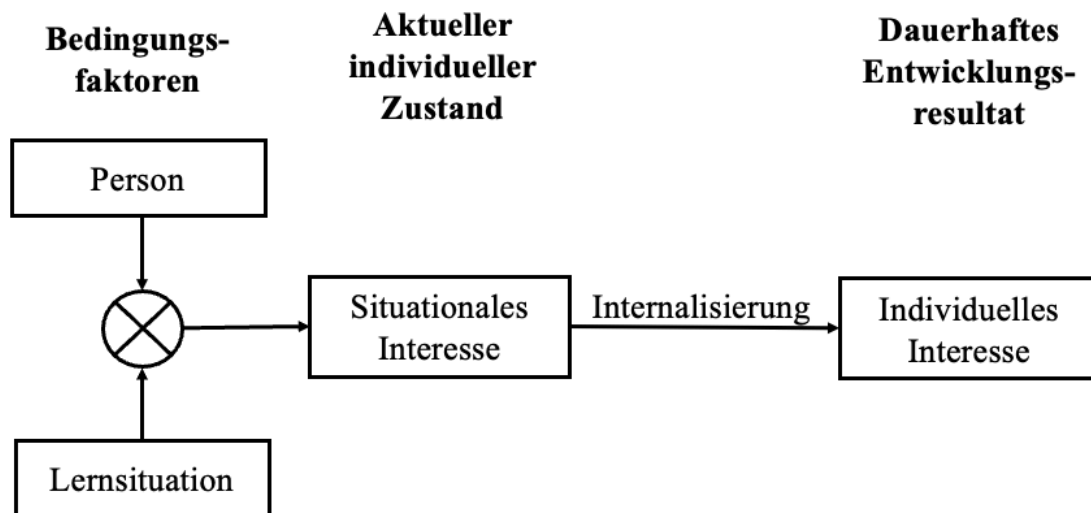


Abbildung 3: Interessensmodell nach Krapp (1998, eigene Darstellung).

Demnach lässt sich zwischen situationalem Interesse und individuellem Interesse unterscheiden. Bei situationalem Interesse handelt es sich um einen temporären Zustand, der insbesondere durch die Relation zwischen der Person und den verschiedenen Faktoren der Situation hervorgerufen wird, wie zum Beispiel die Neuheit einer Umgebung (Krapp et al., 1992). Im Unterricht wird situationales Interesse insbesondere durch die Aufbereitung und Darbietung des Lerngegenstandes geschaffen, wie zum Beispiel durch den Einsatz von Computern (Mitchell, 1993). Damit einher geht eine höhere Aufmerksamkeit und eine bessere mentale Verarbeitung des Gegenstands (Krapp, 1992a; Krapp, 1998). Diese Art des Interesses ist weniger spezifisch und lässt sich in vielen Individuen antreffen. Es motiviert dazu, mit der Umgebung zu interagieren und neue Informationen zu erschließen (Krapp et al., 1992). Individuelles Interesse kann entweder als Disposition oder als Ist-Zustand betrachtet werden. Teilweise wird es auch als „persönliches Interesse“ (Krapp, 1992b, S. 309) beschrieben. Im Sinne einer Disposition bleibt das Interesse über einen langen Zeitraum bestehen. Es nimmt starken Einfluss auf das Lernverhalten

eines Individuums, aber es lässt sich nicht unmittelbar beobachten. In einigen Fällen dient es zur Vorhersage des akademischen Erfolgs. Interesse als aktueller Zustand lässt sich durch eine erhöhte Aufmerksamkeit und das Erleben von Zufriedenheit charakterisieren (Krapp et al., 1992). Der Fokus der Betrachtung liegt hierbei mehr auf konkreten Handlungen und Auseinandersetzungen mit dem Gegenstand. Durch das individuelle Interesse wird demnach deklaratives und prozedurales Wissen angeeignet (Krapp, 1992a).

Die Bildung von situationalem Interesse ist ein Produkt aus den Merkmalen einer Person (individuelles Interesse) und den Merkmalen der Lernumgebung (Interessanz), woraus sich eine interessenorientierte Tätigkeit entwickelt. Daneben besteht das aktualisierte Interesse, das enger in Zusammenhang mit dem individuellen Interesse wirkt (Krapp, 1992a). Individuelles und situationales Interesse beeinflussen sich somit gegenseitig. Bei gering entwickeltem individuellem Interesse spielt die Interessanz der Umwelt, beziehungsweise im konkreten Fall des Lernens die Lernumgebung, eine wichtige Rolle. Das Individuum entscheidet, sich mit dem Lerngegenstand auseinanderzusetzen und entwickelt zunehmendes Interesse. Bei bereits ausgeprägtem individuellem Interesse spielt die Interessanz der Lernumgebung kaum eine Rolle, sodass die Person die Bedingungen seiner Umwelt verändert oder andere Situationen aufsucht (Krapp, 1992a). Situationales Interesse ist einerseits abhängig von individuellem Interesse, andererseits führt es jedoch auch zu individuellem Interesse. Es handelt sich dabei um einen mehrstufigen Prozess, der über zwei Stufen verläuft. Nach Hidi und Renninger (2006) kann auch ein vierstufiges Modell der Interessensentwicklung angenommen werden, auf das an dieser Stelle jedoch nicht eingegangen wird.

Zunächst erfolgt eine Überführung der anfänglichen Neugier in eine andauernde Bereitschaft, sich mit dem Lerngegenstand auseinanderzusetzen. Anschließend muss die Person selbst den Interessensgegenstand ohne äußere Einwirkung als interessant wahrnehmen und sein Wissen darüber erweitern. Bei diesem Prozess reicht es nicht aus, Neugierde zu erwecken (Krapp, 1998). Mitchell (1993) unterscheidet daher zwischen *catch*-Komponenten, die die Aufmerksamkeit des Lernenden auf sich ziehen und einer *hold*-Komponente, die für eine dauerhafte Ausbildung des individuellen Interesses sorgt. Die *catch*-Komponente dient zur Stimulierung der Lernenden, während die *hold*-Komponente zu einer dauerhaften Auseinandersetzung mit dem Lerngegenstand befähigt. Beispiele für eine *catch*-Komponente sind demnach Gruppenarbeiten oder Computer beziehungsweise digitale Medien. Beispiele für die *hold*-Komponente sind Eingebundenheit und

Bedeutsamkeit. Eingebundenheit meint in diesem Zusammenhang das Gefühl, inwieweit Lernende aktiv am Lernprozess teilhaben können. Ochsen et al. (2021) konnten das Modell stützen, indem sie die Passivität der Lernenden als interessenshemmend identifizierten. Unter Bedeutung wird die Relevanz für den Alltag verstanden (Mitchell, 1993). In einer späteren Untersuchung stellte Mitchell (1997) zudem fest, dass eine Lernumgebung mit einem hohen situationalen Interesse einen Einfluss auf das individuelle Interesse nehmen kann. Gleichzeitig nimmt das individuelle Interesse in Lernumgebungen mit geringem situationellem Interesse ab. Neben den von Mitchell (1993) aufgeführten Faktoren wurden noch weitere Einflussfaktoren auf das situationale Interesse gefunden. Dazu gehört unter anderem Neuheit (Palmer, 2004; Palmer, 2009), „hands-on“-Aktivitäten (Palmer et al., 2016), Autonomie oder auch soziale Eingebundenheit, wobei Neuheit als einer der stärksten Einflussfaktoren angesehen wird (Palmer, 2009). Damit existiert eine Vielzahl an Möglichkeiten, das situationale Interesse zu fördern und bei längerer Exposition in individuelles Interesse zu überführen. Insgesamt handelt es sich jedoch um einen sehr komplexen Prozess, der von vielen Faktoren abhängig ist.

## **2.2 Usability**

Nachdem die theoretischen Grundlagen zum ersten Schwerpunkt dieser Arbeit, der Motivation und dem Interesse, gelegt wurden, folgt nun mit dem zweiten Schwerpunkt eine Herleitung des Usabilitybegriffs. Damit verbunden ist ein Überblick über die verschiedenen Methoden zur Messung der Benutzbarkeit und eine Abgrenzung zum Begriff der *User Experience*.

### **2.2.1 Begriffsdefinition**

Der Begriff der Usability spielt in vielen Kontexten der *Human Computer Interaction* eine wichtige Rolle, woher er auch in Zusammenarbeit mit der Psychologie seinen Ursprung verzeichnet (Bevan, 1995a). Ein jedes elektronisches Gerät in unserem Alltag weist eine geringere oder höhere Benutzbarkeit auf. Das Produkt kann damit einfach oder schwierig in der Bedienung sein. Dabei existieren viele Definitionen des Begriffs Usability (Lewis, 2012; Richter & Flückiger, 2016, S. 10). Aus der ursprünglichen Entwicklung heraus wurde zunächst unter den Softwareentwickler:innen der Begriff „Benutzerfreundlich“ („user friendly“; Nielsen, 1993, S. 23) verwendet, um die Nutzerin oder den Nutzer stärker in den Vordergrund zu rücken. Der Begriff wurde allerdings seit den 1980er



Jahren unter unterschiedlichen Konzepten und Definitionen verwendet und daher schließlich von der Usability abgelöst (Bevan et al., 1991). Heutzutage wird die Usability gelegentlich als Gütekriterium der Benutzeroberfläche verstanden und sie ist eng verbunden mit der *User Experience*. Teilweise hat der Begriff der User Experience den Begriff der Usability abgelöst (Richter & Flückiger, 2016, S. 12). Die Benutzbarkeit umfasst hierbei einen engeren Definitionsbereich, der insbesondere die Funktionalität des Systems, mit der der oder die Nutzer:in eine Aufgabe erfolgreich erledigen kann, einschließt. Bei der User Experience hingegen spielen die Ästhetik der Benutzeroberfläche und die hervorgerufenen Gefühle sowie Wahrnehmungen ebenfalls eine Rolle (Richter & Flückiger, 2016, S. 12; Tullis & Albert, 2013, S. 5). Es bestehen, je nach betrachteter Rolle, verschiedene Sichtweisen auf die Usability. Der produktorientierte Ansatz fokussiert die Ergonomie des Produktes, der benutzerorientierte Ansatz bezieht sich auf die mentale Last der Nutzung und die Haltung der Nutzerin oder des Nutzers, und der leistungsbezogene Ansatz nimmt die Interaktion zwischen Produkt und Nutzer:in in den Fokus. Dabei spielen insbesondere die Einfachheit der Nutzung und die Akzeptanz eine wichtige Rolle (Bevan et al., 1991, Bevan, 1995a). Abbildung 4 verdeutlicht die Rolle der Usability in einem sogenannten „Mensch-Maschine-System“ (Richter & Flückiger, 2016, S.11).

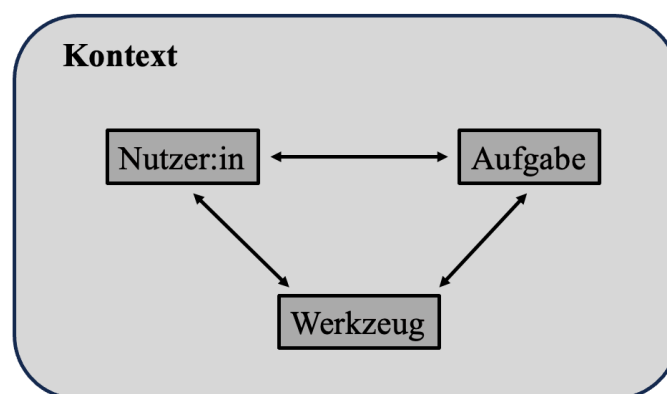


Abbildung 4: Ganzheitliche Betrachtung der Usability (eigene Darstellung nach Richter & Flückiger, 2016).

Dabei möchte der oder die Nutzer:in in einem bestimmten Kontext mithilfe eines Werkzeuges eine Aufgabe erledigen. Im konkreten Fall könnte das bedeuten, dass ein oder eine Schüler:in (Nutzer:in) in einer Chemiestunde (Kontext) mithilfe eines HyperDocs (Werkzeug) Aufgaben (Aufgabe) zur Säure-Base-Chemie bearbeiten muss. Nach Nielsen (1993, S. 24 - 26) ist Usability nicht ein eindimensionales Konstrukt, sondern sie besteht aus mehreren Faktoren: Erlernbarkeit, Effizienz, Einprägsamkeit, Fehlerrate und Zufriedenheit. Er ordnet die Usability ferner der Nützlichkeit eines Systems zu, die ihrerseits

wiederum zur Praktikabilität und schließlich zur Systemakzeptanz zählt (vgl. Abbildung 5). Nach Meinung des Autors handelt es sich bei der Erlernbarkeit um das wichtigste Attribut eines Produktes, damit unerfahrene Benutzer:innen schnell in die zuverlässige Bedienung des Systems finden (Nielsen, 1993, S. 27 - 28).

Eine genaue Spezifikation der Usability wurde in der ISO-Norm 9241-11 definiert (ISO, 2018). In der Norm wird von der Gebrauchstauglichkeit gesprochen. Diese umfasst in der neuesten Fassung von 2016, die die Fassung von 1998 ablöst, im Wesentlichen die Effektivität, Effizienz und Zufriedenheit in einem bestimmten Nutzungskontext. Die Usability gilt auch für die Teilbereiche Lernbarkeit des Systems, regelmäßige Nutzung, Fehlervermeidung, Barrierefreiheit und Wartungsfreundlichkeit (Bevan et al. 2015; Bevan et al. 2016; Richter & Flückiger, 2016, S. 11). Damit zeigen sich Parallelen zu den von Nielsen genannten Attributen. Als Schnittmenge zwischen den verschiedenen Definitionen können die Erlernbarkeit, Effizienz, Zufriedenheit und die Fehlervermeidung genannt werden.

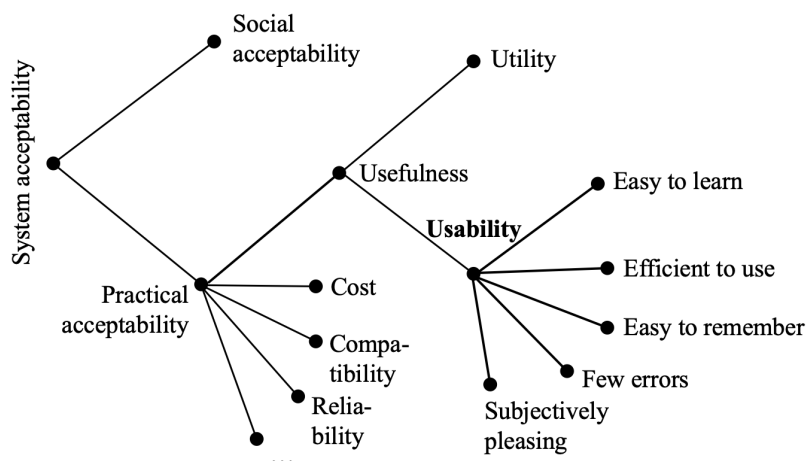


Abbildung 5: Modell der Systemakzeptanz, der die Usability untergeordnet ist (verändert nach Nielsen, 1993, S. 25).

In der neueren Version der ISO-Norm 9241-11 wird insbesondere die Relevanz der User Experience betont. Während in den Anfängen der Internetnutzung (Web 1.0) insbesondere die Performanz einer Nutzer-Produkt-Interaktion (Effektivität und Effizienz) im Vordergrund stand, gewannen mit Voranschreiten des Web 2.0 die subjektiven Reaktionen und Emotionen der Nutzungserfahrung zunehmend an Bedeutung (Bevan et al. 2015). Einige Autor:innen (z. B. Hellman & Rönkkö, 2008; Hassenzahl et al., 2010) sehen die Einfachheit der Nutzung als Teil der Usability daher zunehmend als eine gegebene notwendige Voraussetzung für ein Produkt, welche von den Nutzer:innen erwartet

wird. Sind hingegen Teileaspekte der Usability, wie zum Beispiel die Lernbarkeit, Effektivität oder Effizienz, verletzt, führt das bei den Benutzer:innen zu einer Unzufriedenheit, während eine Erfüllung all dieser Punkte nicht zwangsläufig zu einer hohen Einschätzung der Benutzerbarkeit führt (Hellman & Rönkkö, 2008).

Im Folgenden werden die jeweiligen Aspekte der Usability, die bereits in Teilen genannt wurden, nach der ISO-Norm 9241-11 (ISO, 2018), Barnum (2021, S. 12 - 14) und Bevan et al. (2015) zusammengefasst:

- **Effektivität:** Unter der Effektivität versteht man die Genauigkeit und Vollständigkeit, mit der ein bestimmtes Ziel erreicht werden kann. In der neueren Fassung der ISO-Norm 9241-11 wird zusätzlich die Angemessenheit ergänzt, womit die Genauigkeit und Vollständigkeit verbessert beziehungsweise negative Ergebnisse vermindert werden sollen. Die Angemessenheit bezieht sich dabei auf die Genauigkeit des Resultats und die Vermeidung von Fehlern.
- **Effizienz:** Ursprünglich wurde die Effizienz als Verhältnis zwischen Effektivität und Ressourcenverbrauch definiert. Dieses Maß wurde aufgrund seines begrenzten Verwendungszwecks überarbeitet und definiert als erwarteter Ressourcenverbrauch (Kosten, Zeit, Material und menschliches Bemühen), um ein Ziel zu erreichen.
- **Zufriedenheit:** Die Zufriedenheit wurde ursprünglich gemessen an der positiven Einstellung gegenüber der Verwendung eines Produkts. Dabei soll der oder die Nutzer:in keinerlei Unbehagen oder Unannehmlichkeiten spüren. In der neusten Fassung wird die Zufriedenheit breiter definiert. Sie umfasst alle emotionalen und physiologischen Auswirkungen der Nutzung eines Produktes, Dienstes oder Systems.
- **Kontext der Nutzung:** Der Kontext der Nutzung eines Systems, Dienstes oder Produktes, in dem die Usability gültig ist, kann sich entweder auf den Gesamtkontext, speziellere Kontexte (einzelne Funktionen etc.) oder auch den einzelnen individuellen Kontext eines Nutzers oder einer Nutzerin beziehen.

Die Effektivität, Effizienz und Zufriedenheit lassen sich auch als „Qualität der Nutzung“ („quality of use“; Bevan, 1995a, S. 352) zusammenfassen. Effektivität und Effizienz sorgen dafür, dass Anwender:innen ein bestimmtes Ziel mit einer ausreichenden Genauigkeit und Geschwindigkeit erreichen. Das Produkt soll einen Mehrwert in einem bestimmten Nutzungskontext darstellen, der ohne dieses Produkt nicht gegeben ist. Aufgrund dieses

Funktionszugewinns verwenden Personen schließlich das Werkzeug (Barnum, 2021, S. 12). Die Zufriedenheit wird heutzutage als der wichtigste Faktor des *Usabilityengineerings* angesehen. Zufriedene Benutzer:innen sind motivierter, ein Produkt zu benutzen und akzeptieren dafür auch geringe Probleme in der Effektivität und Effizienz, wobei diese beiden Attribute als gegeben vorausgesetzt werden (Barnum, 2021, S. 13). Die Dominanz der Effektivität und Effizienz in den ursprünglichen Definitionen der Usability wird bereits seit längerem kritisiert, insbesondere die fehlende Betrachtung der User Experience und damit verbunden die emotionale Betrachtung bei der Nutzung von Produkten (Quesenbery, 2003). Aus diesem Grund entwarf Quesenbery (2003) basierend auf der Definition von Nielsen (1993, S. 26 - 27) eine Definition der Usability, bestehend aus den „5Es“: „Effektive“, „Efficient“, „Engaging“, „Error Tolerant“ und „Easy to Learn“ (Quesenbery, 2003, S. 2). In ihrer Definition wird demnach *Satisfaction* mit *Engaging* ersetzt. In der neuesten Fassung der ISO-Norm 9241-11 finden daher auch zunehmend die eben genannten Attribute Einzug. Die Autorin sieht zudem, abhängig vom Nutzungskontext, eine sich ändernde Relevanz der jeweiligen Usabilityfaktoren. In ihren Beispielen besitzt eine Museumsseite andere Schwerpunkte, zum Beispiel sollte sie sehr ansprechend (*engaging*) sein, als ein Registrierungsformular, das in erster Linie keine Fehler aufweisen (*error tolerant*) beziehungsweise mit fehlerhaften Nutzer:inneneingaben umgehen können sollte (Quesenbery, 2003).

### **Pädagogische Usability**

Sowohl Hardware- als auch Softwaresysteme können im Sinne der technischen Usability bewertet werden. Ein sich noch entwickelndes Forschungsfeld beschäftigt sich mit der so genannten *Pedagogical Usability*. Dabei steht die Beachtung von Attributen im Vordergrund, die unmittelbar den Lernprozess in einem digitalen Umfeld unterstützen sollen. Demnach kann sie sich sowohl auf das Lernmaterial selbst als auch auf den gesamten Lernkontext inklusive der verwendeten Hard- und Software beziehen. Die oberste Priorität ist das Erreichen des Lernziels in einem spezifischen (Schul-)Kontext (Nokelainen, 2004). Damit steht die eigentliche technische Usability teilweise in Konkurrenz mit der pädagogischen Usability. Während erstere beispielsweise eine möglichst geringe Gedächtnisbelastung anstrebt, wünscht sich letztere eben eine solche hohe Gedächtnisbelastung, um das Lernen anzureichern (Nokelainen, 2004).

Die pädagogische Benutzbarkeit einer Anwendung ist konzeptionell der *Utility* (vgl. Abbildung 8) zuzuordnen (Nokelainen, 2006). Damit gehört die pädagogische Usability im

engen Sinne in der Softwareentwicklung nicht zur gebräuchlichen Usability. Eine umfangreichere Analyse von Nokelainen (2006) identifiziert zehn Merkmale der pädagogischen Usability:

- *Learner Control*: Lernende müssen ihren Lernprozess anpassen können, um den Vorgaben der Lehrkraft gerecht werden zu können.
- *Learner Activity*: Der Lernkontext sollte zur aktiven Auseinandersetzung anregen und eine gewisse Relevanz für die oder den Lernenden besitzen.
- *Cooperative Learning*: Die Lernsituation sollte kooperatives und kollaboratives Lernen ermöglichen, womit auch gewisse Feedbackmechanismen, zum Beispiel aktueller Lernstand, ermöglicht werden.
- *Goal Orientation*: Lernen sollte zielorientiert und die Lernziele sollte den Nutzer\*innen klar sein, wobei eine gewisse Wahlmöglichkeit gegeben sein sollte.
- *Applicability*: Der Lernprozess muss Fähigkeiten und Wissen fördern, das im späteren Leben Anwendung findet. Dadurch wird wiederum eine gewisse Relevanz des Lerngegenstandes geschaffen.
- *Added Value*: Computereinsatz und die Verwendung von digitalen Medien sollten einen Zugewinn darstellen, wie es zum Beispiel bei kreativen Aufgaben mit mehreren Medien der Fall ist.
- *Motivation*: Der Lernkontext sollte motivierend für die Lernenden sein, indem sowohl die intrinsische als auch die extrinsische Motivation gefördert werden.
- *Valuation of Previous Knowledge*: Die Anwendung von bereits bekannten Inhalten in Kombination mit neu zu erwerbenden Fähigkeiten führt zu einer besseren Auseinandersetzung mit dem Lerngegenstand.
- *Flexibility*: Die Lernumgebung sollte frei erkundbar sein und ermöglichen, auf die Bedürfnisse der Nutzer:innen einzugehen.
- *Feedback*: Der Lernkontext sollte den Schüler:innen sowohl ermutigendes Feedback als auch unmittelbares Feedback zu den bearbeiteten Aufgaben geben.

Auch wenn die oben dargestellten Attribute der pädagogischen Usability bereits erfolgreich untersucht und evaluiert wurden (Nokelainen, 2004; Nokelainen, 2006) sowie das Konstrukt in Studien eingesetzt wurde (Djalev & Bogdanov, 2019), so zeigen sich dennoch einige Probleme der Betrachtung einer solchen Usabilitycharakterisierung. Die Kriterien sind stark kultur- und damit bildungssystemabhängig (Nokelainen, 2006). In manchen Ländern erfolgt beispielsweise der Unterricht, unter Einsatz von Gruppenarbeit, offener, während in anderen Ländern geschlossener Unterrichtsformen bevorzugt werden.

Ein weiterer Kritikpunkt liegt in der Überschneidung der technischen und pädagogischen Usability, die in Teilen auf die gleichen zugrundeliegenden latenten Variablen (Usefulness und Utility) zurückzuführen sind. Als Beispiel seien hier Feedbackmechanismen genannt, die sowohl auf Ebene der Software (Feedback über falsche Benutzer:inneneingabe) als auch auf Ebene des Lernmaterials (Feedback über Aufgabenlösung) bestehen können (Nokelainen, 2006).

### **2.2.2 User Experience in Abgrenzung zur Usability**

Das Forschungsfeld der User Experience (UX) gewinnt in jüngerer Zeit immer mehr an Bedeutung (Lewis & Sauro, 2021, S. 973 - 974.; Tullis & Albert, 2013, S. 6). Während es sich bei der Usability um ein funktionsbezogenes Merkmal handelt, spielen in der UX insbesondere Emotionen und die wahrgenommene Ästhetik einer Anwendung eine große Rolle (Richter & Flückiger, 2016, S. 12; Tullis & Albert, 2013, S. 5). Richter und Flückiger (2016, S. 13) übersetzen die User Experience mit dem Begriff der „Nutzungsqualität“. Die User Experience wird ebenfalls in einer ISO-Norm (ISO 9241-210) spezifiziert und mit „Benutzererlebnis“ (DIN, 2019, S. 11) übersetzt. Sie legt den Fokus auf die „Wahrnehmung und Reaktionen einer Person“ (DIN, 2019, S. 11) vor oder während der Nutzung eines Produktes. Damit wird die besondere Rolle der UX ersichtlich. Ein Produkt kann demnach, im Gegensatz zur Usability, vor der eigentlichen Benutzung eine gewisse anziehende oder abweisende Wirkung auf eine Person auswirken. Nach früheren Auffassungen kann die Usability auch als Teil der User Experience aufgefasst werden, womit ähnliche Methoden der User-Experience-Messung bestehen (Bevan, 2009). Die Methoden der UX-Testung gleichen im Wesentlichen denen der Usabilitytestung. Teilweise werden andere Attribute getestet und demnach andere Fragebögen eingesetzt.

Die User Experience stellt den Menschen in den Mittelpunkt der Betrachtung und konzentriert sich auf die positiven Gefühle, beispielsweise Spaß, einer Anwendungsnutzung (Hassenzahl & Tractinsky, 2006). Einige Autor:innen erachten die Usability als einen notwendigen „Hygienefaktor“ (übersetzt nach Hellman & Rönkkö, 2008, S. 2), der von Nutzer:innen als vorausgesetzt erachtet wird. Effizienz und Effektivität sind demnach wichtig, da sie unmittelbar die Zufriedenheit, das Vertrauen in das Produkt sowie die wahrgenommene Ästhetik und mittelbar die Verbreitung beziehungsweise Nutzung beeinflussen (Lewis & Sauro, 2021, S. 973 - 974). Eine Anwendung besitzt eine ergonomische Qualität (technische Usability) und eine hedonistische Qualität (Merkmale sind nicht aufgabenorientiert, zum Beispiel Originalität, Innovation, Exklusivität), die heute

die Definition der User Experience bestimmen und sich auf die Attraktivität eines Produktes auswirken (Hassenzahl et al., 2000; Hassenzahl 2001). In Ergänzung zu den oben genannten hedonistischen Merkmalen wurde von Hassenzahl (2004) auch die Schönheit beziehungsweise die Ästhetik eines Produktes als eigenständiges Merkmal zur User Experience ergänzt, das ebenfalls Einfluss auf die Produktzufriedenheit nimmt. Dabei unterscheidet sich die wahrgenommene Ästhetik eines Produktes und damit die Auswirkung auf die Usability zwischen den Kulturen (Tractinsky, 1997). Insgesamt scheint ein starker Zusammenhang zwischen der Ästhetik und der Benutzbarkeit eines Produktes zu bestehen (Tractinsky et al., 2000). Daraus ergibt sich die Forderung, sowohl ästhetische als auch funktionelle Betrachtungen bei der Produktgestaltung zu berücksichtigen (Kurosu & Kashimura, 1995).

### **2.2.3 Usabilitydefinition für die vorliegende Arbeit**

Wie aus der theoretischen Herleitung des Usabilitybegriffs hervorgeht, bestehen verschiedene Definitionen der Benutzbarkeit, die teilweise verschiedene Schwerpunkte setzen. Im ursprünglichen Sinne kann sie als Benutzerfreundlichkeit definiert werden (Nielsen, 1993, S. 23). Mit der Zeit wurde jedoch der Begriff weiter ausgeschärft, um verschiedene Aspekte eines Benutzer:innenerlebnisses abbilden zu können. Zudem erforderte die Messung der Usability über verschiedene Softwaresysteme hinweg einheitliche Kriterien, die heutzutage in der ISO-Norm 9241-11 definiert sind (ISO, 2018). Daneben entwickelte sich ein weiteres Forschungsfeld, das wieder mehr die Nutzer:innen und ihre Emotionen in den Vordergrund stellte, die User Experience (Richter & Flückiger, 2016, S. 12; Tullis & Albert, 2013, S. 5). Zusätzlich eröffneten sich weitere Definitionsfelder, die, je nach Kontext der Nutzung, andere Schwerpunkte wählen. Für den Schulkontext wurde beispielsweise die pädagogische Usability definiert, die einen weniger starken Fokus auf die technische Bedienung legt, sondern stattdessen den Lernprozess in das Zentrum rückt (Nokelainen, 2006).

Um die Usability einer Anwendung evaluieren zu können, muss dementsprechend zunächst aus den oben genannten Darstellungen eine geeignete Definition der Benutzbarkeit herangezogen werden. Bei HyperDocSystems handelt es sich um eine neuentwickelte Software, deren technische Usability in einem größeren Umfang noch nicht erprobt wurde. Erst wenn ein Programm technisch reliabel ist und die Aufgaben für den jeweiligen Anwendungszweck geeignet erscheinen, können weitere Aspekte, wie beispielsweise die User Experience, evaluiert werden. Ähnliches gilt für die pädagogische Usability,

zumal dort eine Trennung zwischen der Anwendung selbst und der Darbietung des Lernmaterials schwierig erscheint.

Die verschiedenen technischen Usabilitydefinitionen weisen viele Gemeinsamkeiten und einige wenige Unterschiede auf, die sich vor allem durch einen schwächeren oder stärkeren Fokus auf die emotionale Komponente der Nutzer:innen auszeichnen. Die Gewichtung der Faktoren innerhalb der jeweiligen Definitionen kann ebenfalls variieren. Dabei sollte jedoch keine allgemeingültige Regel formuliert, sondern individuell nach Produkt unterschieden werden. Für die vorliegende Arbeit wird die Definition der Usability nach der ISO-Norm 9241-11 (2018) verwendet, da es sich dabei um eine international gültige Definition der Usability handelt, die über viele Produkte hinweg ihre Gültigkeit besitzt. Im Bereich der Usability handelt es sich bei den drei Attributen Effektivität, Effizienz und Zufriedenheit um die am häufigsten untersuchten Faktoren bei Produkten im E-Learningbereich (Abuhlfaia & Quincey, 2018). Im Sinne der Gewichtung werden alle Faktoren als gleichermaßen wichtig erachtet. Das primäre Ziel von HyperDocSystems sollte aus Sicht der Schüler:innen und Lehrkräfte die Nutzung als Lernwerkzeug beziehungsweise Methodenwerkzeug darstellen. Dabei sollten die Lernenden das Arbeitsblatt effektiv bearbeiten können, sodass eine vollständige lernwirksame Bearbeitung mit möglichst wenigen Fehlern erreicht werden kann. Im Gegensatz zu analogen Arbeitsblättern bieten HyperDocs zum Beispiel den Mehrwert multimedialer Präsentationsform. Gleichzeitig sollten Nutzer:innen möglichst wenig Zeit und mentale Ressourcen auf beispielsweise die Navigation oder andere technisch bedingte Interfaceaktionen verwenden, sodass eine hohe Effizienz des Systems vorliegt. Sind die oberen beiden Bedingungen erfüllt, erfahren die Schüler:innen auch eine hohe Zufriedenheit bei der Nutzung der Anwendung. Die zufriedenstellende Nutzung führt dazu, dass Lernende HyperDocSystems auch in der Zukunft nutzen wollen. Um das individuelle Nutzungserlebnis allerdings feststellen zu können, ist eine Erfassung der Usability notwendig. Dazu existieren verschiedene Methoden.

### **2.2.4 Testung von Usability**

Die Usability eines Systems lässt sich über verschiedene Methoden testen (Holzinger, 2015). Nielsen (1993, S. 175, S. 185) betont die Spezifität der Usabilitymessungen entsprechend der Nutzungsgruppe und des Aufgabenbereichs. Prinzipiell lässt sich in formatives und summatives Testen unterscheiden (Nielsen, 1993, S. 170). Das formative Testen findet während der Produktentwicklung statt und basiert meistens auf mehreren kleineren Messungen über die gesamte Entwicklungszeit (Moser, 2012, S. 224). Das



summative Testen findet mit dem fertigen oder fast fertigen Produkt statt. Die Testung erfolgt mit einer größeren Stichprobe, um eine statistisch repräsentative Aussage zu erhalten (Barnum, 2021, S. 15). Sie dient damit zur Kontrolle der Produktqualität (Richter & Flückiger, 2016, S. 104). Bei Usabilitytestungen können bereits vier bis fünf Nutzer:innen ausreichen, um bis zu achtzig Prozent der Usabilityprobleme aufzufinden (Lewis, 1993; Virzi, 1992). Das beste Kosten-Nutzen-Verhältnis besteht laut Nielsen (1993, S. 173) bereits bei drei Benutzer:innen. Zu einem anderen Schluss kommen hingegen Mochlich (2018) sowie Spool und Schroeder (2001), die mit einer zunehmenden Zahl an Probanden auch zunehmend mehr Fehler fanden und für größere Stichproben plädieren. Die widersprüchlichen Erkenntnisse und Empfehlungen werden teilweise von Barnum et al. (2003) relativiert und in Bezug zu dem jeweiligen Kontext gesetzt, sodass keine allgemeingültigen Aussagen getroffen werden können. Tendenziell hängt die Stichprobengröße vom eingesetzten System und Anwendungszweck ab. Umfangreiche komplexe Programme und große Webseiten erfordern mehr Tester:innen. Das gleiche gilt für sehr relevante oder gar lebenswichtige Programme, die eine hohe Zuverlässigkeit erfordern (Barnum, 2021, S. 18; Richter & Flückiger, 2016, S. 112). Je nach Testmethode (s. unten) empfiehlt Holzinger (2005) zwischen drei und mehr als dreißig Proband:innen. Richter und Flückiger (2016, S. 111) sehen für qualitative Aussagen zwischen vier und mehr als fünfzehn Tester:innen vor, für quantitative Aussagen empfehlen sie mehr als zehn Personen.

Usabilitytests finden für gewöhnlich in einem speziellen Labor oder in der eigentlichen Anwendungsumgebung statt (Richter & Flückiger, 2016, 109 - 110). Während damit Studien unter sehr kontrollierten Bedingungen stattfinden können, ergeben sich einige Nachteile. Zunächst ist die Einrichtung eines Usabilitylabors teuer. Es muss ein spezieller Raum dauerhaft zur Verfügung gestellt werden. Weiterhin ergeben sich Nachteile, die unmittelbar die Messung der Usability beeinflussen können und für viele Laboruntersuchen gelten. Die Umgebung im Labor entspricht einer künstlichen Umgebung. Damit werden zwar experimentelle Gegebenheiten geschaffen, allerdings verhält sich der oder die Nutzer:in im natürlichen Umfeld anders. Ein Grund hierfür könnte die Nervosität sein, die Tester:innen in einer künstlichen Testumgebung entwickeln könnten (Barnum, 2021, S. 83). Auf der anderen Seite bieten Feldmessungen einige Vorteile. Beispielsweise können natürliche Einflüsse auf die Usability, wie Lichteinfall oder Lärm und damit verbundenen Konzentrationsschwankungen, berücksichtigt werden. Ein weiterer Vorteil liegt darin, dass sich die Testperson in ihrem natürlichen Umfeld befindet (Barnum, 2021, S. 84).

Gleichzeitig gehen auch mit der Feldstudie einige Nachteile einher, die wiederum als Vorteile der kontrollierten Testung genannt wurden. Der größte Nachteil besteht in der fehlenden Kontrolle der Testumgebung, sodass Interferenzen mit dem Usabilitytest auftreten können (Barnum, 2021, S. 85).

Holzinger (2015) unterscheidet in Usabilityinspektions- und Testmethoden (vgl. Abbildung 6). Bei Moser (2012) lässt sich äquivalent in Expert:innentests und Nutzer:innentests unterscheiden. Er ergänzt weitere Verfahren, wie zum Beispiel den A/B-Test.

	Inspection Methods			Test Methods		
	Heuristic Evaluation	Cognitive Walkthrough	Action Analysis	Thinking Aloud	Field Observation	Questionnaires
Applicably in Phase	all	all	design	design	final testing	all
Required Time	low	medium	high	high	medium	low
Needed Users	none	none	none	3+	20+	30+
Required Evaluators	3+	3+	1 - 2	1	1+	1
Required Equipment	low	low	low	high	medium	low
Required Expertise	medium	high	high	medium	high	low
Intrusive	no	no	no	yes	yes	no

Abbildung 6: Usabilityinspektions- und Testmethoden (eigene Darstellung nach Holzinger, 2015).

Zur Inspektion der Benutzbarkeit einer Anwendung (Expert:innentests) wird im Wesentlichen zwischen drei Verfahren unterschieden: Heuristische Evaluation, kognitiver *Walkthrough* und Handlungsanalyse (Holzinger, 2005). Diese Verfahren dienen in erster Linie zur Identifizierung von Usabilityproblemen und zur Verbesserung des Produktinterfaces. Sie werden von speziellen Usabilityexpert:innen durchgeführt. Bei der heuristischen Evaluation prüft die Person die Benutzeroberfläche auf spezielle Heuristiken, wie zum Beispiel solche von Nielsen (1993, S. 115 - 155). Der kognitive Walkthrough ist in erster Linie aufgabenorientiert, indem schrittweise das Verhalten des Systems bei Durchführung der Aufgabe geprüft wird. Dabei wird insbesondere auf die Erlernbarkeit der Anwendung geachtet. Die Handlungsanalyse umfasst eine sehr genaue Auswertung der einzelnen Schritte einer Aktionssequenz, wie zum Beispiel die gedrückten Tasten der Tastatur oder die benutzten Zeigerbewegungen. Dabei wird die benötigte Zeit gemessen, um die Handlung auszuführen (Holzinger, 2005). Die eigentliche Messung der Usability erfolgt durch die Endnutzer:innen ebenfalls über drei Methoden (Nutzer:innentests):

*Thinking Aloud*, Feldbeobachtung und Fragebögen. Bei der Thinking-Aloud-Technik müssen die Tester:innen ihre Gedanken und Handlungen konsequent verbalisieren. Dadurch können Fehlvorstellungen und spezielle Sichtweisen auf das System aufgedeckt werden. Gleichzeitig werden die Gründe dafür sofort erkennbar. Dabei reicht schon eine geringe Anzahl an Personen aus, um ein Produkt zu evaluieren. Die Methode kann sich jedoch unnatürlich für Nutzer:innen anfühlen und es handelt sich um eine sehr zeitintensive Methode, die nur eine geringe Anzahl an Datensätzen produziert (Barnum, 2021, S. 402; Holzinger, 2005).

Eine sehr einfache Methode der Usabilitymessung ist die Feldbeobachtung, bei der Anwender:innen in ihrem natürlichen Arbeitsfeld beobachtet werden. Dabei muss jedoch eine interaktionsfreie Hospitation zwischen Beobachter:in und Nutzer:in stattfinden. Das Ziel dieser Methode liegt hauptsächlich in der Identifikation größerer und kritischer Probleme beziehungsweise Fehlfunktionen. Die Beobachtung kann auch per Videografie erfolgen. Der Vor- beziehungsweise Nachteil dieser Methode liegt, je nach Betrachtungsweise, in dem natürlichen Umfeld der Benutzer:innen, das einen Einfluss auf die Produktinteraktion nimmt (Barnum, 2021, S. 84 - 85; Holzinger, 2005; Sonderegger et al., 2021). Generell sind Erhebungen im Feld den Laborbeobachtungen vorzuziehen, da mehr Usabilityprobleme aufgedeckt werden können (Nielsen et al., 2006).

Fragebögen sind eine weitere häufig verwendete Methode der Usabilitymessung (Zaharias & Poulymenakou, 2009), insbesondere im Bereich des E-Learnings (Abuhlfaia & Quincey, 2018) oder *Educational Gaming* (Gao et al., 2019). Bei der Usabilitymessung durch Endnutzer:innen handelt es sich nach Holzinger (2015) um eine unverzichtbare Methode, unter anderem auch, da sie kostengünstig und einfach anzuwenden ist (Root & Draper, 1983). Diese Aussage ist nachvollziehbar, weil die Nutzung des Produktes durch die Anwender:innen erfolgt und dort entsprechend den Usabilityattributen für Zufriedenheit, Effektivität und Effizienz in der Bearbeitung der jeweiligen Aufgaben sorgt. Es handelt sich um eine Vorgehensweise, die zum einen große Stichproben erfordert, aber auch, im Gegensatz zu zum Beispiel der Think-Aloud-Technik, umfangreiche Erhebungen ermöglicht (Richter & Flückiger, 2016, S. 113). Dabei können unerfahrene Nutzer:innen eine Vielzahl der Usabilityprobleme erkennen, die auch Expert:innen finden würden (Ssemugabi & De Villiers, 2007). Die Methode eignet sich insbesondere zur Erfassung der Zufriedenheit (Holzinger, 2005). Es existiert eine Reihe von standardisierten validierten Fragebögen (Barnum, 2021, S. 232), wie zum Beispiel *SUS* (*System Usability Scale*;

Brooke, 1995), *CSUQ* (*Computer System Usability Questionnaire*; Lewis, 1995), *UEQ* (*Usability Evaluation Questionnaire*; Laugwitz et al., 2008) oder *USE* (*Usefulness, Satisfaction and Ease of Use*; Lund, 2001). Standardisierte Fragebögen sollten nichtstandardisierten vorgezogen werden (Lewis, 2014; Moser, 2012, S. 236). Bei den Befragten sollte es sich um eine repräsentative Stichprobe handeln, die das Produkt später auch nutzt (Moser, 2012, S. 236).

## **2.3 Heterogenität und Differenzierung**

Neben der Motivation und der Usability bildet der Einsatz von Lernhilfen den dritten Schwerpunkt dieser Arbeit. Lernhilfen werden oftmals in heterogenen Lerngruppen eingesetzt. Die Verschiedenheit der Schüler:innen stellt Lehrkräfte bei der täglichen Unterrichtsgestaltung vor Herausforderungen. Binnendifferenzierung durch Lernhilfen ist daher eine Möglichkeit der Begegnung mit Heterogenität. Im folgenden Kapitel soll daher eine Begriffsdefinition der Heterogenität und Differenzierung im Schulsystem für diese Arbeit formuliert werden.

### **2.3.1 Der Heterogenitätsbegriff**

Das Adjektiv „heterogen“ leitet sich von den beiden griechischen Wörtern „heteros“ (anders) und „genos“ (Art, Geschlecht) ab und meint „von verschiedener Abstammung, Art, Gattung“ (Trautmann & Wischer, 2011, S. 38). Das gegenteilige Adjektiv dazu lautet „homogen“. Heterogenität liegt nicht als isoliertes Konstrukt vor, sondern sie ergibt sich durch einen Vergleich von Mitgliedern einer Gruppe. Die Verschiedenheit wird durch einen Beobachter oder eine Beobachterin hergestellt. Zudem handelt es sich um eine temporäre Zuschreibung, die sich mit der Zeit oder ändernden Beurteilungskriterien wieder auflösen kann. Der Begriff Heterogenität wird in schulischen Kontexten als Vielfalt oder Verschiedenheit übersetzt. Es handelt sich dabei um kein neues Phänomen, auch wenn das Thema in den letzten Jahren eine immer stärkere Aufmerksamkeit gewinnt. Grund dafür sind unter anderem die Ergebnisse der PISA- und TIMMS-Studien (Trautmann & Wischer, 2011, S. 18; Wenning, 2007).

Im Mittelpunkt der Betrachtung von Heterogenität steht die Verschiedenheit der Lernenden in Bezug auf unterschiedliche Merkmale, wie zum Beispiel die Intelligenz, Interessen oder das Alter. Im Zusammenhang mit dem Heterogenitätsbegriff werden diese Merkmale auch als Heterogenitätsdimensionen zusammengefasst. Als weitere Beispiele sind

die soziale Herkunft und die kognitive Leistungsfähigkeit zu nennen (Trautmann & Wischer, 2011, S. 40). Es existiert noch eine Vielzahl weiterer Unterscheidungen und Einteilungen. Prinzipiell bestehen zwei verschiedene Sichtweisen auf das Thema Heterogenität und Unterrichtsgestaltung: Eine psychologische Sicht im Sinne der Lehr-Lern-Forschung und eine sozialpädagogische Sichtweise (Trautmann & Wischer, 2011, S. 42). Erstere beschäftigt sich mit der Frage, wie Unterschiede in Schulleistungen zwischen Individuen erklärt werden können. Das Angebots-Nutzungsmodell nach Helmke (2004, S. 42) spielt in diesem Zusammenhang eine zentrale Rolle (Trautmann & Wischer, 2011, S. 42 - 43). Kognitive Lernziele stehen bei dieser Betrachtung im Vordergrund. Zudem sind in diesem Modell neben den Merkmalen der Lernenden auch die des Unterrichts zu berücksichtigen. In Bezug auf Schüler:innen sind unter anderem folgende Merkmale relevant: das Vorwissen, die Motivation, die Sprache oder auch die Intelligenz. Die eben genannten Merkmale lassen sich zum Teil im Unterricht gut beobachten und in den Lehr-Lern-Prozessen berücksichtigen. Die Lehrkraft kann zum Beispiel aktiven Einfluss auf das Vorwissen der Individuen nehmen (Trautmann & Wischer, 2011, S. 44). Die eben beschriebene Sichtweise reduziert eine vorliegende Heterogenität auf die Lehr-Lern-Situation. Gleichwohl spielen jedoch soziale Bedingungen ebenso eine wichtige Rolle (Trautmann & Wischer, 2011, S. 47 - 48). Das fachliche Lernen steht dabei eher im Hintergrund. Vorrangig steht die Schule als Institution der Gesellschaft im Vordergrund. Damit liegt die strukturelle Heterogenität aufgrund verschiedener gesellschaftlicher Probleme im Fokus. Die Verschiedenheit zwischen Schüler:innen ist daher als soziales Konstrukt zu interpretieren, das durch eine äußere Beurteilung oder Rahmung entsteht. Ein Beispiel hierfür ist der negativ konnotierte Begriff der „Lernbehinderung“, der im Laufe der Zeit stigmatisiert wurde (Trautmann & Wischer, 2011, S. 48). Unter die sozialpädagogische Betrachtung fallen noch weitere Aspekte, wie beispielsweise das Geschlecht oder die Kulturzugehörigkeit (Trautmann & Wischer, 2011, S. 48). Der Heterogenität wird im Bildungssystem in verschiedener Weise begegnet. Oftmals versucht das Bildungssystem Heterogenität als Chance zu begreifen und im Unterricht im Rahmen der inneren Differenzierung zu nutzen (Wenning, 2007).

Für die vorliegende Arbeit wird ein Heterogenitätsverständnis im Rahmen der Lehr-Lern-Forschung angenommen, bei der vor allem die Lehrkraft die Lernsituation vorkonstruiert und die Leistungsheterogenität im Vordergrund steht.

### 2.3.2 Äußere und innere Differenzierung

Heterogenität stellt gerade unter der Betrachtung der lehr-lern-theoretischen Perspektive Lehrkräfte vor Herausforderungen. In stark heterogenen Klassen können ohne zusätzliche Interventionen nicht alle Schüler:innen gleichermaßen gefördert beziehungsweise gefordert werden - die Heterogenität wird ignoriert. Das deutsche Bildungssystem sieht vor, heterogene Lerngruppen entsprechend ihren Leistungen nach der Grundschule in die verschiedenen Formen weiterführender Schule zu verteilen. Ziel ist es, die Lerngruppen zu homogenisieren. Weiterhin wird in der Praxis innerhalb einer Schulform in verschiedene Klassenstufen unterteilt (Altrichter et al., 2009). Diese so genannte äußere Differenzierung wird zum Teil stark kritisiert (Trautmann & Wischer, 2011, S. 77 - 78; Wenning, 2007). Das liegt unter anderem daran, dass sich Kinder mit ähnlichen Lernvoraussetzungen nach der Grundschule, je nach besuchter weiterführender Schule, unterschiedlich entwickeln. Schüler:innen, die das Gymnasium besuchen, erlangen bessere Fähigkeiten als Jugendliche anderer Schulformen. Es liegt ein so genannter „Institutionseffekt“ (Baumert et al., 2009, S. 37) vor, der auf den speziellen Merkmalen, zum Beispiel Unterrichtskultur, der jeweiligen Schulform beruht (Baumert et al., 2009). Gleichzeitig werden die Schüler:innen indirekt nach sozialer Herkunft separiert, wodurch nach Schulformen ausgerichtete Zusammensetzungen der Klassen entstehen. Dieser Effekt, bei dem sich die Komposition der Lerngruppe auf die Lernkultur und die Interaktion zwischen Lehrkräften und Lernenden sowie innerhalb der Lernenden auswirkt, wird als „Kompositionseffekt“ (Baumert et al., 2009, S. 37) bezeichnet. Die verschiedenen Entwicklungspotentiale von Jugendlichen, bedingt durch die besuchte Schulform, hängen jedoch weniger von der Zusammensetzung der Lerngruppe ab, als vielmehr von dem bereits beschriebenen Institutionseffekt (Baumert et al., 2009). Dennoch sind auch immer die individuellen Voraussetzungen zu berücksichtigen. So kann es sich negativ auf das Selbstkonzept auswirken, wenn ein:e leistungsstarke:r Lernende:r der Grundschule, der sich in einer leistungsschwachen Lerngruppe befunden hat, auf dem Gymnasium in eine leistungsstarke Lerngruppe wechselt. Dieses Phänomen ist unter dem sogenannten „Big-Fish-Little-Pond-Effekt“ (Vock & Gronostaj, 2017, S. 51) bekannt. Der Effekt wurde in einer Metaanalyse von Fang und Kolleg:innen (2018) erneut bestätigt. Sie konnten den Effekt für alle Altersstufen und Kulturkreise bestätigen. Bei Schüler:innen weiterführender Schulen ist der Effekt am größten (Fang et al., 2018).

Die eben beschriebenen Effekte sind auf die äußere Differenzierung zurückzuführen. Wird sie außer Acht gelassen, werden Methoden benötigt, um Individuen in heterogenen Lerngruppen gleichermaßen zu fördern. Die sogenannte innere Differenzierung, auch Binnendifferenzierung genannt, strebt einen produktiven Umgang mit Heterogenität an. Ein Beispiel hierfür ist das Land Rheinland-Pfalz, das in der Mittelstufe einen gemeinsamen Lehrplan für alle Schulformen herausgibt. Dort wird Heterogenität explizit als bereichernd dargestellt. Zudem sind im Orientierungsrahmen Schulqualität die Verschiedenheit und die Differenzierung als ein „Merkmal der Unterrichtsqualität“ (MBWWK, 2014, S. 12) genannt. Die Unterschiede sollen durch entsprechende Anpassungen der Lehrkraft beachtet werden (MBWWK, 2014, S. 12).

Für diese Binnendifferenzierung bestehen verschiedene methodische und didaktische Konzepte der Umsetzung. Das Land Rheinland-Pfalz nennt hierfür die Auswahl der Inhalte und Lernprodukte, „gestufte Aufgabenstellungen und andere Hilfsangebote“ (MBWWK, 2014, S. 9). Zusätzlich wird zur inneren Differenzierung noch in Individualisierung oder individuelles Lernen differenziert, wobei Schüler:innen nach ihren individuellen Voraussetzungen gefördert werden. Methoden der Individualisierung sind zum Teil offene Unterrichtsformen, wie beispielsweise Projektunterricht oder Freiarbeit (Alt Richter et al., 2009). Prinzipiell kann bei der inneren Differenzierung zwischen Methoden unterschieden werden, bei denen die Lernenden entsprechend Anpassungen vornehmen und die Lehrkraft eher als Begleiter fungiert oder bei der die Lehrer:innen an der Lehr-Lern-Forschung ausgerichtete Anpassungen des Lernmaterials vornehmen (Breidenstein et al., 2013). Bei allen Konzepten der inneren Differenzierung spielen jedoch die Voraussetzungen der Lernenden in Bezug auf die Selbststeuerung und Selbstorganisation eine wichtige Rolle. Aber auch Lehrkräfte müssen erfahren sein, um passende Differenzierungsmaßnahmen gestalten (Trautmann & Wischer, 2011, S. 114; Vock & Gronostaj, 2017, S. 80 - 81) oder generell ein lernförderliches Klima in heterogenen Lerngruppen schaffen zu können (Decristan et al., 2017). Gelingt dies durch die Lehrkraft nicht, entstehen in stark heterogenen Lerngruppen teilweise negative Auswirkungen auf das Lernen (Evertson et al., 1981), wobei in anderen Studien (z. B. Decristan et al., 2017) keine negativen Folgen von Heterogenität berichtet wurden. De Giorgi et al. (2012) konnten dieses Phänomen teilweise aufklären: Ein gewisser Grad an Verschiedenheit wirkt lernförderlich, während ein großes Maß an Heterogenität eher negative Folgen trägt.

Der Effekt von Individualisierung (Hattie, 2009, S. 162) und bestimmten Arten von Binnendifferenzierung, beispielsweise Kleingruppenbildung (Lou et al., 1996), auf die Lernleistung ist teilweise gering. Das liegt unter anderem an den vielfältigen Umsetzungsmöglichkeiten und der Qualität der Umsetzung. Oftmals werden die Lernprozesse durch Unterstützungsmaßnahmen nicht angeregt, sondern konkret eine Lösung vorgegeben (Vock & Gronostaj, 2017, S. 69). Eine Übersicht zum Umgang mit Heterogenität auf Ebene der Klassenorganisation bietet Abbildung 7. Die Autoren (Lipowsky & Lotz, 2015) führen dabei beispielsweise offene Unterrichtsformen und adaptives Unterrichten als eigene Konzepte auf.

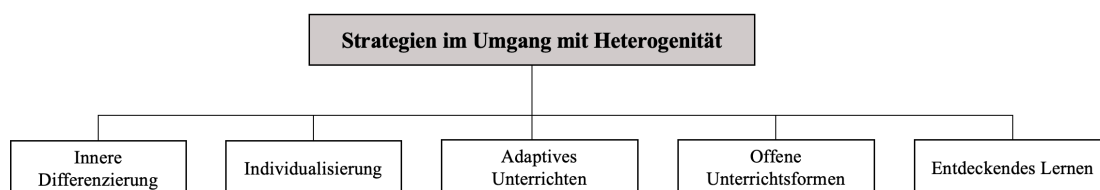


Abbildung 7: Übersicht über Strategien im Umgang mit Heterogenität (eigene Darstellung nach Lipowsky & Lotz, 2015).

Prinzipiell kann Binnendifferenzierung auf drei verschiedenen Ebenen der Unterrichtsplanung bedacht werden. Dazu zählen die organisatorische sowie inhaltliche Ebene und der Grad der Unterstützung selbst. Bei der organisatorischen Differenzierung lernen die Schüler:innen in verschiedenen Sozialformen oder in unterschiedlichem Lerntempo. Bei der inhaltlichen Differenzierung arbeiten die Lernenden zur gleichen Zeit an verschiedenen inhaltlichen Themen oder Schwerpunkten beziehungsweise Lerngegenständen. Bei der Variation der Unterstützung wird ein Lernangebot bereitgestellt, das verschiedene Lernstile und Lernvoraussetzungen berücksichtigt (Joller-Graf, 2010).

Damit bestehen vielerlei Ansatzpunkte zur Binnendifferenzierung im Unterricht. Diese bereits erwähnte Vielfalt macht eine taxonomische Einordnung der Methoden schwierig. Eine Übersicht dazu bieten Pozas und Schneider (2019). In ihrer Untersuchung identifizierten sie insgesamt sechs verschiedene Kategorien von differenziertem Unterricht. Zudem berichten sie über die Wirksamkeit der Methoden. Die erste Kategorie bilden gestufte oder gestaffelte Aufgaben. Hierbei werden die Aufgaben und das dazugehörige Material auf verschiedenen Ebenen, beispielsweise in der Komplexität oder dem Bearbeitungsumfang, variiert. Das Ziel ist dabei, dass die Schüler:innen nach ihren individuellen Bedürfnissen lernen. Die zweite Kategorie bildet das gezielte Zusammensetzen von Schüler:innengruppen, um bewusst homogene oder heterogene Lerngruppen in Bezug auf



das Interesse oder die Lernleistung zu schaffen. Dadurch können in heterogenen Gruppen entweder die leistungsstärkeren Lernenden die leistungsschwächeren unterstützen oder die Lehrkraft unterstützt leistungsschwache homogene Lerngruppen, während die stärkeren Gruppen allein arbeiten. Die dritte Kategorie bilden Tutorensysteme innerhalb der Lerngruppe, bei denen bessere Lernende wiederum schwächere unterrichten. Beide Gruppen profitieren von dem erneuten Erklären der Unterrichtsinhalte. Eine weitere Gruppierung bilden gestufte nonverbale Lernhilfen. Sie sollen den Lernenden bei Lernschwierigkeiten unterstützen. Das sogenannte *Mastery Learning* bildet die fünfte Kategorie. Bei dieser Methode wird ein Mindeststandard festgelegt, den alle Schüler:innen erreichen sollen. Dies ist mit einer engen Kontrolle des Lernfortschritts verbunden. Die letzte Unterteilung ist der offene Unterricht, bei dem Schüler:innen eine große Autonomie über den eigenen Lernprozess gewährt wird. Beispiele hierfür sind das Stationenlernen oder das projektbasierte Lernen (Pozas & Schneider, 2019). Letzel und Otto (2019) untersuchten in einer qualitativen Studie die Nutzungshäufigkeit von Methoden zur Differenzierung des Unterrichts nach der Taxonomie von Pozas und Schneider (2019). Prinzipiell scheinen Lehrkräfte des Gymnasiums weniger häufig zu differenzieren als ihre Kolleg:innen der Gesamtschule. Als häufigste Methode der Unterstützung gaben die Lehrenden abgestufte Aufgaben und Materialien (38 %) an. Darauf folgen gestufte nonverbale Lernhilfen (23 %), Öffnung des Unterrichts (20 %), gezielte Zusammensetzung von Schülergruppen (16 %), Helfer- und Tutorensysteme (2 %) und schließlich das *Mastery Learning* mit einem Prozent (Letzel & Otto, 2019). Auch in den Standards für die Lehrer:innenbildung der Kultusministerkonferenz (KMK, 2019) werden für den Umgang mit Heterogenität explizit Kompetenzen formuliert. Diese beziehen sich sowohl auf die theoretische als auch auf die praktische Ausbildung, wonach die Relevanz für das Bildungssystem ersichtlich wird.

### **2.3.3 Binnendifferenzierung durch Lernhilfen**

Die Differenzierung durch (gestufte, nonverbale) Lernhilfen ist eine gängige Methode der Differenzierung. Dabei kann zwischen weiteren Unterkategorien, wie beispielsweise einfachen Hilfekärtchen oder visuellen Hilfen differenziert werden (Letzel & Otto, 2019). Hilfefkarten können in analoger oder digitaler Form bereitgestellt werden. Es handelt sich um eine flexible Methode der Unterstützung, da Lehrkräfte den Inhalt und die Darstellung der Hilfen leicht anpassen beziehungsweise abändern können. Beispielsweise können Bilder Lernende mit Leseschwierigkeiten unterstützen. Sie werden daher in allen

Klassenstufen eingesetzt (Conderman & Hedin, 2015). Hilfekarten können insbesondere Schüler:innen mit Lernschwierigkeiten unterstützen, indem sie die Metakognition beim Lernen erhöhen. Gleichzeitig wird das selbstständige Lernen gefördert (Conderman & Hedin, 2011). Neben positiven Effekten auf die Selbstwirksamkeit können Lernhilfen auch die kognitive Last beim Bearbeiten von Aufgaben verringern (Arnold et al., 2016).

Einige Konzepte der Differenzierung setzen auf verschiedene Arten der Wissenspräsentation beziehungsweise Modalität, wie zum Beispiel visuelle oder auditive Hilfen (Sanchez & Garcia-Rodicio, 2008). Besonders geeignet für diese Art der Differenzierung sind sogenannte *Computer-based Learning Environments*, in denen aufgrund der Technik problemlos mehrere Modalitäten angeboten werden können. Dabei haben verschiedene Modalitäten unterschiedliche Vor- und Nachteile, die zum Teil auch aufgabenspezifisch sind (Sanchez & Garcia-Rodicio, 2008). Die Art des Inhalts beziehungsweise die generelle Struktur der Lernhilfen kann variieren. Ein Beispiel hierfür sind unstrukturierte Lernhilfen, bei denen weniger strategische Hilfen als vielmehr inhaltliche Unterstützungen angeboten werden. Es handelt sich dabei um Karteikarten mit Texten oder Bildern. Auch bei dieser Methode ist die Selbststeuerung der Schüler:innen wichtig (Großmann & Woest, 2014; Großmann & Woest, 2015). Ein Beispiel für stark systematisierte und strukturierte Lernhilfen sind gestufte Lernhilfen. Dieses Konzept wurde ursprünglich von Leisen (1999) eingeführt und bereits vielfältig erfolgreich im Unterricht erprobt und erweitert (z. B. Arnold et al., 2016, Franke-Braun et al., 2008; Hänze et al., 2010, Stäudel et al., 2007). Die Schüler:innen erleben beim Lernen mit gestuften Lernhilfen Autonomie- und Kompetenzerleben, wodurch eine Verbesserung des Selbstkonzepts angenommen werden kann (Hänze et al., 2007). Bei gestuften Lernhilfen wird in der Regel in Partnerarbeit oder in Gruppenarbeit gelernt. Die Hilfen selbst dienen zur Aktivierung von Vorwissen oder zur Eröffnung der Lernstrategie. Es wird zwischen sechs Stufen von Hilfen unterschieden: „Paraphrasieren“, „Fokussieren“, „Elaboration von Unterzielen“, „Aktivierung von Vorwissen“, „Visualisierung“ und „Verifikation“ (Hänze et al., 2010, S. 67). Die Methode der gestuften Lernhilfen muss eingeübt werden. Gerade zu Beginn neigen die Schüler:innen dazu, sich zunächst die Lösung (letzte Hilfe) zu betrachten. Gleichzeitig werden Erfolge in Bezug auf selbstständiges Lernen erst nach mehrmaligem Einsatz sichtbar (Hänze et al., 2010). Die notwendigen Kompetenzen zum sogenannten Help-Seeking sind noch nicht ausreichend entwickelt.

### 2.3.4 Help-Seeking: Das Aufsuchen von Hilfe

Der Begriff des Help-Seeking umfasst die Beanspruchung einer Hilfe in verschiedenen Kontexten. Die einfachste Form des Help-Seeking beinhaltet das Fragen einer Mitschülerin oder eines Mitschülers, die Lehrkraft oder die Nutzung von Lernhilfen, wobei in der Literatur erstere als Quellen der Unterstützung deutlich häufiger erforscht und beschrieben sind. Es handelt sich damit um einen hochsozialen Prozess. Um den Vorgang des Hilfeersuchens im Sinne der sozialen Kosten, die insbesondere von heranwachsenden Schüler:innen gesehen werden, zu entlasten, können technologieunterstützte Maßnahmen eingesetzt werden. Dadurch kann die Barriere des Help-Seeking verringert werden. Zudem ermöglicht eine asynchrone Bereitstellung von Hilfe die Reflexion über die Unterstützungsmaßnahme (Karabenick & Berger, 2013). Dieser Prozess kann durch intelligente Tutoringsysteme gestützt werden (z. B. Alevan et al., 2016). Allerdings sind diese Anwendungen häufig auf ein bestimmtes Fach und Themengebiet beschränkt. Die Bedeutung der Fähigkeit der adäquaten Nutzung einer Unterstützung ist von zunehmender Bedeutung für Psycholog:innen und Lehrende (Nelson-Le Gall, 1985). In jüngerer Zeit wird das Help-Seeking als eine selbstregulierte Lernstrategie eingeordnet, indem sich ein Individuum bewusst dafür entscheidet, jemanden um Hilfe zu bitten (Karabenick & Berger, 2013). Der Prozess des Ersuchens einer Hilfe beinhaltet vier verschiedene Eigenschaften seitens der Hilfesuchenden. Das Wissen darüber, wann eine Hilfe benötigt wird und wie gegebenenfalls die Frage formuliert wird, erfordert kognitive Kompetenzen. Gleichzeitig sind soziale Kompetenzen erforderlich, um zu wissen, wer am ehesten bei dem Problem unterstützen kann und wie diese Person anzusprechen ist. Affektive-emotionale Einstellungen beeinflussen den gesamten Prozess des Help-Seeking. Beispielsweise wird bei Heranwachsenden dieser Prozess durch soziale Ziele innerhalb der Klasse beeinflusst (z. B. Ryan, 1997). Hinzu kommen kontextual-emotionale Glaubenssätze, die die Zusammenarbeit und die Verhaltensregeln beim Hilfeersuchen beeinflussen (Karabenick & Dembo, 2011).

Der Prozess des Help-Seeking umfasst insgesamt acht Schritte (Karabenick & Berger, 2013). Zunächst muss die oder der Lernende identifizieren, dass ein Problem vorliegt. Anschließend muss entschieden werden, ob eine Hilfe gewollt oder gebraucht wird. Damit verbunden ist schließlich die Entscheidung, generell eine Hilfe in Anspruch zu nehmen. Schließlich folgt die Überlegung über die Art oder das Ziel des Hilfeersuchens. In diesem Zusammenhang kann eine Hilfe geben werden, die den Lernprozess anregt und

unterstützend wirkt (*Instrumental Help-Seeking*) oder es kann direkt nach einer Lösung gefragt werden (*Executive Help-Seeking*). Zum Teil bestehen für diese beiden Arten des Help-Seeking andere Bezeichnungen. Eine dritte Form des Help-Seeking ist das Vermeiden von Hilfen (*Help-Seeking Avoidance*). Wird der Entschluss gefasst, eine Hilfe zu nutzen, muss die oder der Schüler:in entscheiden, wer um Hilfe gefragt wird und schließlich darum bitten. Nachdem die Hilfe erhalten wurde, erfolgt zuletzt die Verarbeitung der erhaltenen Information (Karabenick & Dembo, 2011). Die Art des Hilfeersuchens (instrumental / executive) ist stark abhängig von den Zielorientierungen der Schüler:innen. Während Lernende mit *Mastery*-Orientierung häufiger nach instrumenteller Hilfe fragen, nutzen Lernende mit Performanz-Orientierung häufiger das exekutive Help-Seeking. Einige der Studien zu diesen Zusammenhängen werden im späteren Kapitel des Forschungsstandes aufgeführt.

## 2.4 HyperDocs und HyperDocSystems

Die Binnendifferenzierung mit Lernhilfen bildet ein wichtiges Werkzeug, um heterogene Lerngruppen zu unterrichten. Mit HyperDocSystems lassen sich binnendifferenzierte digitale Arbeitsblätter, sogenannte HyperDocs, erstellen. Im nachfolgenden Kapitel wird daher der Begriff der HyperDocs eingeführt sowie verschiedene Vor- und Nachteile der HyperDocs und Einsatzzwecke im Unterricht näher beleuchtet. Es folgt anschließend eine Auseinandersetzung mit der dahinterstehenden Software HyperDocSystems, die die Administration und Distribution der HyperDocs ermöglicht. Dabei wird der Entwicklungsstand zum Zeitpunkt der im Rahmen dieser wissenschaftlichen Arbeit durchgeführten Studie beschrieben.

### 2.4.1 Der Begriff HyperDoc

Das Wort HyperDocs weist zwei verschiedene Begriffsdefinitionen auf. Zum einen existiert im angelsächsischen Raum eine gleichnamige Methode, mithilfe derer Schüler:innen kollaborativ an einem Google Doc arbeiten. Im Fokus des Lernens steht dabei das kritische Denken und die Problemlösefertigkeit (HyperDocs Admin, 2019). Die zweite Begriffsdefinition wurde von der Fachdidaktik Chemie der RPTU in Kaiserslautern geprägt und versteht unter HyperDocs zunächst vielfach mit **Hyperlinks** ausgestattete Dokumente (**Documents**), genauer Arbeitsblätter. Die Hyperlinks wiederum bildeten ursprünglich, zu Beginn der Entwicklung 2017 und 2018, einen Verweis auf Informationskärtchen,

wie zum Beispiel Lernhilfen und Vertiefungsaufgaben, die über den Browser aufgerufen werden konnten. Durch die digitale Umsetzung dieses Formats, in Abgrenzung zu einem analogen Arbeitsblatt, können Videos, Animationen, farbige Abbildungen oder auch Audiodateien eingebunden werden. Damit schafften HyperDocs multimediale Lernumgebungen. Diese ursprünglichen digitalen Arbeitsblätter wurden mithilfe von Word und eines einfachen eigens dafür entwickelten Content-Management-Systems namens HyperDocSystems umgesetzt, indem die jeweiligen Informationskarten als Hyperlink durch zum Beispiel einzelne Wörter oder spezielle Symbole in einem Worddokument verknüpft und anschließend das gesamte Arbeitsblatt als PDF-Datei exportiert wurde. Die Informationskarten selbst wurden in HyperDocSystems erstellt und verwaltet. Die Schüler:innen konnten die Arbeitsblätter auf einem digitalen Endgerät, vorzugsweise einem Tablet, öffnen und die jeweiligen Hilfen nutzen. Zu diesem Zeitpunkt war es bereits möglich, die Nutzung des zusätzlichen Informationsangebots nachzuverfolgen (Fitting et al., 2018). Gleichzeitig erfolgte die Sicherung jedoch auf einer analogen Kopie des Arbeitsblattes, sodass die HyperDocs eher als digitale Lernhelfer zu verstehen waren. Diese Methode wurde bei Fitting (2018) näher untersucht. Aus der Praxis heraus zeigten sich zum Teil einige empirisch nicht belegte, aber beobachtete Nachteile: Bei der Nutzung des eigentlichen HyperDocs öffneten sich die Informationskärtchen immer gleichzeitig im Browser, wobei ein Wechsel zwischen PDF-Dokument und Kärtchen stattfand. Im Rahmen der Multimedia-Theorie könnte dieser Wechsel zu einem erhöhten Cognitive Load führen (*Split-Attention-Effekt*; Ayres & Sweller, 2014). Der gleiche Effekt könnte eintreten, wenn die Lernenden wiederum von dem HyperDoc, auf dem sie zum Beispiel eine Hilfe erhalten, auf das analoge Arbeitsblatt zur Sicherung der Aufgabe wechseln. Ein technischer Nachteil bestand in der Distribution und Aktualisierung der digitalen Arbeitsblätter, da die PDF-Dateien auf die jeweiligen Tablets verteilt werden mussten. Handelte es sich zudem nicht um die eigenen Geräte der Lernenden, bestand keine direkte Möglichkeit, die HyperDocs wiederum den Schüler:innen zur Verfügung zu stellen. Zuletzt ist noch darauf hinzuweisen, dass die digitalen Arbeitsblätter nicht direkt bearbeitet werden konnten, da keine vorgegebenen Textfelder zur Eingabe von Lösungen eingebunden werden konnten und nur kostenpflichtige Software diesen Funktionsumfang bot. Damit waren die Lerngruppen immer auf die analoge Variante des Arbeitsblattes angewiesen.

Aus den eben beschriebenen Gründen wurde Ende 2018 bis Mitte 2019 an einer volldigitalen Version der HyperDocs gearbeitet. Diese sollte insbesondere die Verwaltung erleichtern und ein volldigitales Bearbeiten und Sichern der HyperDocs ermöglichen. Im

Rahmen dieser Arbeit wurde HyperDocSystems nach der dritten Pilotierung nochmals leicht überarbeitet (vgl. Kapitel 4.4.3), sodass im Folgenden der Stand von Sommer 2020 bis Frühjahr 2021 beschrieben wird. Bei der Entwicklung wurde auf eine einfache Handhabung und eine hohe Usability geachtet, indem unter anderem die Usability-Heuristiken von Nielsen (1994) betrachtet wurden. HyperDocSystems lässt sich auf nahezu allen Endgeräten mit einer Internetverbindung und einem Browser (getestet: Google Chrome, Microsoft Edge, Mozilla Firefox, Apple Safari) öffnen. Die Nutzer:innen müssen sich dazu auf HyperDocSystems anmelden. Dadurch kann die Bearbeitung eines Arbeitsblattes gespeichert und zu einem späteren Zeitpunkt erneut aufgerufen werden.

## 2.4.2 Aufbau und Funktion eines HyperDocs

Die nachfolgende Abbildung (8) zeigt exemplarisch den Funktionsumfang eines HyperDocs zum Zeitpunkt der Studie (2020 - 2021).

The screenshot shows a HyperDoc interface for concentration determination. At the top, there is a navigation bar with three steps: 1. Konzentration für gegebenes Körpergewicht bestimmen, 2. Mit Konzentration von Azorubin im Getränk vergleichen, and 3. Maximal Anzahl an Gläsern berechnen. Below this, the main content area is titled 'Arbeitsblatt: Konzentrationen' and contains a table for concentration determination. The table has columns for concentrations (20 mg/l, 30 mg/l, 40 mg/l, 50 mg/l, 60 mg/l) and a 'Probe' column. The table is highlighted in yellow. Below the table, there is a section titled 'Der Konzentrationsbegriff' with a text description of concentration. The interface also includes various interactive elements like buttons for zooming, editing, and navigation.

Abbildung 8: Ausschnitt eines HyperDoc mit den meisten Funktionen im Überblick. Dunkelgrün: Lernhilfe (vergrößerbar), hellgrün: Trigger für Lernhilfe, gelb: Tabelle, braun: Editierbare Abbildung, blau: Einfache Abbildung, schwarz: Textfeld. Komplettes HyperDoc: vgl. Anhang 9.1.3.

Im oberen Teil des HyperDocs ist eine geöffnete Hilfe (dunkelgrün) zu sehen. Diese Popups ermöglichen das Einblenden verschiedenster Informationen. Die jeweiligen Hinweise lassen sich über die Symbole (hier: „Hilfebuttons“) oder einzelne Wörter öffnen (hellgrün). Sie entsprechen den ursprünglichen Hyperlinks. Gleichzeitig geben die entsprechenden Symbole an, welche Art von Hilfe oder Zusatzinformation, bezogen auf die Informationsdarbietungsform, dargestellt wird. So lässt sich generell zwischen Audiodateien, Videos, Texten oder Bildern unterscheiden. Es existiert außerdem ein Symbol, das Vertiefungsaufgaben angibt (vgl. Abbildung 9). Nach Sicht von Squires und Preece (1999) handelt es sich um ein wichtiges Merkmal von Lernprogrammen, Multimediainhalte abzubilden.



*Abbildung 9: Kennzeichnung der Informationskärtchen auf einem HyperDoc. Blau: Hilfen in verschiedenen Darstellungsformen, rot: Vertiefungsaufgabe. Symbole: Fachdidaktik Chemie TUK (2021).*

Weitere Bestandteile des HyperDoc sind eine Tabelle, in der die editierbaren Abbildungen (braun) und ein einfaches Bild (dunkelblau) eingebettet sind, regulärer Text und ein Textfeld. Bei den Abbildungen und Textfeldern (schwarz), die eine voll-digitale Bearbeitung im Browser zulassen, handelt es sich um das Herzstück der HyperDocs. Die Textfelder können einerseits durch die Tastatur oder andererseits durch die Freihandeingabe mittels eines Apple Pencils oder anderen Stylus beschrieben werden. Während die Beschriftung durch die Tastatur keine umfangreichen Formatierungen zulässt, können beim freien Schreiben verschiedene Farben und Stiftbreiten eingesetzt werden (vgl. Abbildung 10). Dazu öffnet sich beim Tippen auf das Schreibsymbol eine Werkzeugleiste und der Hintergrund wird verdunkelt. Die Nutzer:innen können zwischen „Schreiben“, „Radieren“, „Stiftfarbe“ und „Stiftgröße“ auswählen. Zusätzlich besteht die Möglichkeit, ein ganzes Textfeld zu leeren (Symbol „Besen“).

Es können weitere Plugins, die Funktionsergänzungen zum eigentlichen HyperDoc darstellen, ergänzt werden. Ein Beispiel, das hier nicht gezeigt ist, ist der Datei-Upload, mithilfe dessen die Schüler:innen eigene Videos, Bilder oder Sprachaufnahmen hochladen und in das Arbeitsblatt integrieren können. Ergänzt werden können außerdem weitere interaktive Elemente von externen Webseiten, wie zum Beispiel *Learning Apps* (LearningApps, o. D.) oder *Padlets* (Padlet, 2022). Insgesamt können sowohl auf dem digitalen

Arbeitsblatt als auch auf den Hilfen verschiedene Inhalte (Videos, Bilder, Audios) ergänzt werden, wodurch eine multimediale Lernumgebung entsteht.

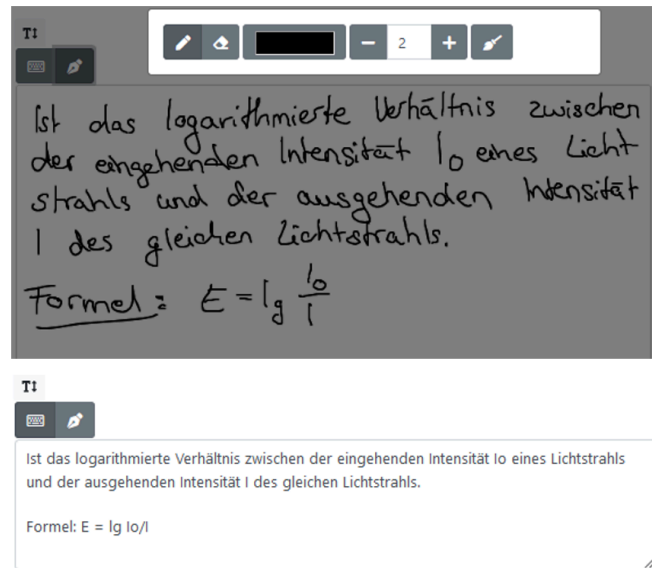


Abbildung 10: Beide Varianten der Textfeldbeschriftung. Oben: Beschriftung mit dem Apple Pencil und geöffnetes Schreibwerkzeug (Antwort einer Schülerin oder eines Schülers). Unten: Beschriftung mit der Tatstatur (keine weitere Formatierung möglich).

Ein HyperDoc besteht nicht aus einzelnen Seiten, die geblättert werden, sondern aus einer kontinuierlichen Seite, wobei zum Beispiel Trennstriche verschiedene Abschnitte markieren können.

### 2.4.3 HyperDocSystems

Mit der kostenlosen Software HyperDocSystems können HyperDocs erstellt, distribuiert und bearbeitet werden. Der generelle Funktionsumfang der HyperDocs wurde bereits im vorherigen Kapitel beschrieben. Die dort beschriebene Funktionalität steht den Schüler:innen zur Verfügung, während Lehrkräfte weitere umfangreichere Funktionen nutzen können. Nachfolgend wird zunächst die technische Umsetzung von HyperDocSystems näher beschrieben. Darauf folgt eine genaue Darstellung des Funktionsumfangs für Lehrkräfte beziehungsweise Ersteller:innen von HyperDocs.

Die aktuelle Version von HyperDocSystems basiert serverseitig auf der Programmiersprache *PHP 7* und wird ergänzt durch das *Zend Framework 3* (Laminas Project, 2022). Der Fokus liegt dabei auf einer objektorientierten Umsetzung und dem *Model-View-Controller*-Konzept. Seit 2020 wird das Zend Framework unter dem *Laminas Project* der *Linux Foundation* (Laminas Project, 2022; The Linux Foundation, 2022) geführt. Das



Frontend setzt sich hauptsächlich aus HTML und JavaScript zusammen. Dabei kommt die Javascript-Bibliothek *jQuery* zum Einsatz, womit insbesondere HTML-Veränderungen, Event-Handling oder auch Ajax-Abrufe deutlich vereinfacht werden (OpenJS Foundation, 2022). Die Datenbank wird mittels *MariaDB* (MariaDB Foundation, 2022) umgesetzt und mit *phpMyAdmin* (phpMyAdmin contributors, 2022) verwaltet. Bei der Programmierung von HyperDocSystems wurde auf eine einfache Bedienung sowie entsprechende Navigationshilfen und Rückmeldungen geachtet. Entsprechend sind die wichtigsten Funktionen, wie zum Beispiel das Erstellen von Arbeitsblättern und Hilfen, über gut sichtbare Buttons schnell erreichbar. Tooltips informieren die Benutzer:innen über entsprechende Funktionen verschiedener Buttons. *Prompts* beziehungsweise Hinweisfenster informieren den oder die Nutzer:in zudem über wichtige Vorgänge oder fordern eine Bestätigung der Aktion ein. HyperDocSystems wird zum aktuellen Zeitpunkt (Stand 09/2023) über einen Server der Rheinland-Pfälzisch Technischen Universität in Kaiserslautern bereitgestellt (Fachdidaktik Chemie TUK, 2021). Das System wird dabei von einem *Apache*-Webserver distribuiert (The Apache Software Foundation, 2022). Eine weitere Nutzung von HyperDocSystems über ein lokales Netzwerk wurde bereits in Fitting (2018) beschrieben. Dazu wird auf einem Laptop mit *Windows*-Betriebssystem<sup>1</sup> *XAMPP* (VMware, 2022) eingerichtet. Es handelt sich um ein Komplettpaket, bestehend aus *Apache*-Webserver, *MariaDB*, *PHP* und *Perl*. *XAMPP* ist für die Betriebssysteme *Windows*, *Linux* und *MacOS* frei verfügbar. In Kombination mit einem WLAN-Router lässt sich HyperDocSystems über ein lokales WLAN nutzen.

#### 2.4.4 Funktionalität von HyperDocSystems

Im folgenden Kapitel wird ein Überblick über die Funktionalität des Systems aus Perspektive der Lehrenden gegeben. Wesentlich kann zwischen drei Schwerpunkten unterschieden werden: Nutzer:innenverwaltung (Login, Registrierung, Anlegung / Löschung von Nutzer:innen), Lernmaterial (Arbeitsblatt, Hilfen) und Monitoring (Statistik zur Nutzung der Informationskärtchen).

---

<sup>1</sup> Auch andere Betriebssysteme (*Linux*, *MacOS*) sollten prinzipiell unterstützt werden. Diese wurden jedoch nicht getestet.

## Nutzer:innenverwaltung

Accounts für Lehrkräfte werden, unter Zuweisung einer Schule, von den Administrator:innen von HyperDocSystems erstellt. Jede Schule wiederum besitzt eine technische Leiterin oder einen technischen Leiter. Mit diesem Accountstatus können weitere Accounts für Lehrer:innen angelegt werden. Jeder Lehrkraft wiederum ist es möglich, Schüler:innenaccounts zu erstellen oder zu ändern. Über einen Account-Generator können zudem ganze Klassensätze von Accounts erstellt werden. Um die HyperDocs besser distribuieren zu können, ist es außerdem möglich, Klassen / Lerngruppen zu erstellen und Schüler:innen diesen zuzuweisen. Damit kann ein HyperDoc direkt mit einer ganzen Klasse oder Lerngruppe geteilt werden. Lernende können zu mehreren Klassen zugeordnet werden.

## Lernmaterial: Arbeitsblätter und Informationskärtchen

Die zentrale Funktion von HyperDocSystems liegt im Erstellen von digitalen Arbeitsblättern. Diese werden über einen Editor, der einen grundlegend ähnlichen Funktionsumfang wie *Microsoft Word* bietet, erstellt (vgl. Abbildung 11).

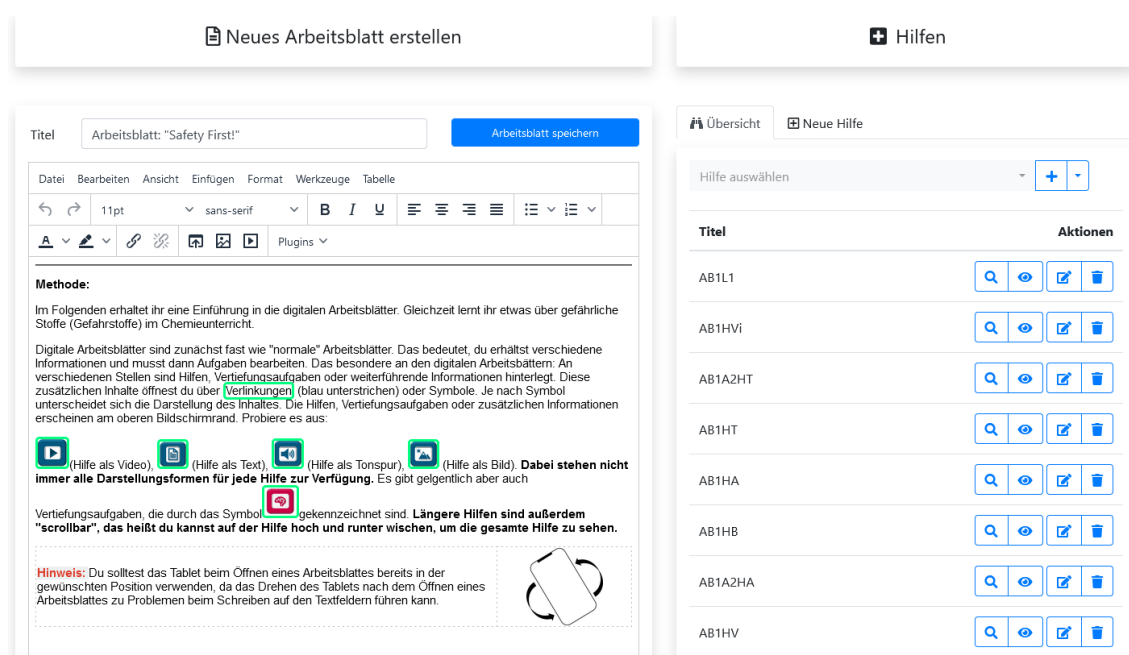


Abbildung 11: Ausschnitt aus dem Editor zum Erstellen und Bearbeiten eines HyperDoc. Links: Inhalte des HyperDocs. Rechts: Organisation der Informationskärtchen.

Der Titel eines HyperDoc ist frei wählbar. Bei den Formatierungen des eigentlichen Inhalts stehen die grundlegenden Funktionen zur Verfügung: Schriftarten, Schriftgröße, Schriftauszeichnungsart, Schriftfarbe, Schriftmarkierung, Textausrichtung, Listen, Aufzählungen und Sonderzeichen. In ein HyperDoc lassen sich zusätzlich verschiedene

Seite | 48

Medien (Video, Audio, Bilder) und andere Inhalte, zum Beispiel Tabellen, integrieren. Für die Medien steht ein Filemanager zur Verfügung, über den Dateien hochgeladen und eingefügt werden können. Es ist jedoch auch möglich, Onlineinhalte direkt mit dem Arbeitsblatt zu verknüpfen. Eine weitere zentrale Funktionalität, die ein einfaches Arbeitsblatt in ein HyperDoc umwandelt, ist die Integration der Informationskärtchen (s. rechte Spalte Abbildung 11). Dabei kann aus zuvor erstellen Kärtchen ausgewählt oder es können direkt neue Kärtchen erstellt werden. Dafür steht der gleiche Editor zur Verfügung. Es können jedoch keine Textfelder oder ein Dateiupload hinzugefügt werden. Ein Informationskärtchen kann über ein Wort oder ein Symbol mit dem Arbeitsblatt verknüpft werden. Im Editor wird die Verknüpfung über einen grünen Rahmen gekennzeichnet.

Erstellte HyperDocs erscheinen in einer Tabelle auf der Startseite. In dieser Übersicht können nun Arbeitsblätter gelöscht und bearbeitet werden. Weitere Funktionalitäten dieser Ansicht beinhalten: Zuweisung eines HyperDoc an einzelne Schüler:innen oder ganzen Klassen und Kopie an andere Lehrkraft. Die Übersicht aller Informationskarten ist ebenfalls in Form einer Tabelle organisiert. In der Übersicht können die Kärtchen bearbeitet, gelöscht oder die Verwendung auf einem HyperDoc gefunden werden. Die Lehrkraft hat außerdem die Möglichkeit, die bereits gespeicherten Lösungen der Schüler:innen zu einem Arbeitsblatt einzusehen.

### **Monitoring: Statistik zur Nutzung der Informationskärtchen**

Eine zentrale Funktion von HyperDocSystems ist das Monitoring der Informationskärtchen beziehungsweise Hilfen. Sobald eine oder ein Nutzer:in eine der Informationskarten öffnet, werden der Startzeitpunkt, der Accountname und der Name der Karte in der Datenbank hinterlegt. Wird die Karte geschlossen, erfolgt die Eintragung des Schließzeitpunktes. Dadurch können Annahmen über die generelle Nutzung und die spezifische Nutzung nach Verweildauer getroffen werden. Die Daten können wahlweise als csv-Datei exportiert oder direkt in der Übersicht die ersten fünf Hilfen beziehungsweise Nutzer:innen betrachtet werden. In den nachfolgenden Beispielen wird explizit von „Hilfen“ gesprochen, wobei alle Informationskärtchen in dieser Übersicht erscheinen.

Eine Übersicht zeigt alle genutzten Hilfen in einer detaillierten Tabelle (vgl. Abbildung 12). Dort können auch die gesamten Daten heruntergeladen werden, um sie zum Beispiel für Forschungszwecke zu nutzen. Diese Übersicht zeigt zudem die einzelne Verweildauer eines jeden Hilfeaufrufs sowie den Zeitpunkt der Öffnung. Um eine bessere Übersicht zu

erreichen, kann außerdem ein spezifisches Datum ausgewählt werden, wodurch die Hilfen gefiltert werden.

### Monitoring

Sehen Sie hier die Nutzungshäufigkeiten aller Hilfen [↗](#)

---

Datum:

📄

Arbeitsblatt	Hilfe	Benutzer	geöffne...	für [Zeit]
Arbeitsblatt: Farbstoffe in Lebensmitteln	AB2MA3HV	Igp1275	10.09.2020 12:02:06Uhr	0m 12s
Arbeitsblatt: Farbstoffe in Lebensmitteln	A2MV	Igp1274	10.09.2020 12:01:12Uhr	0m 3s
Arbeitsblatt: Farbstoffe in Lebensmitteln	A2MV	Igp1270	10.09.2020 12:01:08Uhr	0m 7s
Arbeitsblatt: Farbstoffe in Lebensmitteln	A2MV	Igp1268	10.09.2020 12:00:53Uhr	1m 0s

Abbildung 12: Detaillierte Übersicht der Hilfenutzung in HyperDocSystems.

Die bereits erwähnten detaillierteren Statistiken, die insbesondere für Lehrkräfte interessant sind, können Abbildung 13 und Abbildung 14 entnommen werden. Abbildung 13 zeigt die am meisten geöffneten Hilfen in absteigender Reihenfolge. Weiterhin ist die durchschnittliche Nutzungshäufigkeit aller Hilfen abgebildet.

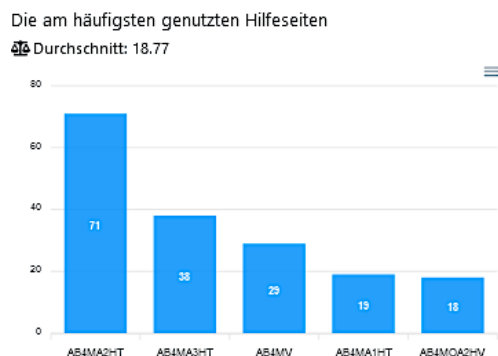


Abbildung 13: Die Übersicht zeigt die am häufigsten genutzten Hilfen.

Abbildung 14 gibt die Nutzungshäufigkeit der Hilfen aufgeschlüsselt nach Accounts in absteigender Reihenfolge wieder. In dieser Ansicht ist außerdem die durchschnittliche Nutzung der Hilfen über alle Benutzer:innen hinweg zu sehen.

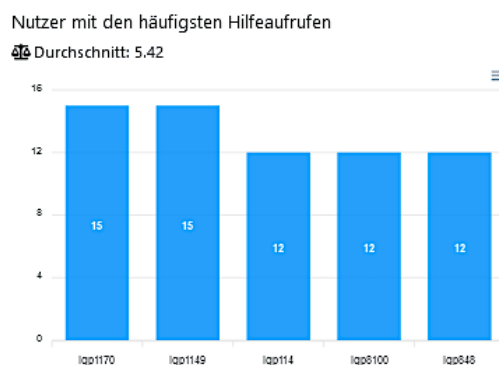


Abbildung 14: Die Übersicht zeigt die Nutzung der Hilfen, aufgeschlüsselt nach Account.

### 2.4.5 Theorie des Multimedialen Lernens und Cognitive Load

Das Lernen mit digitalen Medien folgt bestimmten psychologischen Grundvoraussetzungen, die im Rahmen verschiedener Theorien zusammengefasst wurden. Als zwei bedeutende Beispiele sind die Theorie des Multimedialen Lernens nach Mayer (2014) und die Cognitive Load Theory nach Sweller (1988) zu nennen. Während erstere eine Rahmung für die Gestaltung von multimedialen Lernumgebungen für das Lernen vornimmt, beschäftigt sich letztere mit konkreten Vorgängen bei der Verarbeitung von Informationen. Die Theorien sind aufgrund ihrer theoretischen Grundlagen eng miteinander verzahnt und sind in Verbindung mit HyperDocSystems, als ein multimediales Lernmedium, zu berücksichtigen.

#### Theorie des Multimedialen Lernens

Die Theorie des multimedialen Lernens von Richard E. Mayer ist ein theoretischer Ansatz, der verschiedene Regeln im Umgang mit dem Lernen von multimedialen Inhalten umfasst. Grundlegend geht es um das Lernen mit Bildern und Worten. Es wurde festgestellt, dass der Lernzuwachs größer ist, wenn beide Präsentationsformen, Text und Bild, gleichzeitig präsentiert werden, anstelle einer Einzelpräsentation dieser beiden Darstellungsformen. Diese theoretische und empirisch geprüfte Grundannahme ist unter der „multimedia learning hypothesis“ (Mayer, 2014a, S. 1) bekannt. Multimediale Inhalte begegnen uns überall im Alltag, insbesondere bei Lernprozessen, aber auch beim freizeitlichen Medienkonsum. Als Beispiele sind hierbei YouTube-Videos zu nennen, die sowohl gesprochenen Text als auch Bilder, oft Animationen, beinhalten. Nach Mayer (2014a, S. 2) ist der Begriff „Multimedia“ definiert als die gleichzeitige Präsentation von Wörtern und Bildern. Wörter können dabei als geschriebener oder gesprochener Text präsentiert werden. Unter Bildern versteht der Autor sowohl statische Bilder und

Abbildungen als auch dynamische Graphiken und vollständige Videos. Die Verarbeitung dieser Inhalte erfolgt nach der Kognitiven Theorie des Multimedialen Lernens (vgl. Abbildung 15). Das Modell beruht auf den Erkenntnissen der Verarbeitung von visuellen und verbalen Inhalten. Die Theorie geht auf das Arbeitsgedächtnismodell von Baddeley und die Duale Kodierungstheorie von Paivio zurück (Nieding et al., 2015, S. 66).

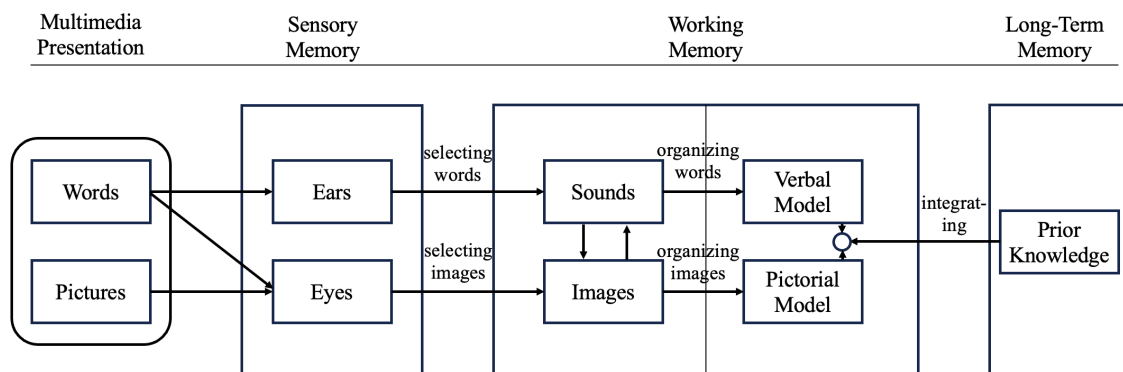


Abbildung 15: Kognitive Theorie des multimedialen Lernens (verändert nach Mayer, 2014b).

Das Modell umfasst in der Summe drei Gedächtnisspeicher: Das sensorische Gedächtnis, das Arbeitsgedächtnis und das Langzeitgedächtnis (vgl. Abbildung 15). Zwischen den verschiedenen Speichern finden die Prozesse der Selektion, Organisation und Integration statt. Die Selektion umfasst die Wahrnehmung lernrelevanter Inhalte basierend auf der multimedialen Repräsentation und der Überführung dieser Inhalte in das Arbeitsgedächtnis. Die Organisation beinhaltet die Strukturierung und Verknüpfung der Repräsentationen. Schließlich erfolgt ein Abgleich mit dem Vorwissen und eine Integration in das Langzeitgedächtnis. Die Informationen über einen Gegenstand werden nach der Cognitive-Load-Theorie (Sweller, 1988) und weiteren kognitiven Modellen in Form von Schemata im Langzeitgedächtnis gespeichert (Van Merriënboer & Kester, 2014). Schemata beinhalten Informationen über Objekte und Gegenstände, Situationen oder Ereignisse. Ein Schema selbst wird als einzelnes Informationsstück im Arbeitsgedächtnis, unabhängig von seinem Informationsgehalt, wahrgenommen und dient zur Speicherung der Information im Langzeitgedächtnis. Damit reduzieren Schemata die kognitive Belastung (Nieding et al., 2015, S. 58).

Jeder wahrgenommene Inhalt beansprucht eine bestimmte Menge an kognitiven Ressourcen. Während die Kapazitäten des sensorischen Speichers und des Langzeitgedächtnisses nahezu unbegrenzt sind, ist die Kapazität des Arbeitsgedächtnisses begrenzt (Mayer, 2014b). Der Lerninhalt kann auf verschiedenen Ebenen kognitive Ressourcen beanspruchen: Intrinsisch (*engl.: intrinsic*), lernirrelevant (*engl.: extraneous*) und lernrelevant

(*engl.: germane*). Auf dieser Grundlage wurde die Cognitive-Load-Theorie gegründet. Eine theoretische Herleitung der Theorie erfolgt im nächsten Absatz. Damit umfasst das Lernen mit multimedialen Inhalten insgesamt drei Annahmen: Die duale Kodierung der Inhalte, die Limitierung der Kapazität des Arbeitsgedächtnisses und die Voraussetzung der aktiven Verarbeitung von Inhalten (Mayer & Moreno, 2003). Aufgrund dieser verschiedenen Annahmen existiert auch Kritik an der Theorie. Insgesamt können die Ergebnisse nicht auf alle Fächer und Inhalte übertragen werden, da die Erforschung der Regeln zur Gestaltung multimedialer Lernumgebungen meistens aus Studien in den Naturwissenschaften gewonnen wurden. Weiterhin kann nicht für alle hypermedialen Lernumgebungen ein Wissenszuwachs festgestellt werden. Zuletzt sei erwähnt, dass sich die Theorie primär auf die Prozesse und Limitationen des Arbeitsgedächtnisses fokussiert und die Prozesse im Langzeitgedächtnis kaum beachtet werden (Nieding et al., 2015, S. 79).

### **Cognitive-Load-Theorie**

Die Cognitive-Load-Theorie geht auf Sweller (1988) zurück und umfasst die drei bereits vorgestellten Loadtypen. Alle drei Arten finden Relevanz bei der Speicherung und Verarbeitung von sekundärem biologischem Wissen. Dieses Wissen beinhaltet das grundlegende Fachwissen aller Fachrichtungen und ist, im Gegensatz zu Verhaltensweisen, nicht biologisch angeboren. Für das Lernen gelten in Bezug auf das sekundäre biologische Wissen zwei Annahmen (Sweller, 2016). Das Wissen ist bereichsspezifisch und muss entsprechend explizit angeeignet werden. Es ist im Langzeitgedächtnis gespeichert und muss bei entsprechenden Problemlösungen abgerufen werden (Tricot & Sweller, 2013). Die zweite Annahme besagt, dass der Lernprozess ausreichend unterstützt werden sollte. Erst bei hohem Vorwissen beginnen sich die Vorteile einer geführten Instruktion zu verringern (Kirschner et al., 2006).

Sweller (1988) unterscheidet in die drei bereits genannten Loadtypen *intrinsic*, *extraneous* und *germane*. Die *intrinsic* kognitive Last geht von dem Lerninhalt selbst aus und kann daher nicht beeinflusst werden, ohne den Lerninhalt selbst zu ändern. Eine hohe *intrinsic* kognitive Last liegt vor, wenn beispielsweise die Lerninhalte in einer Beziehung miteinander stehen. Als Beispiel führen Paas und Sweller (2014) das Lernen von neuen Vokabeln an. Eine Vokabel für sich besitzt eine geringe Last, wohingegen ein zusammenhängender Satz eine höhere Last verursacht, da er mehr Kapazität des Arbeitsgedächtnisses beansprucht. Der *Intrinsic Load* wird außerdem durch das Vorwissen der Person beeinflusst (Sweller et al., 2019). Der *Extraneous Cognitive Load* hingegen geht von

dem Lernmaterial selbst aus und sollte daher möglichst geringgehalten werden, um die Ressourcen des Arbeitsgedächtnisses nicht übermäßig zu belegen. Mit ihm ist die Mühe verbunden, die Darbietung der Informationen zu erfassen. Je mehr überflüssige oder schwer verständliche Inhalte dargeboten werden, desto höher ist der Extraneous Cognitive Load. Dabei spielt die Vernetzung mit anderen Inhalten, wie auch beim Intrinsic Load, eine Rolle (Sweller et al., 2019). Zuletzt ist der Germane Load zu beschreiben. Diesen müssen Lernende aufbringen, um die Inhalte in Schemata zu überführen. Der lernrelevante Load sollte daher, im Gegensatz zu den anderen beiden Typen, möglichst hoch sein, um einen hohen Lernzuwachs zu gewährleisten. Ursprünglich wurde er in Form einer Rechnung als die übrige freie Kapazität des Arbeitsgedächtnisses beziffert, die nicht auf die intrinsische oder die lernirrelevante Last entfallen (Sweller, 1988). Nach neuerer Ansicht dient der Germane Load im Arbeitsgedächtnis zur Verschiebung der Ressourcen zur Verarbeitung des Extraneous Load hin zur Verarbeitung des Intrinsic Load, womit der gesamte Load verringert wird (Sweller et al., 2019). Dadurch kann durch eine Verringerung des Extraneous Load die Lernqualität erhöht werden. Eine Überlastung des Arbeitsgedächtnisses tritt insbesondere durch eine hohe Interaktivität des Lernmaterials (Intrinsic Load), gepaart mit unübersichtlichem oder überfrachtetem Lernmaterial (Extraneous Load) auf. Ist jedoch die Interaktivität gering, kann die lernirrelevante Last hoch sein, ohne dass es zu einer Überanstrengung der oder des Lernenden kommt (Paas & Sweller, 2014).

Die Menge an Informationen, die im Arbeitsgedächtnis zur gleichen Zeit gespeichert werden kann, ist auf circa sieben Informationseinheiten begrenzt (Miller, 1956). Im Rahmen der Verarbeitung dieser Informationen können nicht mehr als vier Einheiten prozessiert werden (Cowan, 2001). Gleichzeitig ist die Dauer der Informationsspeicherung ebenfalls beschränkt. Sie beträgt in etwa 18 Sekunden, falls keine Wiederholung erfolgt (Peterson & Peterson, 1959). Um die Verarbeitung der Lerninhalte in einem größtmöglichen Umfang zu gewährleisten, sollte der Extraneous Load geringgehalten werden. In der Theorie des multimedialen Lernens wird dies über drei Ziele erreicht. Zunächst sollte die lernirrelevante Belastung geringgehalten werden, indem beispielsweise unnötige Informationen oder Visualisierungen vermieden werden (Kohärenzprinzip). Zudem sollte die essenzielle Verarbeitung im Arbeitsgedächtnis maximiert werden, indem zum Beispiel Text gesprochen und nicht gedruckt repräsentiert wird (Modalitätsprinzip). Zuletzt sollte die Integration der verbalen und visuellen Repräsentationen durch beispielsweise Lernhilfen (Prinzip der gelenkten Entdeckung) in das Langzeitgedächtnis gefördert werden (Mayer,



2014b). Es existieren weitere Prinzipien, nach denen das Lernen mit multimedialen Inhalten gestaltet werden sollte, siehe dazu Mayer (2014, S. 63) und Sweller (2016).

### **3. Forschungsstand und Forschungslücken**

Nachdem im vorherigen Kapitel die theoretischen Grundlagen und Begriffsverständnisse dargelegt wurden, erfolgt nun eine Ausweitung des im ersten Kapitel aufgeworfenen Studienüberblicks. Dadurch sollen Forschungslücken, die zur Motivation dieser Dissertation und Formulierung der Forschungsfragen beigetragen haben, aufgedeckt und diskutiert werden. Diese Forschungsfragen werden mit Detailfragen im letzten Teilkapitel formuliert. Auf einen Forschungsstand bezüglich HyperDocSystems wird verzichtet, da zum aktuellen Zeitpunkt (Stand 09/2023) lediglich Ergebnisse dieser (Fitting et al., 2021; Fitting & Hornung, 2022; Fitting et al., 2023) und einer weiteren Arbeit (Loth & Döhrmann, 2022; Loth, 2023) publiziert wurden. Bei letzterer steht HyperDocSystems nicht als Forschungsgegenstand im Mittelpunkt.

#### **3.1 Motivations- und interessenförderndes Potential digitaler Medien und multimedialer Lernumgebungen**

Der Einfluss der intrinsischen Motivation und des Interesses auf die Lernwirksamkeit ist hinreichend belegt (Lazowski & Hulleman, 2016). Daher sollte das Ziel guten Unterrichts sein, diese Persönlichkeitsmerkmale zu fördern. Ein breites Forschungsfeld beschäftigt sich mit der Frage, welchen Einfluss digitale Medien auf die Förderung der Motivation und des Interesses nehmen. Es existieren dazu viele Studien. Insgesamt besteht ein positiver Einfluss von digitalen Medien auf die Motivation und das Interesse, der in Teilen jedoch auch geschlechtsspezifisch ist (Higgins et al, 2019; Conradt & Bogner, 2016). Das erste Teilkapitel bietet daher einen Überblick über Studien, die vor allem Unterschiede zwischen den Geschlechtern feststellen konnten. Zum Teil ist die Wirksamkeit von Technologie auf die Motivation beziehungsweise das Interesse auf die Förderung des situativen Interesses zurückzuführen (z. B. Mitchell, 1997). Unter anderem weckt der Einsatz digitaler Medien die Neugier, was zu dem sogenannten „Neuheitseffekt“ (z. B. Schulz, 2020, S. 91) führt. Verschiedene Autor:innen erweitern die Theorie der Selbstbestimmung daher um den Aspekt der Neuheit aufgrund ihres motivationsfördernden Potentials (González-Cutre et al., 2016). Einige Metaanalysen fassen die Erkenntnisse in Bezug auf das digitale Lernen und die Motivation zusammen. Sie zeigen generell einen positiven Einfluss der digitalen Medien auf die Motivation und werden im letzten Teilkapitel vorgestellt. Im Folgenden werden nun zunächst einige allgemeinere Studien vorgestellt, die den Einfluss von Technologieeinsatz auf die Motivation untersuchten.

Speziell für den Chemieunterricht existieren wenig Studien, die explizit den bereits beschriebenen Einfluss der Motivation erforschten. Eine Studie von Lee und Osman (2012) sowie Osman und Lee (2014) untersuchte mit 127 Schüler:innen (weiblich: 77, männlich: 50; Alter: 16) einer weiterführenden Schule den Einfluss eines interaktiven multimedialen Programms zum Thema Elektrochemie. Es handelte sich um eine quasi-experimentelle Feldstudie. Als Variablen wurden der Lernstand und die Motivation erhoben. Zwischen den Gruppen zeigten sich in Bezug auf den Lernzuwachs knapp signifikante Unterschiede zugunsten der Experimentalgruppe ( $t(105.06) = 2.30, p = 0.024$ ). Die Gruppe erzielte insbesondere in Bezug auf die Erklärungen auf Teilchenebene bessere Ergebnisse. Auf Ebene der Motivation konnten zwischen den beiden Gruppen keine signifikanten Ergebnisse gefunden werden. Hervorzuheben sind die hohen Mittelwerte von  $M = 4.16$  (Experimentalgruppe) und  $M = 4.04$  (Vergleichsgruppe) der extrinsischen Motivation. Aufgrund des kulturellen Hintergrundes (Malaysia) zeigen die Schüler:innen generell wenig intrinsische Motivation (Lee & Osman, 2012). Auf der anderen Seite konnte eine Studie aus Indonesien zum Thema Säure-Base-Titration tendenziell einen positiven Einfluss des Einsatzes multimedialer Lernmaterials auf die Motivation feststellen (Sary et al., 2018). Die Studie zeigt allerdings methodische Schwächen, wie beispielsweise die fehlende Angabe der genauen t-Werte und Stichprobenszusammensetzung.

Eine deutsche Studie fand keine eindeutigen Unterschiede zwischen den Geschlechtern in der Motivation beim Einsatz einer multimedialen Lernumgebung. Die Autor:innen (Conradt & Bogner, 2016) untersuchten im Fach Biologie den Einfluss einer multimedialen Hypertext-Lernumgebung auf das Fachwissen und die Motivation zu drei Messzeitpunkten (Pre, Post und Follow-Up: 6 Wochen). An der quasi-experimentellen Untersuchung nahmen 393 Gymnasiast:innen teil (weiblich: 46 %, männlich: 54 %; Durchschnittsalter: 13.6). Insgesamt wurden vier Gruppen gebildet: Eine Gruppe mit einer computergestützten multimedialen Lernumgebung und einer computergestützten Sicherungsphase (CALC), eine Gruppe mit einer computergestützten multimedialen Lernumgebung und einer lehrkraftgestützten Sicherungsphase (CALT), eine Gruppe mit analogem Textbuch und lehrkraftgestützter Sicherung (TAL) und einer lehrkraftgestützten Sicherungsphase und eine Kontrollgruppe, die nicht an der Studie teilnahm. Die Schüler:innen erhielten eine zehnminütige Einführung, gefolgt von einer 60-minütigen Bearbeitungsphase, die in Partnerarbeit durchgeführt wurde. Die Sicherungsphase dauerte zehn Minuten und wurde entsprechend den Gruppen durch ein computergestütztes oder lehrkraftgestütztes Quiz abgeschlossen. In Bezug auf den Lernfortschritt zeigte die TAL-Gruppe die

größten Zuwächse. Sie unterschieden sich signifikant zu denen der CALT-Gruppe ( $Z = -4.43, p < 0.001$ ). Nach sechs Wochen waren diese Unterschiede nicht mehr signifikant. Zwischen den Geschlechtern konnten ebenfalls Unterschiede festgestellt werden. In der CALT- und TAL-Gruppe ( $Z = -2.39, p = 0.017$ ;  $Z = -3.78, p < 0.001$ ) schlossen die Schülerinnen signifikant besser ab, in der CALC-Gruppe die Jungen. In Bezug auf die Motivation, operationalisiert nach der Selbstbestimmungstheorie, konnten lediglich unter Kontrolle des Geschlechts signifikante Unterschiede zwischen den Gruppen gefunden werden. Diese Unterschiede betreffen die Schülerinnen. Während in der CALT-Gruppe Schülerinnen ein größeres Interesse als in der CALC-Gruppe zeigen ( $Z = -2.74, p = 0.006$ ), zeigt sich bei der wahrgenommenen Kompetenz ein umgekehrter Effekt ( $Z = -2.13, p = 0.033$ ). Für die TAL-Gruppe liegen in beiden Fällen keine signifikanten Unterschiede vor. In Bezug auf den Druck zeigt sich bei den Schülerinnen zwischen der CALT- und der TAL-Gruppe ein signifikanter Unterschied ( $Z = -3.08, p = 0.002$ ), da die TAL-Gruppe weniger Druck erfährt (Conradty & Bogner, 2016). Die Studie zeigt daher keinen Vorteil des Einsatzes von computergestützten Lernumgebungen in Hinblick auf die Motivation. Teilweise ergeben sich durch den Einsatz Nachteile hinsichtlich des Wissenszuwachses und Drucks.

In einem einfacheren Studiendesign untersuchten Jenö und Kolleg:innen (2017) den Einfluss des Einsatzes einer mobilen App bei der Erschließung botanischer Inhalte unter anderem auf die intrinsische Motivation, die wahrgenommene Kompetenz, die Autonomie, und die Lernwirksamkeit. An der Untersuchung nahmen 71 Zweitsemesterstudierende teil (weiblich: 65 %, männlich: 35 %; Alter 21 – 22). Die Kontrollgruppe arbeitete mit einem traditionellen Textbuch. Die Dauer der Intervention betrug drei Tage. Zwischen der Kontroll- und Experimentalgruppe zeigten sich signifikante Unterschiede in Bezug auf die intrinsische Motivation ( $t(69) = 7.24, p < 0.001, d = 1.73$ ) und die wahrgenommene Kompetenz ( $t(69) = 3.44, p < 0.001, d = 0.82$ ) sowie die Lernleistung ( $t(69) = 3.41, p < 0.05, d = 0.54$ ) zugunsten der Experimentalgruppe. Die Ergebnisse konnten durch die Gestaltung der App erklärt werden, da diese die wahrgenommene Kompetenz durch unmittelbares Feedback und die Autonomie durch dynamische, spielerische Funktionen unterstützte (Jenö et al., 2017). Die Ergebnisse der Studie konnten von der Forschergruppe in einer weiteren Studie ( $n = 58$ ) bestätigt werden. Während bei dieser Untersuchung zusätzlich signifikante Ergebnisse in Bezug auf die Autonomie gefunden wurden, zeigten sich keine signifikanten Unterschiede für die Lernwirksamkeit (Jenö et al., 2019b).

Schon Kandler (2002, S. 106, 132) hat sich mit dem Einfluss von Computern auf das Interesse von Schüler:innen beschäftigt. In ihrer Befragung mit 816 Schüler:innen (weiblich: 411, männlich: 405) der fünften bis siebten Jahrgangsstufe und 22 Lehrkräften aus nahezu allen Schultypen konnte gezeigt werden, dass die Schüler:innen bei Lernsoftware insbesondere die Autonomie- und Kompetenzunterstützung als wichtig erachten. Zudem fördert der technische Komfort das Interesse bei der Nutzung einer Lernsoftware. Im Gegenzug sorgen Programmängel zu einer Abnahme des Interesses. Die Instruktionsqualität wird zudem als wichtiger erachtet als die Multimedialinhalte. Die Studie konnte weiterhin zeigen, dass Mädchen signifikant häufiger eine Lernsoftware verwenden ( $t = 4.17$ ,  $df = 782.49$ ,  $p < 0.001$ ) und Jungen häufiger Computerspiele spielen ( $t = -6.99$ ,  $df = 757.17$ ,  $p < 0.001$ ).

### **3.1.1 Geschlechtsspezifische Unterschiede in Bezug auf die Motivation und das Interesse beim Lernen mit digitalen Medien**

Einige aufgeführte Studien zeigten signifikante Unterschiede in der Motivation beim Lernen mit digitalen Medien zwischen Jungen und Mädchen. Diese Unterschiede sind möglicherweise auf die unterschiedliche Nutzungshäufigkeit und Einstellungen gegenüber Technologie zurückzuführen, wie bereits bei Kandler (2002, S. 106, 116) zu sehen war. Insbesondere im E-Learningbereich bestehen zum Teil größere technische Hürden bei der Nutzung der Anwendungen. In anderen Anwendungsbereichen, wie beispielsweise *Social-Media* oder *E-Commerce*, können keine Unterschiede zwischen den Geschlechtern festgestellt werden oder die Ergebnisse sind widersprüchlich (Goswami & Dutta, 2016). In der bereits vorgestellten Studie von Conradt und Bogner (2016) zeigten sich ebenfalls Unterschiede zwischen Schülerinnen und Schülern, wobei signifikante Unterschiede zwischen den einzelnen Interventionen in Bezug auf die Motivation lediglich bei Mädchen festgestellt werden konnten. Gleichzeitig lagen die Werte für das Interesse und die wahrgenommene Kompetenz bei der rein computergestützten Lernumgebung unter denen der Jungen und auch beim Wissenstest schlossen die Schülerinnen in der voll digitalen Umgebung tendenziell schlechter ab.

Bruckermann und Kolleg:innen (2016) fanden ein motivationsförderndes Potential von digitalen Medien lediglich für Jungen. Hierbei handelte es sich jedoch nicht um eine vollständig digitale Lernumgebung, sondern um eine tabletgestützte Intervention in einem Schülerlabor. Dazu führten 100 Schüler:innen (weiblich: 59, männlich: 41; Durchschnittsalter: 13.40) ein Experiment zur Photosynthese durch. Neben einem Pre- und

Post-Test wurde nach drei Wochen ein Follow-Up-Test erhoben. Während die Kontrollgruppe in analoger Form arbeitete, führte die Experimentalgruppe folgende Schritte mittels eines Tablets durch: Messung der O<sub>2</sub>-Produktion, Dokumentation mittels Video und Auswertung mittels automatisch erstellter Graphen der eingesetzten App. Zwischen den beiden Gruppen lagen in Bezug auf die intrinsische Motivation ( $t(98) = 2.217, p = 0.029, d = 0.45$ ) und die wahrgenommene Wahlmöglichkeit ( $t(98) = 2.371, p = 0.02, d = 0.48$ ) signifikante Ergebnisse zugunsten der Experimentalgruppe vor. In Bezug auf das Interesse ergab sich kein signifikanter Unterschied. Unter Kontrolle des Geschlechts zeigten sich jedoch, ebenfalls zugunsten der Experimentalgruppe ( $t(39) = 3.152, p = 0.003, d = 1.01$ ), signifikante Unterschiede zwischen den Jungen beider Gruppen. Auch beim Wissenszuwachs zeigten die Schüler größere Zugewinne als die Schülerinnen ( $t(39) = 2.114, p = 0.04, d = 0.68$ ) im Post-Test. Im Follow-Up-Test wurden allerdings keine signifikanten Unterschiede zwischen den Geschlechtern oder Gruppen festgestellt (Bruckermann et al., 2016). Die Ergebnisse stehen dabei in Teilen in Widerspruch mit denen von Conradt und Bogner (2016), bei denen Schülerinnen in den nicht vollständig computergestützten Lernumgebungen besser abschlossen als die Jungen. Auch wenn beide Studien in Deutschland und in etwa zur gleichen Zeit durchgeführt wurden, ist das Untersuchungssetup unterschiedlich. Damit zeigt sich bereits die Relevanz, unterschiedliche Arten von technologischen Lernwerkzeugen in unterschiedlichen Kontexten zu erforschen. Verdeutlicht wird dieser Sachverhalt durch eine Studie von Liu und Kolleg:innen (2011), die beim Einsatz einer problemorientierten multimedialen Lernumgebung keine Unterschiede zwischen den Geschlechtern ( $n = 220$ ) in Bezug auf das Interesse, die wahrgenommene Kompetenz, die erbrachte Anstrengung und die Nützlichkeit feststellten.

Eine finnländische Untersuchung zeigte signifikante Unterschiede zwischen Jungen und Mädchen beim Lernen mit Tablets im Mathematikunterricht (Tossavainen & Hirsto, 2017). An der Befragung nahmen 138 Schülerinnen und 118 Schüler von der ersten bis zur achten Klasse teil. Eine Intervention fand nicht statt. Die Schüler:innen nutzten iPads im Mathematikunterricht im Rahmen des regulären Unterrichts. Die Befragung zur Motivation wurde lediglich mit einem Item abgedeckt. Insgesamt kam ein Fragebogen mit 92 Fragen und einer fünfstufigen Antwortskala vom Likert-Typ zum Einsatz. In der Befragung zeigte sich, dass in Hinblick auf die Motivation Jungen mehr Spaß beim Lernen mit iPads haben als Mädchen ( $t(253) = 2.05, p < 0.05, d = 0.26$ ). Der Effekt ist allerdings als klein anzusehen. Zusätzlich gab es zugunsten der Schüler signifikante Unterschiede in Bezug auf die Kompetenzen, Tablets zu nutzen ( $t(253) = 2.62, p < 0.01, d = 0.33$ ). Die

beiden eben genannten Items korrelieren positiv auf einem signifikanten Niveau miteinander ( $r = 0.25$ ,  $p < 0.001$ ). Auf der anderen Seite stimmten die Teilnehmenden eher mäßig zu, dass iPads das Lernen von mathematischen Inhalten einfacher gestalten ( $M_m = 2.68$ ,  $M_w = 2.28$ ). Bei Schülerinnen wurde die Unterstützung als signifikant schlechter wahrgenommen ( $t(172) = 2.28$ ,  $p < 0.05$ ,  $d = 0.33$ ). Außerdem bevorzugten Schülerinnen das analoge Lernen, während Jungen lieber mit dem iPad mathematische Inhalte erschließen (Tossavainen & Hirsto, 2017).

### 3.1.2 Neuheitseffekt im Zusammenhang mit digitalen Medien

Das motivations- und interessensfördernde Potential digitaler Medien kann in Teilen durch den sogenannten Neuheitseffekt erklärt werden (Silvia, 2005). Teilweise kann in verschiedenen Metaanalysen das lernförderliche Potential digitaler Medien durch diesen Effekt begründet werden, da länger andauernde Interventionen mit digitalen Werkzeugen geringere Effektstärken aufweisen (Higgins et al., 2019; Hillmayr et al., 2020). Im Speziellen konnten Higgins und Kolleg:innen (2019) zeigen, dass die Effektstärken der Interventionen in Bezug auf die Motivation kontinuierlich abnehmen, wobei die Abnahme zunehmend geringer ausfällt. Die Metaanalysen werden im Detail im Folgekapitel vorgestellt. Während manche Autor:innen (z. B. Hillmayr et al., 2020) als eine Erklärung für die bereits erwähnte nachlassende Wirksamkeit von digitalen Medien über die Zeit den Neuheitseffekt in Betracht ziehen, konnten andere Gruppen keinen eindeutigen Einfluss feststellen. Ein Beispiel hierfür ist die bereits vorgestellte Studie von Jenö und Kolleg:innen (2017), die unter Berücksichtigung der Vorerfahrung und Nutzung digitaler Medien keine Unterschiede auf die Ergebnisse der Intervention fanden und daher den Einfluss des Neuheitseffektes ablehnen. Die gleiche Forschergruppe beschäftigte sich später mit dem Neuheitseffekt in einem ähnlichen Untersuchungssetup (Jenö et al., 2019a). In der Studie mit 69 Biologiestudierenden (weiblich: 47.8 %; Durchschnittsalter: 21.80) wurde der Einfluss von Neuheit auf die Motivation und die Lernleistung untersucht. Die Studie fand im Rahmen eines Bestimmungskurses in Biologie statt. Die Teilnehmenden wurden in drei Gruppen eingeteilt: Eine Gruppe arbeitete mit dem regulären Textbuch, eine Gruppe führte die Aufgabe mit einem digitalen Textbuch mit den gleichen Inhalten durch und die letzte Gruppe führte die Bestimmung mit einer App durch, die unter anderem Hinweise und Rückmeldungen enthielt. Sowohl das digitale Textbuch als auch die App wurden als signifikant interessanter eingestuft ( $F(2,68) = 46.21$ ,  $p < 0.01$ ) als das traditionelle Textbuch, wobei die App den höchsten Neuheitswert aufwies. Gleiches gilt für die

wahrgenommene Kompetenz ( $F(2,65) = 7.02, p < 0.01$ ) und die Autonomie ( $F(2,66) = 12.89, p < 0.001$ ), wenn auch nicht in gleichem Maße. In einer Pfadanalyse wurde der Zusammenhang zwischen *Need-Satisfaction*, eigenständiger Motivation (intrinsisch und selbstreguliert), Internalisierung und Leistung modelliert. Die App hatte einen positiven Einfluss auf die *Need-Satisfaction* und die Neuheit. Während *Need-Satisfaction* ein positiver Prädiktor für die eigenständige Motivation darstellte, zeigte sich die Neuheit als ein negativer Prädiktor für die eigenständige Motivation. Zudem nimmt die eigenständige Motivation einen positiven Einfluss auf die Internalisierung. Gleichzeitig ist die Neuheit ein negativer Prädiktor für die Internalisierung und damit für die intrinsische Motivation. Dadurch zeigte sich insgesamt, dass die App durch ihre verschiedenen Mechanismen die *Need-Satisfaction* förderte und damit auch die eigenständige Motivation, während die Neuheit der Anwendung eher einen negativen Einfluss auf die eigenständige Motivation nahm. Bestätigt wurde dies durch den negativen Einfluss der Neuheit auf die Leistung der Studierenden. Die Autor:innen sehen daher weniger die Produkteigenschaften, zum Beispiel auch die Nützlichkeit oder die Effizienz, als positive Prädiktoren für die Lernwirksamkeit, sondern die zugrundeliegenden motivationalen Mechanismen der Nutzer:innen (Jeno et al., 2019a).

Die Ergebnisse werden zum Teil durch eine Studie von Kopinska (2020) unterstützt, die den Einsatz von digitalen Medien im Fremdsprachenunterricht erforschte. Über einen Zeitraum von 18 Monaten nutzten 77 Schüler:innen (weiblich: 48, männlich: 29; Alter: 16 – 17) einer spanischen weiterführenden Schule in einer von drei wöchentlichen Schulstunden digitale Technologien. Zu den Tätigkeiten zählten zum Beispiel das Erstellen und Lesen von Blogeinträgen, die Nutzung von PowerPoint oder auch Webquests. Bei der Untersuchung wurden die Einstellungen der Lernenden gegenüber der Nutzung von digitalen Medien im Unterricht erfragt. Das Instrument umfasst sechs Variablen: ICT<sup>2</sup> als Lernunterstützung, ICT als Mittel zur Verbesserung der englischen Sprache, ICT-Nutzung im Klassenraum, ICT als Mittel zur Verbesserung der Fähigkeiten und Karriere, ICT-Zugang in der Schule und Schwierigkeiten der ICT-Nutzung. In Bezug auf die Ergebnisse konnten zwischen den Geschlechtern keine Unterschiede festgestellt werden. Unterschiede zwischen dem Pre- und Post-Test lagen lediglich bei einer Variable, der Ausstattung in den Schulen, vor. Auch unter Berücksichtigung von leistungsstarken und

---

<sup>2</sup> Informations- und Kommunikationstechnologie (engl. *Information and Communication Technology*)



-schwachen Schüler:innen konnten keine Unterschiede in den Variablen festgestellt werden. Eine weitere Analyse, bei der die Lernenden in Bezug auf ihre Einstellung gegenüber der Nutzung von digitalen Medien im Unterricht in „neutral“ und „positiv“ eingeteilt wurden, zeigte signifikante Unterschiede in nahezu allen Variablen im Pre-Test. Im Post-Test hingegen zeigten sich diese Unterschiede nicht mehr. Während die neutral eingestellten Schüler:innen ihre Haltung verbesserten, verschlechterte sich die Haltung der positiv eingestellten Lernenden etwas. Die Ergebnisse zeigen, dass sich die Haltung der Lerngruppen gegenüber der Nutzung von digitalen Medien im Unterricht kaum ändert und, je nach Situation, eher positiv zunimmt. Unter Voraussetzung des Neuheitseffekts müsste sich jedoch die Einstellung verschlechtern, was jedoch nicht beobachtet werden konnte (Kopinska, 2020). Stoa und Chu (2023) fordern daher in Bezug auf das Bedürfnis nach Neuheit mehr Studien im Bereich des Schulunterrichts, die sowohl den kurzfristigen als auch langfristigen Einsatz in den Blick nehmen.

### **3.1.3 Metaanalysen zum Einfluss von Technologieeinsatz auf die Motivation und das Interesse**

Aufgrund des großen Interesses und der damit verbundenen hohen Anzahl an Studien zu dem Thema des Einflusses der Nutzung digitaler Technologien auf die Motivation und das Interesse, existieren verschiedene Metastudien, die diese Studien zusammenfassen. Dabei zeigen sich positive Eigenschaften der digitalen Medien auf die Lernleistung und die Motivation, die jedoch zum Teil von den eingesetzten Methoden und Werkzeugen abhängig sind.

Die Metaanalyse von Higgins und Kolleg:innen (2019) untersuchte die Wirkung des Technologieeinsatzes auf die Leistung, die Motivation und auf die Einstellung im Fach Mathematik. Die Analyse berücksichtigte am Ende des Auswahlprozesses 24 Studien mit insgesamt 4522 Proband:innen. Bei den Studien handelte es sich um Interventionen zwischen dem Kindergarten und der achten Klasse im Zeitraum zwischen 1985 und 2013. Die durchschnittliche gewichtete Effektstärke der Interventionen betrug für die Motivation  $d = 0.30$ , die Leistung  $d = 0.68$  und die Einstellung  $d = 0.59$ . Innerhalb der Gruppen waren die Effektstärken mit  $d = 0.49$  (Leistung) und  $d = 0.53$  (Motivation) mittelstark. Unter Berücksichtigung des Zwischengruppendesigns zeigten sich bei der Motivation zum Teil deutliche Unterschiede. Beim Einsatz von Experimentalgruppen mit verschiedenen Interventionen lag die durchschnittliche Effektgröße bei  $0.64$ , während bei Interventionen ohne Kontrollgruppe die Effektstärke bei  $0.26$  lag und bei einem klassischen

Experimental-/Kontrollgruppendesign keine signifikanten Unterschiede zwischen den Gruppen gefunden werden konnten. Auch bei der Dauer der Studien zeigten sich Unterschiede in den Effektstärken. Kürzere Interventionen über eine bis fünf Sitzungen zeigten mit  $d = 0.91$  die größte Wirkung auf die Motivation. Bei sechs bis zehn Sitzungen lag der Effekt nur noch bei  $d = 0.28$  und bei elf oder mehr Sitzungen lag der Effekt bei  $d = 0.22$  (Higgins et al., 2019). Damit zeigt sich ein abnehmender Effekt des Einsatzes digitaler Medien auf die Motivation. Auch Hillmayr und Kolleg:innen (2020) verweisen in ihrer Metaanalyse auf den geringeren Effekt digitaler Medien auf die Lernleistung bei länger andauernden Studien und begründen diesen Effekt mit der abnehmenden Motivation. Sie fanden außerdem große Effektstärken für den Einsatz von Tutoringsystemen und geringe Effektstärken für die Verwendung von Hypermediasystemen. Während bei ersteren insbesondere Feedbackfunktionen, Aktivierung des Vorwissens und eine strukturierte Erschließung Anwendung finden, sorgen letztere oftmals für eine Überforderung der Lernenden durch die fehlende Struktur und die Informationsfülle. Eine weitere Feststellung der Arbeitsgruppe liegt in den größeren Effektstärken bei Nutzung der Geräte in Partnerarbeit, anstelle von Einzelarbeit (Hillmayr et al., 2020). Die bisherigen Ergebnisse in Bezug auf die Lernwirksamkeit digitaler Medien werden von neueren Metaanalysen, zumindest für den Mathematikunterricht, gestützt (Ran et al., 2021; Ran et al., 2022).

Wie Hillmayr und Kolleg:innen (2020) bereits zeigten, führen Hypermediasysteme im Vergleich zu anderen Werkzeugen eher zu geringeren Lerneffekten. Moos und Marroquin (2010) haben sich genauer verschiedene Studien zu multimedialen, hypermedialen und hypertextualen Lernumgebungen unter Berücksichtigung der Motivation betrachtet und in ihrem Review zusammengefasst. Insgesamt konnten 35 Artikel nach dem Auswahlprozess gewonnen werden. Davon konnten 20 für die Beantwortung der Frage genutzt werden, wie Lernumgebungen die Motivation beeinflussen. Acht Studien wurden genutzt, um den Einfluss der motivationalen Variablen in der jeweiligen Umgebung auf die Lernleistung zu untersuchen und sieben Studien wurden herangezogen, um den Lernprozess beziehungsweise Lernweg näher zu untersuchen. In Bezug auf die Motivation zeigen hypermediale und multimediale Lernumgebungen einen interessensfördernden Charakter. Verantwortlich dafür ist zum einen wahrscheinlich der bereits beschriebene Neuheitseffekt, wodurch bei längerer Nutzung dieses Potential abnimmt. Zum anderen haben Lernende in solchen Lernumgebungen eine bestimmte Autonomie in Bezug auf den Lernweg und die Darstellung der Lerninhalte, wodurch ebenfalls das Interesse und die Motivation gefördert werden. Bei Schüler:innen mit geringerem Vorwissen kann jedoch das Fehlen

von Struktur und Führung die Motivation beeinträchtigen, da es zu einer Überforderung kommt. Diesen Sachverhalt haben bereits Hillmayr und Kolleg:innen (2020) angeführt. In Bezug auf die Lernwirksamkeit der verschiedenen Lernumgebungen zeigen sich ähnliche Ergebnisse, die durch die gleichen Sachverhalte begründet werden können. In Bezug auf die Lernförderung durch ein erhöhtes Interesse sorgen hingegen zusätzliche Funktionen, die dieses Interesse erhöhen, jedoch nicht zwangsläufig für den Lernprozess von Relevanz sind, für eine zusätzliche Belastung des Arbeitsgedächtnisses. Die letzte Forschungsfrage der Autor:innen lässt sich in ähnlicher Weise wie die vorherigen beantworten. Der Lernweg innerhalb einer multimedialen oder hypermedialen Lernumgebung ist stark abhängig von dem Interesse, dem Vorwissen und der Selbstwirksamkeit der Lernenden (Moos & Marroquin, 2010).

## 3.2 Usability

Die Usabilityforschung weist eine lange Historie auf. Schon früh wurde die Bedeutung der Bedienbarkeit von Softwaresystemen erkannt (Root & Draper, 1983). Primär richtet sich die Forschung auf Programme, die außerhalb des Bildungskontextes eingesetzt werden. Größere Unternehmen evaluieren ihre Software in einer Vielzahl von Usabilitystudien. Ein Beispiel hierfür sind Siemens und IBM (z. B. Lewis, 1995). Im Bereich der Schul- und Universitätsbildung spielt die Usability ebenfalls eine wichtige Rolle. Dazu existieren verschiedene Studien, die im Folgenden dargestellt werden. Darüber hinaus werden Forschungsergebnisse präsentiert, die allgemeine Erkenntnisse zur Benutzbarkeit liefern und im Rahmen dieser Arbeit relevant sind.

### 3.2.1 Zusammenhang zwischen Usability, Interesse, Motivation und Lernwirksamkeit

Eine Studie, die explizit den Zusammenhang zwischen Usability, Leistung und Interesse sowie weitere Parameter in Zusammenhang mit der Motivation untersuchte, fand eine mittelstarke signifikante Pearson-Korrelation zwischen Interesse und Usability ( $r = 0.396$ ) sowie Leistung und Usability ( $r = 0.384$ ). Die Autor:innen (Karapanos et al., 2018) der Studie untersuchten diese Faktoren im Rahmen einer quasi Feldstudie in zwei Biologieleistungskursen mit insgesamt 31 Schüler:innen zwischen 16 und 19 Jahren. Dazu erarbeiteten die Lernenden in einer 90-minütigen Einheit in Partnerarbeit in einem virtuellen Labor Inhalte zu Zellkulturen. Die Usability wurde mittels des *SUS*, das Interesse

durch eine Subskala des *FAM (Fragebogen zur aktuellen Motivation)* und die Lernleistung durch einen eigens konzipierten Multiple-Choice-Test erhoben. Der Zusammenhang zwischen Usability und Leistung konnte weiterhin in einer Regressionsanalyse gezeigt werden. Die Autor:innen gehen von einer Wirksamkeit der Usability auf den Lernerfolg unter Berücksichtigung der Cognitive-Load-Theorie aus, sodass eine schlechtere Usability zu einer höheren kognitiven Last führt. In Bezug auf den Zusammenhang zwischen Interesse und Benutzbarkeit konnten die Forscher:innen keine eindeutige Wirkungsrichtung angeben. Die theoretischen Zusammenhänge sind diesbezüglich nicht eindeutig. Beispielsweise könnten eine positivere Einstellung und ein höheres Interesse an der Lernumgebung zu einer höheren Bewertung der Usability führen. Unterschiede zwischen den Geschlechtern wurden in Bezug auf das Interesse und die Usability festgestellt. Schüler schätzen beide Faktoren mit einem signifikanten mittelstarken Effekt höher ein als Schülerinnen. In der Veröffentlichung wurden weitere Forschungsfragen in Bezug auf die Wahrnehmung der Usability formuliert. Ein Zusammenhang zu den computerbezogenen Kompetenzen ist nach Ansicht der Autor:innen zu erwarten (Karapanos et al., 2018).

Die Bedeutung der Usability für das Lernen wird ebenfalls bei Simon und Kollegen (2008) hervorgehoben. In einer breiten Studie befragten sie Studierende, die eine Lernplattform im Rahmen des Studiums nutzen, bezüglich deren Usability und Zuverlässigkeit, ihrer Motivation, der Medienpräferenz, der Zufriedenheit und der Lernunterstützung. Die Autor:innen konnten in ihrem Modell zeigen, dass die Medienpräferenz, analog oder digital, einen Einfluss auf die Zufriedenheit und die Lernunterstützung aufwies. Die einfache Bedienbarkeit und die Qualität der präsentierten Inhalte hatten ebenfalls einen signifikanten positiven Einfluss auf die Zufriedenheit. Darüber hinaus konnte ein schwacher Effekt ( $p < 0.1$ ) der Motivation auf die Lernunterstützung identifiziert werden. In ihrer Zusammenfassung identifizieren Simon et al. (2008, S. 729) die Usability „als bedeutenden Einflussfaktor für die Zufriedenheit der Nutzer“.

Eine weitere Studie, die die Relevanz der Usability für das Lernen hervorhebt, untersuchte den Zusammenhang zwischen der Benutzbarkeit zweier Lernplattformen und der Lernwirksamkeit. Tselios et al. (2001) erforschten bei Erstsemesterstudierenden die Auswirkungen der unterschiedlichen Usability zweier Lernplattformen im Rahmen eines Distanzlernkurses. Die Studierenden ( $n = 88$ ; keine weiteren Angaben) absolvierten am Ende des Kurses einen Leistungstest, der überwiegend aus Multiple-Choice-Fragen bestand. Diejenigen, die die besser nutzbare Lernplattform verwendeten, erzielten signifikant

bessere Ergebnisse im Leistungstest. Auf Ebene der Individuen konnte allerdings keine Korrelation zwischen Usabilitybewertung und Leistung festgestellt werden. Die Autor:innen heben in ihrer Konklusion ebenfalls die Wichtigkeit der Betrachtung der Usability bei Lernprogrammen hervor. Die gleiche Fragestellung wie bei Tselios und Kolleg:innen, wurde im Rahmen eines fünfzehnwöchigen hybriden Lernkurses mit 181 Studierenden (54.70 % weiblich) untersucht (Meiselwitz & Sadera, 2008). Die Inhalte wurden zu 50 Prozent digital unterrichtet. Die meisten Lernenden (76 %) hatten bereits zwei bis fünf Onlinekurse belegt. Die subjektive Lernleistung und Usability wurden mittels eines speziell entwickelten Fragebogens erhoben. In der Untersuchung zeigte sich ein signifikanter positiver Zusammenhang zwischen der Usability und der Lernwirksamkeit ( $F(3,156)=113.46, p < 0.005$ ). Das Regressionsmodell konnte 68 % der Varianz der Lernleistung durch die Usability erklären. Die Forscher:innen konnten außerdem zeigen, dass eine höhere Kompetenz in computerbezogenen Tätigkeiten, wie beispielsweise Internetsuche oder der Umgang mit E-Mails, zu einer höheren Bewertung der Usability führte. Die Studie zeigt damit ähnliche Ergebnisse wie Karapanos und Kolleg:innen (2018). Im Gegensatz dazu konnte sie jedoch keine Unterschiede zwischen den Geschlechtern feststellen. Die Gründe dafür könnten nach den Autor:innen in der homogenen Zusammensetzung der Gruppe in Bezug auf das Geschlecht liegen (Meiselwitz & Sadera, 2008). Der Einfluss der Usability auf die Lernwirksamkeit konnten in Teilen auch Davids et al. (2014) zeigen. Sie untersuchten zwei Versionen eines Systems (Standardversion und verbesserte Version). Es handelte sich dabei um ein spezifisches E-Learningprogramm im Bereich der Medizin. Die Proband:innen, die die überarbeitete Version nutzten, erzielten nicht-signifikant bessere Ergebnisse bezüglich der Lernleistung. Ein Grund für die fehlende Signifikanz sehen die Autor:innen in dem hohen Ausbildungsstand der Teilnehmer:innen, wodurch eine schlechte Usability, die zu einem höheren Extraneous Cognitive Load führt, die Lernwirksamkeit nicht ausreichend reduziert. Sie schlagen daher weitere Studien mit Novizen unter Einbezug weiterer Messungen, wie zum Beispiel dem Cognitive Load oder der Motivation vor (Davids et al., 2014).

Eine umfangreiche Studie von Venkatesh (2000) untersuche den Zusammenhang zwischen der Einfachheit der Nutzung (*Ease of Use*), der intrinsischen Motivation (bezogen auf die Computernutzung) und der Kontrollierbarkeit als ein Teil des *Technology Acceptance Model* (Davis et al., 1989). Das postulierte Modell wurde im Rahmen dreier Erhebungen untersucht. Über drei Monate wurden insgesamt drei Messungen erhoben. Die Stichprobengröße umfasste zwischen 52 und 160 Personen, die im Rahmen ihres Berufes

verschiedene Softwaresysteme nutzen. Die Zusammenhänge zwischen den Variablen des Modells variierten teilweise zu den unterschiedlichen Messzeitpunkten. Der Autor konnte mit seinem Modell den Einfluss der intrinsischen Motivation, der computerbezogenen Selbstwirksamkeit und der Computerangst auf die Einfachheit der Nutzung aufzeigen (Venkatesh, 2000). Der Einfluss der *Computer Playfulness*, operationalisiert als die intrinsische Motivation, zeigt in den ersten beiden Testzeitpunkten (Einführung und einmonatige Nutzung) einen signifikanten Einfluss auf die Einfachheit der Nutzung. Insgesamt sollte nach Meinung des Autors mehr Fokus auf die individuellen Unterschiede der Nutzer:innen anstelle der objektiven Messung der systembedingten Usability gelegt werden. Der Autor betont die Rolle der intrinsischen Motivation und computerbezogenen Selbstwirksamkeit für das Arbeiten mit Softwaresystemen (Venkatesh, 1999; Venkatesh, 2000). In Anlehnung an Venkatesh (2000) konnten auch Hackbarth et al. (2003) den negativen Einfluss von Computerangst auf die Einfachheit der Nutzung feststellen. Die Ergebnisse bezüglich der computerbezogenen Selbstwirksamkeit konnten zudem von Orfanou et al. (2015) in einer Studie mit 769 Studierenden (weiblich: 487, männlich: 282; Durchschnittsalter: 21.78) verschiedener Fachrichtungen gezeigt werden. Die Lernenden arbeiteten während eines Semesters mit zwei verschiedenen Lernplattformen und bearbeiteten nach Abschluss der Lehrveranstaltung einen Fragebogen. Die Forscher:innen fanden einen signifikanten Zusammenhang ( $r = 0.326, p < 0.01$ ) zwischen der Internet-selbstwirksamkeit und der Usability. Weiterhin wurden die Studierenden ausgehend von den Analysen in Gruppen mit niedriger und hoher Selbstwirksamkeit eingeteilt. Die Gruppen unterschieden sich in Bezug auf die wahrgenommenen Usability der Lernplattform signifikant ( $r = 0.31, p < 0.001$ ). Zudem besteht ein signifikanter Zusammenhang zwischen der Einstellung gegenüber dem Internet und der Bewertung der Benutzbarkeit ( $r = 0.223, p < 0.01$ ). Eine Korrelation zwischen Alter und Usability konnte nicht festgestellt werden (Orfanou et al., 2015). Unterschiede zwischen den Geschlechtern in Bezug auf die Usabilitybewertung konnten ebenfalls nicht festgestellt werden.

#### **Cognitive Load**

Der Cognitive Load spielt eine wichtige Rolle beim Lernen mit digitalen Medien. Insbesondere der Germane Cognitive Load und der Extraneous Cognitive Load können beim Einsatz von interaktiven digitalen Lernumgebungen beeinflusst werden (Koć-Januchta et al., 2022). Ein Zusammenhang zwischen dem Cognitive Load und der Usability ist weitestgehend bekannt. Bereits Bevan (1995b) erkannte die Relevanz des Cognitive

Workload und des *Mental Effort* bei der Messung der Usability. Zunehmend findet daher auch die Cognitive-Load-Theorie Beachtung im Forschungsfeld der Human-Computer-Interaction (Harrison et al., 2013; Hollender et al., 2010). Insgesamt ist der Anteil der Studien, die den Cognitive Load als Usability-Attribut beachten, immer noch gering (Weichbroth, 2020). Davids und Kolleg:innen (2015) erachten die Usability in Verbindung mit der kognitiven Last für wichtige Evaluationsmerkmale im Bereich des E-Learnings. Huang (2019) leitet aus einem systematischen Review insgesamt sieben Usability-Heuristiken ab, wovon eine den Cognitive Load repräsentiert.

In einer Untersuchung von Ibili und Billinghamurst (2019) wurde der Zusammenhang zwischen Usability und Cognitive Load in Abhängigkeit vom Geschlecht untersucht. Dazu wurde eine AR-Anwendung in einer weiterführenden Schule eingesetzt, die das Lernen von dreidimensionaler Geometrie unterstützen sollte (weiblich: 30, männlich: 29; Alter: 13 – 14). Es handelt sich um eine Feldstudie, bei der die Anwendung über vier Wochen im Schulunterricht eingesetzt wurde. Es bestanden keine signifikanten Unterschiede zwischen den Geschlechtern in Bezug auf die drei Load-Typen. Das Gleiche gilt für die Usability. Unter Berücksichtigung des Zusammenhangs zwischen Usability und Cognitive Load ergaben sich je nach Geschlecht signifikante Ergebnisse. Sowohl bei den Schülerinnen als auch bei den Schülern zeigten sich signifikante positive Korrelationen zwischen der Nützlichkeit und dem Germane Load sowie zwischen der Einfachheit der Nutzung und dem Germane Load ( $r = 0.382 - 0.782$ ,  $p = 0.037 - 0.001$ ). In Bezug auf den Intrinsic Load zeigte sich ein signifikanter negativer Zusammenhang zur Nützlichkeit bei Schülerinnen, jedoch nicht bei Schülern. Umgekehrt lag bei Letzteren ein signifikant negativer Zusammenhang zwischen dem Extraneous Load und der Nützlichkeit vor ( $r = -0.371$ ,  $p = 0.048$ ). Die Autor:innen führen die Unterschiede unter anderem auf die unterschiedliche Bewertung der Usability und auf den erhöhten Germane Load bei Schülerinnen zurück (Ibili & Billinghamurst, 2019).

Eine Voruntersuchung, die einen Vergleich zwischen einem AI-basierten Textbuch und einem eBook mit 16 Biologiestudierenden eines Einführungskurses durchführte, zeigten einen starken negativen Zusammenhang zwischen der Schwierigkeit des Lerngegenstandes und der Usability (Koć-Januchta et al., 2020). Ein Zusammenhang zwischen der Usability und dem Mental Effort konnte nicht festgestellt werden. Die Studie wurde zu einem späteren Zeitpunkt mit einem anderen Instrument zur Erfassung der drei verschiedenen Load-Typen wiederholt (Koć-Januchta et al., 2022). An der Studie nahmen zu Beginn 42

Studierende Teil (weiblich: 69 %; Durchschnittsalter: 26.28). Die absolute Anzahl der Befragten variierte zwischen 25 und 41, je nach Messzeitpunkt ( $n = 4$ , wovon zwei mit Cognitive-Load- und Usabilityerhebung). Die Student:innen konnten über den gesamten Kurszeitraum (ca. 1.5 Monate) auf ein AI-basiertes Textbuch zurückgreifen. Zum ersten Messzeitpunkt korrelierte der Germane Load positiv mit der Usability ( $r = 0.51, p < 0.01$ ). Dieser positive Zusammenhang blieb auf einem ähnlichen Niveau zum zweiten Messzeitpunkt bestehen. In Bezug auf den Extraneous Load zeigte sich ein signifikant negativer Zusammenhang zur Usability bei beiden Messzeitpunkten ( $r = -0.69, p < 0.01$ ;  $r = -0.76, p < 0.01$ ). Zum zweiten Messzeitpunkt zeigte sich außerdem eine signifikant negative Korrelation zwischen dem Intrinsic Load und der Usability ( $r = -0.5, p < 0.01$ ). Die Ergebnisse der Studie entsprechen damit in großen Teilen denen von Ibili & Billinghurst (2019).

### 3.2.2 Evaluation und Messung von Usability

Insgesamt besteht ein zunehmendes Interesse an der Evaluation der Usability von Softwaresystemen. Speziell im Bereich des *World Wide Webs* spielt die Benutzbarkeit eine zunehmende Rolle. In einer systematischen Mapping-Studie von Fernandez und Kolleg:innen (2011) mit 206 Publikationen wurde die Zunahme von Evaluationsstudien ab dem Jahr 2004 festgestellt. Dieser Sachverhalt verdeutlicht die zunehmende Relevanz der Usability. Weiterhin konnten die Autor:innen Fragebögen als eine häufig eingesetzte Methode der Usabilityerhebung identifizieren. Zu einem ähnlichen Schluss kommen Abuhlfaia und Quincey (2018), die ebenfalls eine systematische Mapping-Studie mit 61 Veröffentlichungen im Bereich der Hochschulbildung durchführten. In 50 Prozent der Studien wurden Fragebögen zur Evaluation der Benutzbarkeit eingesetzt. Sie fanden zudem als meist untersuchte Faktoren der Usability folgende Verteilung: Effektivität (23 %), Zufriedenheit (19 %), Effizienz (17 %), Einfachheit der Nutzung (14 %), Lernbarkeit (13 %), Probleme (12 %), Nützlichkeit (11 %), Erinnerungsfähigkeit (8 %) und Fehler (1 %). Die Verteilung entspricht damit der grundlegenden Definition der Usability. Bei dieser Erhebungsform sollte jedoch auf standardisierte Instrumente zurückgegriffen werden (Hornbæk & Law, 2007). Untersuchungen der Usability sollten insgesamt stärker auf einem größeren Untersuchungsbereich stattfinden, der sich über mehrere Stunden und Erhebungszeitpunkte erstreckt sowie komplexere Aufgaben in den Blick nimmt. In diesem Zusammenhang sollten zudem weitere Parameter, die in Verbindung mit der Usability stehen könnten, untersucht werden. Dadurch soll ein besseres Verständnis der



Beeinflussung der Benutzbarkeit einer Anwendung gewonnen werden (Hornbæk, 2006; Hornbæk & Law, 2007).

Es existieren in der Literatur zudem weitere Einflussfaktoren auf die Usability, die bei einer Erhebung berücksichtigt werden müssen. Ein möglicher Faktor ist das kulturelle beziehungsweise kontinentale Umfeld, in dem die Erhebung stattfindet. Wallace und Yu (2009) stellten in ihrer Studie mit 23 Teilnehmer:innen, zusammengesetzt aus 13 taiwanischen (weiblich: 4, männlich: 9; Durchschnittsalter: 26) und zehn nordamerikanischen Tester:innen (weiblich: 4, männlich: 6; Durchschnittsalter: 28), signifikante Unterschiede in der Bewertung der Usability eines MP3-Players fest. Zu einem ähnlichen Schluss kam eine Studie von Shen und Kolleg:innen (2006). Sie untersuchten den Einfluss des Interfacedesigns durch Variation der Icons auf die Nutzung eines künstlichen webbasierten Betriebssystems. Es zeigten sich, je nach sozio-kulturellem Hintergrund, Unterschiede.

Ein weiterer Einflussfaktor auf die Usabilitybewertung besteht durch den Zeitpunkt der Messung. Dabei existieren zum Teil widersprüchliche Ergebnisse. In der bereits vorgestellten Studie von Orfanou und Kolleg:innen (2015) zeigte sich eine steigende Bewertung der Benutzbarkeit des Lernmanagementsystems mit zunehmender Verwendungsdauer. Zu dieser Erkenntnis kommt auch Wang (2021), der jedoch einen differenzierten Blick auf den Einfluss des Testzeitpunktes über die Erhebungszeit warf. In der Studie mit 30 Teilnehmenden (weiblich: 19, männlich: 11; Alter: 20 - 26) wurde der Einfluss des Messzeitpunktes auf die Usabilitybewertung untersucht. Die Proband:innen verwendeten dazu das Programm „Adobe InDesign CS6“. Insgesamt wurden zehn Testzeitpunkte festgelegt. T1, T4 und T7 wurden unmittelbar nach der erforderlichen Aufgabe, in Abständen von drei Tagen (drei verschiedene Stufen von Aufgaben), erhoben. Die anderen Erhebungen erfolgten bei Bedarf remote. Es handelte sich dabei um retrospektivische Befragungen in Abständen von einem Tag. Den letzten Messzeitpunkt (T10) bildete ein Follow-Up-Test. Die Messung der Usability erfolgte über den *SUS*. Insgesamt nahm die Bewertung der Benutzbarkeit von der ersten bis zur dritten Stufe (T1 zu T7) zu. Die Ergebnisse entsprechen damit denen von Orfanou et al. (2015). Der Effekt nahm jedoch mit zunehmender Einsatzdauer ab, was sich durch eine abflachende Lernkurve erklären lässt (Wang, 2021). Die Ergebnisse der retrospektivischen Befragungen (T2, T3, T5, T6, T8, T9) fluktuieren teilweise stark, sodass es manchmal zu einer Abnahme der Usability, im Vergleich zur vorherigen Einschätzung, kam. Das betraf insbesondere die Messzeitpunkte unmittelbar nach Erreichen der nächsten Aufgabenstufe (T5 und T8). Als Grund

dafür wird der „recency effect“ (Wang, 2021, S. 233) aufgeführt, der auf Hassenzahl und Sandweg (2004) zurückgeht. Hierbei bewerten Nutzer:innen ihre aktuellste und nicht die gesamte Erfahrung mit dem Produkt. Ein weiterer Grund liegt womöglich an der „peak-end rule“ (Wang, 2021, S. 233), wodurch Erinnerungen, die einen erreichten Meilenstein betreffen, besser gespeichert werden. Die Befragung zum Messzeitpunkt T10, einen Monat später, kann als ein Durchschnittswert der gesamten Erfahrung angesehen werden, da das jüngste Erlebnis nicht mehr erinnert werden kann. Dabei entfällt der Recency Effect.

Eine weitere Studie von Lee und Ha (2019) zeigt tendenziell die gleichen Ergebnisse in Bezug auf die Zunahme der Usabilitybewertung über die Zeit. Sie untersuchten allerdings einen differenzierteren Ansatz, indem sie die Usability und die Ästhetik der Anwendung, eine Webseite über Filme, variierten und über fünf Messzeitpunkte innerhalb einer Nutzungsperiode erfassten. Sie wählten dazu ein *2x2x5 mixed-Design*. Die Usability und die Ästhetik waren die *between-subjects*-Variablen mit zwei Ausprägungen (niedrig und hoch) und die Zeit mit fünf Zeitpunkten die *within-subjects*-Variable. Die Usability wurde über den *PSSUQ*, einen standardisierten Fragebogen, erfasst. An der Studie haben 64 Studierende (weiblich: 28, männlich: 36; Durchschnittsalter: 22.75) teilgenommen, die im Schnitt 3.5 Stunden am Tag vor dem Computer verbrachten. Dabei führte eine wiederholte Nutzung der Webseite zu einer höheren Bewertung der Usability. Dieser Effekt war allerdings nur signifikant für den Untersuchungsansatz mit der Webseite, die eine geringe Usability und geringe Ästhetik aufwies. Gleiches Ergebnis gilt für die Zufriedenheit in Bezug auf die Nutzung. Als Limitation ist bei dieser Studie die Einfachheit der Aufgabe, eine Recherche über Filmtitel, aufzuführen. Die Effekte sollten daher für komplexere Aufgaben getestet werden (Lee & Ha, 2019).

Eine weitere Studie von Yuhui und Kolleg:innen (2019) untersuchte den Einfluss des Messzeitpunktes auf die Usability genauer, indem sie die Usability zweier Produkte, eines mit einer geringen und eines mit einer hohen Usability, in bestimmten zeitlichen Abständen untersuchten. Insgesamt nahmen an der Studie 44 Personen im Alter von 19 bis 28 Jahren teil, die mindestens zwei Jahre Erfahrung im Design-Bereich hatten. Die Usability wurde über den *SUS* erhoben. Bei den untersuchten Anwendungen handelt es sich um Videoeditoren (Premiere CS6: geringe Usability, iJianJi: hohe Usability). Die Messzeitpunkte umfassen T1 (Ende der Aufgabe), T2 (fünf Minuten nach T1), T3 (eine Stunde nach T1), T4 (24 Stunden nach T1) und T5 (sieben Tage nach T1). Die Autor:innen konnten Unterschiede zwischen den beiden Anwendungen und zwischen den Erhebungen

feststellen, jedoch keinen signifikanten Einfluss des Haupteffektes „Zeit“ (T1 – T5). Bei Betrachtung der beiden ersten Messzeitpunkte (T1 zu T2) zeigten sich signifikante Ergebnisse in der Bewertung der Usability. Während bei der Verwendung des Programms mit der geringen Usability (Premiere CS6) die Benutzbarkeit abnahm, stieg sie bei Verwendung des besser benutzbaren Programms (iJianJi). Im weiteren Verlauf der Messzeitpunkte nimmt jedoch die Usability bei Messzeitpunkt T3 auf T4 bei iJianJi (hohe Usability) wiederum ab, während sie bei Premiere CS6 leicht zunimmt. Die Autor:innen schließen aus ihrer Studie, dass die Usabilitybewertung einer Anwendung in einem engen Zeitfenster nach der Nutzung erhoben werden sollte. Zudem sollte der Messzeitpunkt bei verschiedenen Nutzer:innengruppen konstant gehalten werden (Yuhui et al., 2019). Bei der Studie ist allerdings als Limitation anzuführen, dass die Nutzer:innen das Produkt lediglich einmalig verwendeten und bereits ausgeprägte Erfahrungen im Aufgabenbereich aufwiesen. Die eben berichteten Ergebnisse konnten bereits Kortum und Johnson (2013) aufzeigen, indem sie ebenfalls zwei Anwendungen mit verschieden guter Benutzbarkeit (hoch: MS Publisher, niedrig: MATLAB) untersuchten. Während bei MS Publisher die Usability über drei Messzeitpunkte kontinuierlich stieg, sank sie bei Verwendung von MATLAB von T1 zu T2, um bei T3 wieder das Ausgangsniveau zu erreichen. Damit zeigt sich eine Abhängigkeit des Messzeitpunktes auf die Usability des Produkts.

Ein weiterer Faktor für die Bewertung der Usability ist das Geschlecht. Teilweise bestehen in der Konzeption von Webseiten spezielle Attribute, die zwischen den Geschlechtern unterschiedlich wahrgenommen werden (Barnum, 2021, S. 117). Unterschiede bei der Bewertung der Usability zwischen Frauen und Männern konnten bereits in der Studie von Karapanos et al. (2018) aufgezeigt werden. Als Gründe wurden die unterschiedlich stark ausgeprägten computerbezogenen Kompetenzen herangezogen. Zu einem ähnlichen Schluss kam eine weitere Untersuchung, in der insbesondere der Einfluss des Geschlechts auf die Bewertung der Usability einer VR-Umgebung untersucht wurde (Schmidt et al., 2013). An der Studie nahmen 28 Psychologiestudierende teil (weiblich: 14, männlich: 14; Durchschnittsalter: 24.93). Die Teilnehmenden wurden zufällig auf zwei verschiedene VR-Umgebungen (sonnig, regnerisch) verteilt. Sie konnten sich nach einer kurzen Einweisung für fünf Minuten frei darin bewegen. Danach wurde ein Fragebogen erhoben, der unter anderem die wahrgenommene Nützlichkeit, die Einfachheit der Nutzung, die Immersion und die technologiebezogene Ängstlichkeit erfasste. Der Effekt der VR-Umgebung wird an dieser Stelle nicht weiter erläutert. Es wurden signifikante Unterschiede in der wahrgenommenen Immersion und der technologie-bezogenen Ängstlichkeit

zwischen Männern und Frauen gemessen. Frauen zeigten eine höhere Immersion und technologiebezogene Ängstlichkeit als Männer. Gleichzeitig erachteten Männer die Technologie als nützlicher als ihre Kolleginnen. Ähnliche Ergebnisse in Bezug auf die technologiebezogene Ängstlichkeit konnten auch schon von anderen Autor:innen festgestellt werden (Schumacher & Morahan-Martin, 2001). Bestärkt werden diese Ergebnisse durch eine Erhebung von Ong & Lai (2004). In einer Studie mit Angestellten (weiblich: 67, männlich: 89; Durchschnittsalter: 30.8) großer Software- und Hardwareunternehmen untersuchten sie die Bewertung des firmeninternen E-Learningkurses unter anderem in Hinblick auf die Selbstwirksamkeit, Nützlichkeit und Einfachheit der Nutzung. Es zeigten sich bei allen erhobenen Variablen signifikante Unterschiede zwischen den Geschlechtern. Männer bewerteten die Usability der Kurse höher und zeigten eine höhere Selbstwirksamkeit. Die Autor:innen sahen insbesondere die computerbezogene Selbstwirksamkeit als eine wichtige Einflussgröße auf die Nützlichkeit und die Einfachheit der Nutzung einer Anwendung (Ong & Lai, 2004). Auf der anderen Seite konnten Orfanou und Kolleg:innen (2015) in der bereits vorgestellten Studie keine Unterschiede zwischen den Geschlechtern in Bezug auf die Usabilitybewertung feststellen. Auch andere Autor:innen konnten keine Unterschiede in der Wahrnehmung der Benutzbarkeit einer Anwendung zwischen den Geschlechtern feststellen (Billestrup et al., 2016; Ibili & Billingham, 2019).

Insgesamt zeigen sich verschiedene Einflussfaktoren auf die Messung der Usability, wie das Geschlecht, der Messzeitpunkt oder der kulturelle Hintergrund. Oftmals werden die Erhebungen mit Proband:innen durchgeführt, die zwar einen aufgabenspezifischen Bildungshintergrund haben, allerdings nicht im Bereich der Usabilityforschung tätig sind. Es stellt sich daher die Frage, ob solche Personen geeignet sind, um Usabilityprobleme und -evaluationen durchzuführen. Eine Studie von Ssemugabi und Villiers (2007) untersuchte diese Fragestellung genauer. Mit 64 Studierenden der Informatik untersuchten sie die Identifikation verschiedener Usabilityheuristiken. Die Gruppe der Experten wurde von vier Lehrenden auf diesem Themengebiet gebildet. Untersucht wurde ein Kurssystem der Universität (Info3Net), das den Proband:innen bekannt war. Insgesamt zeigte sich, dass die Studierenden, die keinen Hintergrund in der Usabilityforschung aufwiesen, 92 Prozent der großen Usabilityprobleme identifizieren konnten (Experten: 92 %). Kleinere Schwierigkeiten wurden zu 43 Prozent identifiziert (Experten: 64 %). Insgesamt wurden somit 73 Prozent der Probleme durch Novizen erkannt (Experten: 77 %), womit auch

Lernende, die ein Programm im Lernprozess nutzen, zur Evaluation der Benutzbarkeit der Anwendung herangezogen werden können (Ssemugabi & Villiers, 2007).

### **3.3 Einsatz von Lernhilfen und Lernunterstützungen**

Zum aktuellen Zeitpunkt existieren, im Vergleich zu den anderen Forschungsthemen dieser Arbeit, im Bereich der Binnendifferenzierung mit Lernhilfen, unter Berücksichtigung der Nutzungsfrequenz und Motivation, bisher wenig Studien. Obwohl viele verschiedene Methoden der *Differentiated Instruction* existieren, gibt es nur wenige hochqualifizierte Studien (Smale-Jacobse et al., 2019). Insbesondere der isolierte Einsatz von einfachen, nicht-gestuftem Lernhilfen wurde bisher wenig erforscht. Zu einem ähnlichen Schluss kommen Pozas und Schneider (2019), die darauf verweisen, dass hauptsächlich Effekte in Bezug auf behinderte Schüler:innen, beschreibende Veröffentlichungen oder theoretisch konzeptionelle Ansätze beschrieben werden. Im Folgenden wird daher der Literaturüberblick geweitet und in der Breite diskutiert. Es werden auch Studien einbezogen, die keinen digitalen Schwerpunkt aufweisen oder gestufte Lernhilfen erforschen.

#### **3.3.1 Wirksamkeit von Lernhilfen**

Speziell für den Einsatz in Schülerlaboren wurde der Einsatz von gestuften Tippkarten (Erklärung, Experimentiertipps und sprachliche Hilfe) erprobt (Affeldt et al. 2018). Die Methode wurde bisher lediglich deskriptiv beschrieben. Von 1032 befragten Schüler:innen geben ca. 80 Prozent an, dass sie die Tippkarten als hilfreich erachteten. Die Autor:innen der Studie legen nahe, dass das Konzept auch mit multimedialen Hilfekarten (z. B. Wolff & Wolff, 2016) verknüpft werden könnte. Bei diesem Konzept können mittels QR-Code gestufte multimediale Hilfekarten über das digitale Endgerät geöffnet werden. In einer ebenfalls deskriptiven Ergebnisdarstellung eines Biologie Leistungskurses zeigte sich ebenfalls ein positives Bild gegenüber der eingesetzten Methode (Wolff & Wolff, 2016). Generell werden Lernhilfen im Chemieunterricht beziehungsweise naturwissenschaftlichen Unterricht häufig beim Experimentieren eingesetzt. Bei einer Untersuchung von Großmann und Woest (2014, 2015) wurden sogenannte unstrukturierte Lernhilfen beim Lösen von theoretischen Aufgaben und Experimentieraufgaben eingesetzt. Die Hilfen boten dabei hauptsächlich inhaltliche Unterstützungen. Dabei konnte gezeigt werden, dass die Hilfen vermehrt bei den theoretischen Aufgaben eingesetzt wurden. Zudem wurden die Hilfen insbesondere bei Berechnungen verwendet. Zudem sollte der Einsatz der

Lernhilfen über einen längeren Zeitraum eingeführt und erprobt werden, da die Schüler:innen die Hilfen nicht direkt zielführend zur Auswertung der Aufgaben heranziehen. Die Autor:innen konnten weiterhin zeigen, dass Lernende mit einem positiven Selbstkonzept seltener zu einer Hilfe griffen als jene, mit einem geringeren Selbstkonzept. Insbesondere leistungsschwächere Schüler:innen verwendeten Hilfen als Kopiervorlage für die eigene Lösung der Aufgaben. Ein Zusammenhang zwischen der Motivation und der Nutzung der Lernhilfen konnte nicht eindeutig belegt werden. Zum Teil hängt die Nutzung der Lernhilfen mit tagesabhängigen Faktoren zusammen (Großmann & Woest, 2014; Großmann & Woest, 2015).

Auch Arnold et al. (2016) haben den Einsatz von Lernhilfen beim Experimentieren erforscht. In einer 2x2 quasi-experimentellen Studie untersuchten die Autor:innen den Einfluss von verschiedenen *Scaffolds* auf die Experimentierkompetenz von 220 Schüler:innen der 11. Jahrgangsstufe in einer Unterrichtsreihe zum Thema Enzyme. Als Lernunterstützung wurden gestufte Lernhilfen, sogenannte Forschertipps, und Concept Cartoons verwendet. Die Hilfen wurden den Schülergruppen (2 bis 3 Personen) in Umschlägen zur Verfügung gestellt. Als abhängige Variablen erhoben die Autor:innen das Methodenwissen, das Fachwissen, die kognitive Belastung und das wissenschaftliche Denken. Die Lernenden arbeiteten über 8 Wochen im Regelunterricht mit den verschiedenen Scaffolds. Der Unterricht wurde von den Lehrkräften durchgeführt. In den beiden Interventionen „Forschertipps“ und „Concept Cartoons“ konnte ein signifikanter Zuwachs beim wissenschaftlichen Denken festgestellt werden. In der Doppelintervention (Forschertipps und Concept Cartoons) blieb der Zuwachs aus. Weitere signifikante Unterschiede konnten nicht festgestellt werden. Tendenziell schienen jedoch die Concept Cartoons das Fachwissen und die Forschertipps das Methodenwissen zu fördern. Die Ergebnisse zeigten zudem, dass alle Interventionsgruppen eine signifikante Verringerung der kognitiven Last gegenüber der Vergleichsgruppe berichteten (Arnold et al., 2016).

Eine weitere Studie mit 63 Studierenden (weiblich: 46) des Lehramts untersuchte ebenfalls in einem 2x2 quasi-experimentellen Design den Einfluss von metakognitiven und multimedialen Unterstützungen auf die Experimentierkompetenz (Bruckermann et al., 2017). Alle vier Gruppen erhielten Unterstützung in Form von kognitiven Hilfekarten. Die Interventionsgruppen wurden zusätzlich durch Einführung einer metakognitiven Strategie, durch multimediale Unterstützung (Dokumentation des Experimentierprozesses mittels eines Videos) oder durch eine Kombination aus den eben genannten

Unterstützungen gefördert. Die Performanz wurde in Form eines qualitativen Beobachtungsprozesses gemessen und ausgewertet. Die Vergleichsgruppe zeigte im Vergleich zu allen drei Interventionsgruppen den größten Kompetenzzuwachs. Die Autor:innen erklärten den ausbleibenden Zuwachs gegenüber der Vergleichsgruppe insbesondere über den zusätzlichen Cognitive Load, der durch die Intervention verursacht wurde. Im Falle der multimedialen Unterstützung stieg der Extraneous Load an, wodurch die Kapazität des Arbeitsgedächtnisses verringert wird. Bruckermann und Kolleg:innen schlossen aus ihrer Studie, dass eine Kombination aus mehreren Unterstützungsmöglichkeiten nicht zwangsläufig zu einem erhöhten Lernzuwachs führt. Der Cognitive Load wurde allerdings nicht durch ein Messinstrument erhoben (Bruckermann et al., 2017).

Zwei weitere Studien, die zum Zeitpunkt des eigenen Studiendesigns noch nicht zur Verfügung standen, sollen im Folgenden dargestellt werden. Sie zeigen die Relevanz des Themas für diese Dissertation. Zwei der Studien beschäftigen sich ebenfalls mit dem Experimentierprozess im Biologieunterricht. In einer Studie von Kleinert und Kolleg:innen (2021) mit 35 Schüler:innen (weiblich: 15, männlich: 20; Durchschnittsalter: 12.50) der Gesamt- und Realschule wurde der Einfluss der Nutzung von gestuften Lernhilfen auf die Experimentierkompetenz mittels qualitativer Inhaltsanalysen untersucht. Die Autor:innen konnten feststellen, dass leistungsstärkere Schüler:innen seltener zu einer Hilfe griffen. Zudem nutzten die meisten Lernenden die Musterlösung zur Korrektur und Überprüfung der eigenen Lösung, während allerdings auch einzelne Schüler:innen direkt zur Musterlösung griffen, um den Prozess des Experimentierens zu verkürzen und schneller zu einer Vollendung der Aufgabe zu gelangen. Als Limitation der Studie wurde die fehlende Kontrolle des genauen quantitativen Nutzungsverhaltens der Lernunterstützungen genannt (Kleinert et al., 2021).

Eine weitere Feldstudie von Kleinert und Kolleg:innen (2022) untersuchte den Einfluss von gestuften Lernhilfen auf motivationale Faktoren der Schüler:innen ( $n = 75$ ; weiblich: 55,2 %; Durchschnittsalter: 16.50). Dabei wurden konkret folgende Variablen erhoben: Die wahrgenommene Kompetenz und das Interesse/Vergnügen (*Kurzskala Intrinsischer Motivation*) sowie als Kovariate der Selbstbestimmungsindex (auf Grundlage der *Skalen zur motivationalen Regulation beim Lernen*). Die Studie setzte sich aus drei Gruppen zusammen: Kontrollgruppe (keine gestuften Lernhilfen), Experimentalgruppe I (gestufte Lernhilfen im Naturwissenschaftsunterricht) und Experimentalgruppe II (gestufte Lernhilfen im Naturwissenschaftsunterricht und Mathematikunterricht). Die Inhalte

(Naturwissenschaftsunterricht: Osmose, Mathematikunterricht: Lineare Funktionen) wurden in acht bis zehn Unterrichtsstunden erarbeitet. Für den Mathematikunterricht wurden die aus dem Naturwissenschaftsunterricht gewonnenen experimentellen Daten verwendet. Die Daten wurden mittels ANCOVA analysiert. Der Effekt der Kovariate erreichte lediglich in der Unterrichtseinheit für den Naturwissenschaftsunterricht auf die Variable „Wahrgenommene Kompetenz“ keine Signifikanz ( $p = 0.051$ ). Zwischensubjekteffekte zeigten in der Unterrichtseinheit für den Mathematikunterricht bei der wahrgenommenen Kompetenz einen signifikanten Unterschied ( $p = 0.048$ ). Es zeigten sich jedoch zum Teil deskriptive Unterschiede zwischen den Gruppen. In Experimentalgruppe I erreichten die wahrgenommene Kompetenz und das Interesse in beiden Fächern das höchste Niveau, während das Interesse in Experimentalgruppe II in beiden Fächern unter dem der Kontrollgruppe lag. Die Autor:innen begründeten dieses Ergebnis mit der Cognitive-Load-Theorie. Die gestuften Hilfen verringerten in der Experimentalgruppe I die kognitive Last, wodurch das Kompetenzerleben und das Interesse stiegen. In Experimentalgruppe II hingegen wurde die Komplexität der Unterrichtsreihe durch die zusätzlichen gestuften Hilfen im Mathematikunterricht verringert, wodurch der Anspruch des Unterrichts und damit das Interesse sanken (Kleinert et al., 2022). Die Forscher:innen legen nahe, in weiteren Studien die Hilfenutzung zu erfassen und in getrennten Untersuchungsdesigns den Einfluss von Lernhilfen auf die Motivation zu untersuchen.

Mit den dargestellten Studien zeigt sich zunächst die Wirksamkeit von Lernhilfen für den Unterricht. Zum Teil können die Effekte durch motivationale und lernpsychologische Faktoren (z. B. Cognitive Load) erklärt werden. Andere Faktoren, wie beispielsweise das Geschlecht oder der Leistungsstand der Lernenden wurden noch nicht dargestellt. Der Einfluss dieser Variablen wird häufiger bei Studien im Bereich des Help-Seeking untersucht.

#### **3.3.2 Help-Seeking**

In Verbindung mit dem Einsatz von Lernhilfen steht auch immer die Fähigkeit der Lernenden, Hilfen adäquat zu beanspruchen. Dieses Verhalten wird unter dem Begriff des Help-Seeking zusammengefasst. Es existieren verschiedene Studien zu diesem Thema, die im Folgenden dargestellt werden. Insbesondere mit dem zunehmenden Einsatz von computerbasierten interaktiven Lernumgebungen gewinnt dieses Forschungsbiet zunehmend an Relevanz.



Es ist bekannt, dass Lernende sich in der Methode der Hilfesuche unterscheiden. Während manche keine Hilfe nutzen, verwenden andere die Unterstützung als direkten Lösungsansatz (Vaessen et al., 2014).

Mit dem Zusammenhang zwischen den Zielen von Lernenden und der Nutzung von Hilfen haben sich in der Vergangenheit bereits andere Autor:innen beschäftigt. In einer Studie von Alevan und Koedinger (2000) wurde der Einsatz eines intelligenten Tutoringsystems im Rahmen des Geometrieunterrichts (Lernzeit insgesamt circa 500 Minuten) mit 41 Schüler:innen (9. Klasse) evaluiert. Die Autor:innen interessierte insbesondere die Nutzung von Hilfen. Dazu konnten die Lernenden auf ein Glossar zurückgreifen (unintelligente Hilfe) oder gestufte Hilfen durch den Tutor (intelligente Hilfe) erhalten. Während das Glossar mathematisches Fachwissen bereitstellte, gab die intelligente Hilfe Unterstützung in Form von metakognitiven Lösungsstrategien. Die Lernenden mussten sich in beiden Fällen aktiv für das Ersuchen der Hilfe entscheiden. Es zeigte sich, dass die Schüler:innen das Glossar selten nutzten, auch wenn sie dadurch mehr Fehler verursachten. Der Grund dafür lag einerseits in den fehlenden metakognitiven Fähigkeiten des Hilfeaufsuchens und in der Fehleinschätzung der Aufgabenschwierigkeit. Ein weiteres Problem könnte im Verständnis des Glossars selbst liegen, in dem grundlegendes mathematisches Fachwissen bereitgestellt wurde. In Bezug auf die intelligenten Hilfen zeigt sich ein ähnliches Verhalten. Zwar nutzten die Schüler:innen die intelligenten Hilfen öfter als das Glossar, allerdings ebenfalls erst nach mehreren falschen Lösungseingaben. Zusätzlich klickten sich die Lernenden direkt zu der obersten Stufe der Hilfe durch und nutzten seltener die Zwischenhilfen. Insgesamt sind für eine erfolgreiche Nutzung von Lernhilfen sowohl ein bestimmtes fachwissenschaftliches Grundverständnis als auch metakognitive Strategien des Aufsuchens von Hilfen notwendig (Alevan & Koedinger, 2000). In weiteren Studien mit dem intelligenten Tutoringsystem konnten Alevan et al. (2006) feststellen, dass leistungsschwächere Schüler:innen häufiger zu einer Hilfe griffen. Gestützt wird diese Erkenntnis beispielsweise von Newman und Schwager (1995), die ebenfalls in ihrer Studie ( $n = 118$ ) mit Drittklässler:innen (weiblich: 41, männlich 18; Durchschnittsalter: 9) und Fünftklässlern (weiblich: 30, männlich: 29; Durchschnittsalter: 12.10) feststellten, dass leistungsschwächere Schüler:innen häufiger eine Hilfe in Anspruch nahmen. Die Intervention fand über drei Tage statt, in denen die Lernenden Mathematikaufgaben lösten. Dabei konnten lediglich beim dritten Tag Fragen gestellt und damit eine Hilfe in Anspruch genommen werden. An diesem Tag fand ebenfalls eine Differenzierung der Probanden in eine Gruppe mit *Learning Goals* und *Performance Goals* statt. Die

Ergebnisse zeigten, dass die Fähigkeit, nach einer gezielten Hilfe zu fragen, mit dem Alter stieg. Zudem erfragten die Kinder mit Learning Goals häufiger eine Rückmeldung zu ihren Aufgaben als die andere Gruppe (Newman & Schwager, 1995).

Auch in einer Erhebung von Bartholomé und Kolleg:innen (2006) mit 74 Studierenden (weiblich: 58, männlich: 14, keine Angabe: 2; Durchschnittsalter: 21.25) zeigten sich ähnliche Ergebnisse. Die Lernenden nutzten in Zweiergruppen eine Computerlernumgebung über vier Kurssitzungen, um Pflanzen zu bestimmen. Dabei konnten sie über einen Button in jedem Bestimmungsschritt Hilfen anfordern und zu jeder Zeit auf ein Glossar zurückgreifen. Über Log-Dateien konnten die Performanz und das Help-Seeking-Verhalten ausgelesen werden. Die Analyse der Daten zeigte einen positiven Zusammenhang zwischen dem Help-Seeking und der Richtigkeit der Antworten ( $r = 0.53$ ,  $p = 0.01$ ) und einen negativen Zusammenhang zu den Folgefehlern in den weiteren Pflanzenbestimmungen ( $r = 0.40$ ,  $p = 0.013$ ). Die Zusammenhänge erwiesen sich als nicht signifikant für die Nutzung des Glossars. In Bezug auf das Vorwissen nutzten Studierende mit einem geringeren Vorwissen häufiger eine Hilfe und erzielten am Ende bessere Ergebnisse als Studierende mit hohem Vorwissen. Signifikante Abhängigkeiten zwischen dem Interesse und dem Nutzen einer Hilfe wurden nicht gefunden. Für weitere Studien sollen insbesondere umfangreichere Log-Dateien in Kombination mit weiteren Daten in größeren Erhebungsgruppen gesammelt werden (Bartholomé et al., 2006).

Die Ergebnisse der genannten Arbeitsgruppen stehen in Widerspruch zu den Ergebnissen anderer Autor:innen, wie zum Beispiel Wood und Wood (1999). Die Gründe dafür, in Bezug auf die Untersuchung von Wood und Wood, könnten unter anderem in den verschiedenen eingesetzten Tutoringsystemen, der Umgebung (Klassenraum gegen Labor) und in der Dauer der Studie liegen (Aleven et al., 2006). In Bezug auf andere Studien, die im weiteren Verlauf dargestellt werden, könnte der Grund auch in der Quelle der Hilfe liegen (Beal et al., 2008), sodass in computergestützten Lernumgebungen mit virtuellen Lernhilfen ein anderes Verhalten in der Nutzung von Unterstützungen eintritt als in dem regulären unterrichtsgesprächsfokussierten Unterricht. In der Untersuchung ( $n = 90$ , weiblich: 48, männlich: 42) von Beal und Kolleg:innen (2008) konnte in drei Highschools ebenfalls festgestellt werden, dass schwächere Schüler:innen häufiger eine Hilfe nutzten als leistungsstarke. Zu Anwendung kam dabei ein Tutoringsystem zur Vermittlung von Geometrie. Gestützt wird die Hypothese von Beal et al. (2008) und unter anderem von einer älteren Studie von Karabenick und Knapp (1988). Dabei mussten 28

Psychologiestudent:innen (weiblich: 12, männlich: 16) eine fiktive Begriffsbildungsaufgabe an einem Computer lösen. Nach jeder falschen Eingabe konnten die Studierenden eine Hilfe anfordern. Die Individuen wurden randomisiert auf zwei Gruppen verteilt. Eine Gruppe erhielt die Hilfe direkt von dem Computer, während für die andere Gruppe zwar ebenfalls Hilfe über den Computer beanspruchte, jedoch dabei eine Konversation mit einer anderen Person über eine zeitliche Verzögerung von 20 Sekunden simuliert wurde. Die Simulation wurde von 12 aus 14 Teilnehmenden unwissentlich als real erachtet. Unterschiede zwischen den Geschlechtern konnten in Blick auf das Help-Seeking nicht festgestellt werden. Zwischen den beiden Gruppen (Computerhilfe und menschliche Hilfe) lag ein signifikanter Unterschied ( $t(26) = 2.45, p < 0.01$ ). Dabei nutzen 86 Prozent der Computergruppe eine Hilfe, während es in der anderen Gruppe lediglich 36 Prozent taten. Die durchschnittliche Anzahl der angeforderten Hilfen betrug 3.50 (PC) und 0.80 (Person). Gründe<sup>3</sup> für das Auslassen jeglicher Nutzung von Hilfen bei den Proband:innen waren insbesondere das selbstständige Problemlösen ( $M = 6.10$ ), Fehleinschätzung der Aufgabenschwierigkeit, ( $M = 3.30$ ) eine längere Bearbeitung der Aufgabe bei Nutzung einer Hilfe ( $M = 2.40$ ), Angst vor Herausfinden der Hilfenutzung durch andere ( $M = 1.50$ ) und das Gefühl von Versagen bei Nutzung einer Hilfe ( $M = 1.50$ ). Folglich waren in dieser Situation keine beziehungsweise nur geringe Gründe in Bezug auf soziale Ziele vertreten. Dabei besteht jedoch ein Widerspruch zu den Ergebnissen der Studie in Bezug auf die beiden Bedingungen. Womöglich gaben die Teilnehmenden bewusst andere Gründe an oder waren sich der genauen Situation der Untersuchungsbedingungen im Nachhinein bei Beantwortung der Fragen nicht mehr bewusst (Karabenick & Knapp, 1988).

An der bereits genannten Studie von Wood und Wood (1999) nahmen 42 Schüler:innen im Alter von 14 bis 15 Jahren teil. Die Lernenden wurden in zwei Gruppen, eine leistungsstarke ( $n = 22$ ) und eine leistungsschwache ( $n = 20$ ), unterteilt. Im Rahmen von zwei 30- bis 50-minütigen Lerneinheiten zum Thema der quadratischen Funktion wurde die Hilfenutzung in einem intelligenten Tutoringsystem untersucht. Dabei zeigte sich, dass, nach einem Fehler, leistungsstärkere Schüler:innen häufiger zu einer Hilfe griffen als leistungsschwächere. Insgesamt liegt jedoch in beiden Gruppen ein positiver Zusammenhang zwischen der Nutzung der Lernhilfen und der Lösungswahrscheinlichkeit einer Aufgabe vor, wobei dieser Effekt bei den leistungsschwächeren Schüler:innen stärker ausgeprägt

---

<sup>3</sup> Skala vom Typ „Likert“ 0 bis 9

war (Wood & Wood, 1999). Zu einem ähnlichen Ergebnis in Bezug auf die Nutzung von Lernhilfen und den Leistungsstand der Lernenden kamen Ryan et al. (1997). In der Befragung mit 443 Fünftklässler:innen (weiblich: 231, männlich: 212) untersuchten die Autor:innen den Einfluss von sozialen und akademischen Zielen auf die Vermeidung von Help-Seeking. Die Schüler:innen wurden dazu unmittelbar in den Klassen über Fragebögen befragt. Anschließend wurde eine Regressionsanalyse durchgeführt. Die Ergebnisse zeigten unter anderem nichtsignifikante deskriptive Unterschiede zwischen den Geschlechtern. Weiterhin zeigten leistungsschwächere Schüler:innen häufiger die Tendenz, Hilfen nicht zu nutzen ( $\beta = -0.24, p < 0.001$ ). Bei Lernenden mit starken Statuszielen (*Social Goals*) zeigte sich ebenfalls ein signifikanter Einfluss auf das Vermeiden von Hilfenutzung ( $\beta = 0.11, p < 0.05$ ). Verstärkt wurde dieser Effekt mit dem Eintritt in die Pubertät. Unter Kontrolle der akademischen und sozialen Ziele zeigte sich ein signifikanter Einfluss des Geschlechts ( $\beta = 0.12, p < 0.05$ ). So sehen Schülerinnen häufiger eine Gefahr in der Nutzung von Hilfen als Schüler. Als Limitationen bei dieser Studie sind unter anderem der einzige Erhebungszeitpunkt, das Verwenden von Daten durch Selbsteinschätzung und die fehlende Information über den Kontext der Hilfenutzung aufzuführen (Ryan et al., 1997). Gerade in Bezug auf die sozialen Ziele ist die Bereitstellung der Lernhilfen von großer Bedeutung. Beispielsweise erfordert das Fragen einer Mitschülerin oder eines Mitschülers weniger Aufwand und Offenbarung, im Sinne der Notwendigkeit der Nutzung einer Hilfe, als das Holen einer Hilfekarte vom Pult oder das Fragen der Lehrkraft im Plenum. Die Ergebnisse wurden jedoch durch eine weitere Studie von Ryan und Shin (2011) gestützt. In dieser Untersuchung wurden 217 Sechstklässler (weiblich: 47 %) und deren jeweiligen Lehrkräfte, zwei pro Klasse beziehungsweise Individuum befragt. Die Lehrer:innen gaben Auskunft über das Help-Seeking-Verhalten der Schüler:innen. Dabei wurde in adaptives / angemessenes (*adaptive / appropriate*), vermeidendes (*avoidant*) und abhängiges (*dependent*) Verhalten unterscheiden. Lernende möchten ihre Autonomie größtenteils erhalten. Unter adaptivem Help-Seeking wird das Aufsuchen eines Hinweises anstelle einer Lösung verstanden. Unter avoidant Help-Seeking wird das Vermeiden von Hilfeersuchen verstanden und unter dem abhängigen Help-Seeking das Aufsuchen von Hilfe, sobald eine Schwierigkeit festgestellt wird (Ryan et al., 2005). Zusätzlich wurden über Fragebogen die akademische Selbstwirksamkeit und soziale Ziele erhoben. Weiterhin wurde der Notendurchschnitt im ersten und dritten Quartal des Schuljahres ermittelt. Die Variablen wurden mittels einer Regression analysiert. In den Analysen zeigten sich der Leistungsstand ( $\beta = 0.41, p < 0.001$ ) und die akademische

Selbstwirksamkeit ( $\beta = 0.24, p < 0.001$ ) als positive Prädikatoren für das adaptive Aufsuchen von Hilfe. Damit griffen leistungsstärkere Schüler:innen häufiger zu einer Hilfe in Form eines Hinweises, der zur weiteren Bearbeitung der Aufgabe dient. Die sozialen Ziele zeigten sich als schwacher, aber signifikanter negativer Prädiktor für das Aufsuchen von Hilfen ( $\beta = -0.13, p < 0.05$ ). Im Umkehrschluss waren die genannten Faktoren Leistungsstand ( $\beta = -0.39, p < 0.001$ ) und die akademische Selbstwirksamkeit ( $\beta = -0.14, p < 0.001$ ) negative Prädikatoren für das Vermeiden der Beanspruchung von Hilfen. Insgesamt hatten leistungsschwächere Schüler:innen größere Probleme zu identifizieren, wann sie eine Hilfe in Anspruch nehmen sollten. In diesem Zusammenhang konnten sie ebenfalls schlechter Fragen an die Lehrkräfte stellen, um eine geeignete Hilfe zu erhalten. Zudem zeigte sich die Lernwirksamkeit von adaptivem Help-Seeking. In Bezug auf die sozialen Ziele konnten die Ergebnisse von Ryan et al. (1997) bestätigt werden (Ryan & Shin, 2011). Auch bei dieser Studie wurde nicht detailliert die Unterstützungsform offengelegt. Es ist anzunehmen, dass sich die Ergebnisse primär auf mündliche Unterstützungen durch die Lehrkraft bezogen.

Neben dem Kontext, in dem der Prozess des Hilfeersuchens stattfindet, existieren weitere Einflussfaktoren auf den Prozess des Nutzens einer Hilfe. Eine Studie von Kessel und Steinmayr (2013) untersuchte den Einfluss des Geschlechts auf den Help-Seeking-Vorgang. An der Befragung nahmen 83 Schülerinnen und 99 Schüler zweier Gymnasien teil (Durchschnittsalter: 16.36). Dabei wurden das Help-Seeking-Verhalten, das Selbstkonzept der Geschlechterrolle, die Intelligenz und die Schulleistung erhoben. Die Befragung dauerte insgesamt 120 Minuten. Dabei zeigten sich in Bezug auf das Ersuchen von Hilfe zwischen Mädchen und Jungen signifikante Unterschiede. Schülerinnen hatten eine höhere Intention als Jungen, Hilfe in Anspruch zu nehmen ( $M_w = 5.53, M_m = 5.04; p = 0.002, \eta^2 = 0.05$ )<sup>4</sup> und zeigten ein geringeres Vermeidungsverhalten in Bezug auf die Nutzung einer Hilfe ( $M_w = 2.18, M_m = 2.54; p = 0.018, \eta^2 = 0.03$ ). Außerdem nahmen die Schülerinnen die wahrgenommenen Kosten als signifikant geringer an. In Bezug auf die Geschlechterrollen zeigte sich generell ein positiver Zusammenhang zwischen der Hilfenutzung und positiver Feminität und ein negativer Zusammenhang zu einer negativen Maskulinität. Neben dem Geschlecht konnten andere Arbeitsgruppen (Ryan et al., 2009) zudem Unterschiede zwischen Jugendlichen mit unterschiedlichem kulturellem

---

<sup>4</sup> Ergebnisse der ANOVA; 7-stufige Likert-Skala

Hintergrund feststellen. Während zwischen den Geschlechtern bei Individuen mit afrikanisch-amerikanischem Hintergrund keine Unterschiede im Vermeiden von Hilfeersuch auftraten, zeigten europäisch-amerikanische Mädchen geringere Vermeidung der Nutzung von Lernhilfen. Die Ergebnisse stehen damit in Einklang mit denen von Kessel und Steinmayr (2013). Zu den gleichen Schlüssen kamen Schenke et al. (2015), die unter anderem auch die Unterschiede zwischen den Geschlechtern in Bezug auf das Help-Seeking untersuchten. An der Erhebung nahmen 8189 Schüler:innen (männlich: 48 %) aus 448 Mathematikklassen teil. Es handelte sich dabei um siebte bis elfte Klassen. In den Klassen lag eine hohe ethnische Diversität vor. Die Befragung wurde zu Beginn und zum Ende des Schuljahres durchgeführt. Zum Einsatz kam ein Fragebogen mit einer fünfstufigen Likert-Skala. Dabei wurde unter anderem das Hilfenutzungsverhalten erhoben. Dabei zeigte sich, dass Jungen zu beiden Testzeitpunkten seltener bei Lehrkräften oder Mitschüler:innen um Hilfe baten. Zudem zeigten sie eher ein Verhalten, das durch ein direktes Erfragen der Antwort geprägt ist (*Expendient Help-Seeking*) und weniger ein Hilfeersuchen im Rahmen der Verständniskerngewinnung (*Instrumental Help-Seeking*). Die Autor:innen verwiesen auf den möglichen Einfluss der kulturellen Hintergründe, auch wenn dieser Forschungsstand nicht explizit in der Studie untersucht wurde (Schenke et al., 2015). In Bezug auf den kulturellen Hintergrund und das Geschlecht fanden weitere Forschungsgruppen Unterschiede, die in großen Teilen die bisher aufgeführten Ergebnisse stützen. Cheong und Kolleg:innen (2004) befragten 314 Schüler:innen (weiblich: 64, männlich: 250) aus der achten bis zur zwölften Klasse in Informatik unter anderem in Bezug auf ihr Help-Seeking-Verhalten. Die Ergebnisse wurden mittels einer hierarchischen Regressionsanalyse ausgewertet. Dabei zeigte sich ein signifikanter Einfluss des Geschlechts in Bezug auf das instrumentelle Hilfeersuchen ( $\beta = -0.12, p < 0.01$ ) und die wahrgenommenen Vorteile ( $\beta = -0.14, p < 0.05$ ). Beide Variablen wurden von Schülerinnen höher eingestuft. In Bezug auf das exekutive Hilfeaufsuchen und das Vermeiden der Hilfenutzung zeigte sich kein signifikanter Einfluss des Geschlechts (Cheong et al., 2004). Die bisherigen Ergebnisse in Bezug auf das Nutzungsverhalten von Hilfen bei Schülerinnen und Schülern werden durch neuere Studien gestützt (z. B. Davison et al., 2022).

Bisher wurden bereits einige Studien vorgestellt, die einen Bezug zwischen den akademischen Zielen und dem Help-Seeking-Verhalten aufzeigen konnten. Im Folgenden werden nun weitere Untersuchungen vorgestellt, die den Einfluss motivationaler Variablen auf den Prozess der Hilfenutzung aufzeigen. Newman (1990) konnte in Bezug auf die motivationalen Voraussetzungen bei Schüler:innen verschiedener Klassenstufen (3, 5 und

7) unterschiedlich starke Zusammenhänge der genannten Variablen feststellen. Die Stichproben setzten sich aus 60 (3. Klasse - weiblich: 25, männlich: 35; Durchschnittsalter: 9.00), 57 (5. Klasse - weiblich: 33, männlich: 24; Durchschnittsalter: 11.10) beziehungsweise 60 Lernenden von Mathematikklassen (7. Klasse - weiblich: 35, männlich: 25; Durchschnittsalter: 13.10) zusammen. Als Variablen wurden unter anderem die positiven und negativen Überzeugungen bezüglich des Help-Seeking im Mathematikunterricht, die intrinsischen und extrinsischen Orientierungen (Vorliebe für Herausforderungen, Neugier / Interesse und selbständige Bewältigung) sowie die wahrgenommene Kompetenz erhoben. Die Daten wurden mittels einer Strukturgleichungsmodellierung untersucht. Die Ergebnisse zeigten, dass die Intention, eine Hilfe zu nutzen von der dritten bis zur siebten Klasse abnahm. In den späteren Klassen bestimmten vor allem die wahrgenommene Kompetenz und die intrinsischen Orientierungen die Hilfenutzung. Der Vorgang wurde zudem durch die positiven und negativen Glaubenssätze gegenüber dem Help-Seeking geprägt (Newman, 1990). Gestützt werden die Ergebnisse von Ryan und Pintrich (1997), die in ihren Studien unter anderem auch einen positiven Einfluss der wahrgenommenen Kompetenz auf die Hilfenutzung bei Siebt- und Achtklässlern fanden. Auf der anderen Seite fand Butler (1998) gemischte Ergebnisse in Bezug auf den Zusammenhang zwischen wahrgenommener Kompetenz und Help-Seeking. Eine Studie mit 437 Studierenden (weiblich: 312, männlich: 125; Durchschnittsalter: 22.54) konnte signifikante Zusammenhänge zwischen motivationalen Variablen und der Hilfenutzung beziehungsweise Vermeidung der Beanspruchung von Hilfen feststellen (Algharaibeh, 2020). Die instrumentelle Hilfenutzung korrelierte positiv mit einem schwachen Effekt mit der intrinsischen Motivation ( $r = 0.291, p < 0.01$ ) und der extrinsischen Motivation ( $r = 0.265, p < 0.01$ ). Das exekutive Help-Seeking korrelierte positiv mit der Vermeidung der Hilfenutzung ( $r = 0.739, p < 0.01$ ) und der Amotivation ( $r = 0.709, p < 0.01$ ). Zudem bestand ein signifikanter Zusammenhang zwischen Amotivation und der Vermeidung der Nutzung von Hilfen ( $r = 0.738, p < 0.01$ ) (Algharaibeh, 2020). Auch Yang und Cao (2013) fanden positive Zusammenhänge zwischen der intrinsischen Motivation und dem Aufsuchen von Lernhilfen. Im Gegensatz zu den eben genannten Studien, untersuchten sie den Help-Seeking-Vorgang im Bereich des E-Learnings. An der Studie nahmen 93 Studierende teil (weiblich: 75%). Die Teilnehmer:innen wurden am Ende des Semesters bezüglich verschiedener motivationaler Variablen und ihrem Help-Seeking-Verhalten befragt. Die Ergebnisse wurden über Korrelations- und Pfadanalysen ausgewertet. Auch hierbei zeigten sich signifikante positive Zusammenhänge zwischen der intrinsischen

( $r = 0.30$ ,  $p < 0.01$ ) sowie extrinsischen Motivation ( $r = 0.44$ ,  $p < 0.01$ ) und dem Help-Seeking. Die Pfadanalysen bestätigten diese Ergebnisse in zwei verschiedenen Modellen. Dabei überwog der Effekt der extrinsischen Motivation leicht (Yang & Cao, 2013). Aktuelle Studien aus anderen Ländern und Fächern zeigten ebenfalls einen Zusammenhang zwischen der intrinsischen Motivation und Help-Seeking (z. B. Averina & Kuswandono, 2022). Eine andere Studie hingegen fand keinen Zusammenhang zwischen den bereits genannten Variablen (Lynch & Dembo, 2004).

Insgesamt bestehen viele Einflussvariablen auf den Prozess des Help-Seeking. Neben dem Geschlecht bestehen zudem weitere Personenmerkmale, die die Nutzung von Lernhilfen beeinflussen. Als weiterer Einflussfaktor hat sich der Kontext der Hilfenutzung erwiesen. In diesem Zusammenhang ist beispielsweise von Relevanz, welche Person um Hilfe gebeten wird (Lehrkräfte / Mitschüler:innen) oder in welcher Form die Hilfe erfolgt (personell / maschinell). In Bezug auf Letzteres konnte beispielsweise eine Studie von Ruf und Ploetzner (2014) nachweisen, dass ein Unterschied in der generellen Bereitstellung von Hilfen in einer multimedialen Lernumgebung bestand. Die Lernhilfen wurden dazu entweder statisch, dynamisch (Hilfe weitet sich nach wenigen Sekunden aus) oder eingeklappt am rechten Bildschirmrand präsentiert. Die Lernenden nutzen im dynamischen Setting die Lernhilfe am häufigsten, gefolgt von der statischen Bereitstellung. Am wenigsten wurden die Lernhilfen im Falle der eingeklappten Integration verwendet. Die Studie verdeutlicht die feinen Nuancen, die im Bereich des Help-Seeking bestehen, so dass neben personellen Merkmalen auch ein starker Einfluss der verwendeten Lernmaterialien besteht.

### **3.4 Zusammenfassende Darstellung der Forschungslücken**

Aus dem aktuellen Forschungsstand können verschiedene Forschungslücken abgeleitet werden. Zudem zeigt sich je nach Forschungsgegenstand eine hohe Relevanz für das Lernen im Unterricht. In den nachfolgenden Unterkapiteln werden daher explizit Forschungslücken und die Relevanz des Untersuchungsgegenstandes aufgeführt. Das Kapitel gliedert sich damit in drei Teile: Motivation und Interesse, Usability und Einsatz von Lernhilfen.



### 3.4.1 Motivations- und interessenförderndes Potential digitaler Medien

Eine Vielzahl von Studien beschäftigte sich bereits mit der Beforschung des Zusammenhangs zwischen der Motivation sowie dem Interesse und dem Einsatz digitaler Medien (Higgings et al., 2019). Ein Großteil dieser Studien zeigt einen positiven Zusammenhang zwischen den Variablen. Zum Teil sind diese Vorteile bisher allerdings nur für einzelne Fächer oder Klassenstufen untersucht. Beispielsweise fokussierten viele Studien den Einsatz im Mathematikunterricht, gerade im Hinblick auf intelligente Tutoringsysteme (Hillmayr et al., 2020). Speziell für den Chemieunterricht zeigen manche Studien allerdings kein einheitliches Ergebnis in Bezug auf das motivationsfördernde Potential von digitalen Medien (Lee & Osmann, 2012; Osman & Lee, 2014; Sary et al., 2018). Auch unter Betrachtung der Geschlechter zeigt sich kein einheitliches Bild (Bruckermann et al., 2016; Conradt & Bogner, 2016; Higgins et al., 2019; Liu et al., 2016; Tossavainen & Hirsto, 2017), was weitere Studien in Hinblick auf diese Faktoren erfordert. Genauso spielt die Art des digitalen Tools eine Rolle (Hillmayr et al., 2020). Aus diesem Grund sollte jedes funktional neue Werkzeug, wie es bei HyperDocSystems der Fall ist, in Hinblick auf diese Variablen untersucht werden. Zwar bieten manche Tools ähnliche Funktionen, allerdings kann es durch zusätzliche Funktionen oder die Art der Implementierung zu einer Veränderung der motivationalen Merkmale einer Person kommen, wenn beispielsweise eine Überlastung des Arbeitsgedächtnisses vorliegt, wie es für manche multimedialen Systeme (Moos & Marroquin, 2010) oder multimediale Lernhilfen (Bruckermann et al., 2017) der Fall ist. Für eine genauere Untersuchung dieser motivationalen Variablen schlagen manche Autor:innen daher auch die stärkere Berücksichtigung von Log-Files vor, die die Interaktion mit multimedialen Inhalten nachvollziehbar machen (Moos & Marroquin, 2010). Im Falle von HyperDocSystems wird die Nutzung der Lernhilfen protokolliert und kann mit den motivationsbezogenen Variablen in Verbindung gebracht werden. Bei vielen der Studien handelte es sich um einzelne Sitzungen, in deren Rahmen der motivationale Einfluss digitaler Werkzeuge untersucht wurde. Daher fordern manche Forscher:innen länger andauernde Studien, um unter anderem auch den Einfluss von Neuheit auf die Lernenden zu untersuchen (Jung et al., 2019; Van Loon et al., 2012; Stoa & Chu, 2023). Dabei bleibt die Wirksamkeit des Neuheitseffektes uneindeutig (González-Cutre et al., 2016; Jenó et al., 2019a; Kopinska, 2020). Allgemein plädieren andere Forschungsgruppen für Studien auf dem Gebiet der digitalen Medien, um den Einfluss dieser auf den Lernprozess zu untersuchen. Dabei wird insbesondere ein Kontrollgruppendesign gefordert. Zudem sollen auch Studien veröffentlicht werden, die keinen oder einen negativen

Zusammenhang zwischen der Motivation und dem Einsatz digitaler Tools zeigen, um der Publikationsverzerrung entgegenzuwirken (Hillmayr et al., 2020).

### 3.4.2 Usability

Unter Berücksichtigung des aktuellen Forschungsstandes zeigen sich verschiedene Forschungslücken und eine hohe Relevanz in Bezug auf die Usability von Softwareanwendungen. Insgesamt existieren wenige Studien im Bereich der Sekundarstufe I und II, die die Usability berücksichtigen oder als expliziten Forschungsschwerpunkt in den Fokus nehmen. Dabei betonen jedoch Simon et al. (2008) und Tselios et al. (2001) die Wichtigkeit der Benutzbarkeit für das Lernen mit digitalen Programmen, da unter anderem ein Einfluss auf die Lernwirksamkeit besteht (Davids et al., 2014; Meiselwitz & Sadera, 2008). In Zusammenhang mit dem Lernen mithilfe von verschiedenen Softwareanwendungen steht der Cognitive Load, da er eine wichtige Einflussgröße auf den Lernprozess darstellt. Insgesamt berücksichtigen jedoch bisher nur wenige Usabilitystudien die kognitive Last bei mobilen Anwendungen (Weichbroth, 2020). Dabei sehen allerdings Karapanos et al. (2018) unter anderem einen Zusammenhang zwischen der Usability und dem Extraneous Cognitive Load. In deren Studie wurde dieser Zusammenhang allerdings nicht näher untersucht. Eine aktuelle Studie von Koć-Januchta und Kolleg:innen (2022) konnte schließlich den negativen Zusammenhang zwischen Extraneous Cognitive Load und Usability aufzeigen. Auch eine andere Studie konnte einen Zusammenhang zwischen der Usability und dem Cognitive Load aufzeigen (Ibili & Billinghamurst, 2019). Damit wird die hohe Relevanz der Betrachtung des Cognitive Load bei der Evaluation von Anwendungen im Bereich des Lernens ersichtlich. Gleichzeitig sind die Ergebnisse von Koć-Januchta und Kolleg:innen (2022) bisher nur für ein AI-gestütztes Textbuch verifiziert, wodurch weitere Untersuchungen mit anderen Arten von digitalen Anwendungen erforderlich sind.

Die bereits aufgeführte Studie von Karapanos et al. (2018) untersuchte weiterhin den Einfluss der Usability auf die Lernwirksamkeit und das Interesse. In der Studie formulieren die Autor:innen als eine Forschungslücke den Einfluss von Persönlichkeitsmerkmalen, wie zum Beispiel die computerbezogene Selbstwirksamkeit und die Computerangst, auf die Usability. Einen ähnlichen Schluss zieht Venkatesh (2000), indem er den Fokus zukünftiger Untersuchungen auf die Erhebung individueller Variablen, wie zum Beispiel die intrinsische Motivation, und die computerbezogene Selbstwirksamkeit richtet. Ebenso betonen Davids und Kolleg:innen (2014) unter anderem die Relevanz der

Motivation und des Spaßes bei weiteren Untersuchungen der Benutzbarkeit von Multimedia-Anwendungen. Hornbæk (2006) zeigt auf, dass für die Erklärung von Usability-messungen auch Korrelationen zwischen den verschiedenen Variablen berechnet werden sollten. Zudem betont er die Notwendigkeit der Untersuchung von komplexeren Anwendungen und longitudinalen Studien, um Grundfragen der Usability zu erforschen. In Bezug auf die Problematik des Testzeitpunktes bei Testungen der Usability zeigen Lee und Ha (2019) auf, dass in weiteren Untersuchungen insbesondere komplexere Anwendungen evaluiert werden sollten. Auch Meiselwitz und Sadera (2008) sowie Orfanou und Kolleg:innen (2015) weisen darauf hin, heterogenere Probandengruppen bei Usabilitystudien von digitalen Lernumgebungen zu berücksichtigen. Andere Autor:innen (z. B. Wallace & Yu, 2009) konnten zudem einen kulturabhängigen Einfluss auf die Usabilitybewertung feststellen, wodurch jede Anwendung in ihrem spezifischen Nutzungsumfeld getestet werden sollte. Zudem zeigt sich in der Studie von Orfanou et al. (2015) in Bezug auf die mehrmalige Testung der Usability ein Widerspruch zu anderen Untersuchungen (z. B. Kortum & Johnson; Yuhui et al., 2019). Insgesamt liegt die Literatur in Bezug auf den Messzeitpunkt kein eindeutiges Ergebnis vor, was eine weitere Untersuchung erfordert. Gleiches gilt für die geschlechtsabhängige Bewertung der Usability, da teilweise Unterschiede zwischen den Studien bestehen und teilweise nicht (Billestrup et al., 2016; Ibili & Billinghurst, 2019; Ong & Lai, 2004; Orfanou et al., 2015).

### **3.4.3 Einsatz von Lernhilfen und Lernunterstützungen**

Zum Teil ist der Einsatz von Scaffolding und spezieller der Einsatz von Lernhilfen in Bezug auf die Lernwirksamkeit im Unterricht hinreichend erprobt (z. B. Arnold et al., 2016). Zu spezielleren Forschungsfragen in Bezug auf die Nutzung von Lernhilfen, beispielsweise die Quantität der Nutzung, konnten bisher nur wenige Studien gefunden werden. Daraus lässt sich ableiten, dass die Nutzung und die Erfassung der Nutzungsgründe von Lernhilfen bisher noch nicht ausreichend erforscht sind. Zu einem ähnlichen Schluss kommen Pozas et al. (2019). Die Wirksamkeit von gestuften Lernhilfen wurde bereits ausführlicher untersucht. Dabei steht jedoch der Kompetenzzuwachs beim Experimentieren im Vordergrund (z. B. Arnold et al. 2016, Stiller & Wilde, 2021). Untersuchungen für den Regelunterricht Chemie, die insbesondere die Nutzung von Lernhilfen zur Erschließung von Fachwissen fokussieren, sind in einer geringeren Anzahl vorhanden. Dabei existieren allerdings verschiedene Forschungslücken anderer Studien, die zu den relevanten Forschungsfragen und Detailfragen der vorliegenden Arbeit in Bezug auf

Lernhilfen geführt haben. In dieser Forschungsarbeit steht allerdings nicht die Lernwirksamkeit im Vordergrund, sondern andere Variablen, wie beispielsweise die Motivation oder der Cognitive Load.

Allgemein betonen verschiedene Autor:innen die Relevanz von Log-Files beziehungsweise Protokollierung zur Erfassung der genauen Hilfenutzung (Bartholomé, 2006; Kleinert et al., 2021; Stiller & Wilde, 2021). Kousa und Aksela (2019) betonen zudem die generelle Relevanz von Werkzeugen, die die Erstellung und den Einsatz differenzierten Unterrichtsmaterials im Fach Chemie ermöglichen. Damit zeigt sich die Bedeutung von HyperDocSystems als Forschungswerkzeug. Zwar existieren bereits schon länger Tutoringsysteme und spezielle Anwendungen für einzelne Fachrichtungen und Themengebiete (z. B. Alevén et al., 2006; Ruf & Ploetzner, 2014), allerdings ist zum jetzigen Zeitpunkt kein allgemeines Tool bekannt, das die Integration von Lernhilfen in digitale Arbeitsblätter plattformübergreifend in einer Anwendung ermöglicht. Insgesamt existieren viele Studien bezüglich des Help-Seeking im Bereich der Mathematik (z. B. Newman, 1990) und wenige im Fach Chemie, wodurch mit dieser Arbeit ein Beitrag für den Chemieunterricht geschaffen werden soll. Auch in Bezug auf die ausgewählten Altersgruppen fokussierten viele Studien eher den Primarbereich (z. B. Newman & Schwager, 1995) oder die universitäre Ausbildung (z. B. Vaessen et al., 2014). Wenige Studien beachteten bisher die Mittelstufe oder Oberstufe. Damit zeigt sich die Relevanz von weiteren Untersuchungen für diese Altersgruppen. Viele Studien befassen sich zudem mit dem Help-Seeking von Schüler:innen in Bezug auf die Lehrkraft oder Mitschüler:innen (z. B. Ryan, 2005). Studien, die explizit die Nutzung von Lernhilfen unter Berücksichtigung verschiedener Variablen untersuchten, konnten kaum gefunden werden. Die meisten dieser Studien existieren für den erwähnten universitären Bereich und beziehen sich auf E-Learningangebote und nicht auf klassisch unterrichtliche Schulstunden im 45-Minuten-Takt. Solche Variablen sind beispielsweise intrinsisch und extrinsisch veranlagte Motive, die zum Teil bereits Berücksichtigung in den Untersuchungen zur Hilfenutzung gefunden haben. Insgesamt bezogen sich diese Studien oft auf akademische Ziele und nicht auf das Interesse oder die intrinsische Motivation. Zum Teil existieren keine eindeutigen Ergebnisse diesbezüglich (z. B. Algharaibeh, 2020; Großmann & Woest, 2015; Lynch & Dembo, 2004). Einige Autor:innen geben daher weitere Desiderata in diesem Bereich (z. B. Kleinert et al., 2022; Stiller & Wilde, 2021). Auch das Vorwissen beziehungsweise die Leistungsstärke von Lernenden haben Einfluss auf deren Hilfenutzungsverhalten. Aber auch auf diesem Teilgebiet der Forschung existieren uneindeutige Erkenntnisse, wie

bereits oben dargestellt (z. B. Alevén et al., 2006; Wood & Wood, 1999). Ein aktuelles Review unterstützt den generellen Forschungsbedarf im Hinblick auf weitere Variablen, die den Prozess des Hilfeersuchens berücksichtigen (Martín-Arbós et al., 2021). In manchen Studien (z. B. Arnold et al., 2016; Dong et al., 2020) konnte ein Zusammenhang zwischen Cognitive Load und Help-Seeking festgestellt werden. In diesem Zusammenhang kann jedoch gerade der Einsatz multimedialer Unterstützungen die kognitive Last erhöhen (Bruckermann et al., 2017), weshalb weitere Studien in diesem Bereich angestellt werden sollten (Mayer et al., 2002). Nicht zuletzt bestehen außerdem Unterschiede zwischen den Geschlechtern, die insbesondere unter verstärkter Berücksichtigung des sozialen Kontexts in Lerngruppen untersucht wurden (z. B. Kessel & Steinmayr, 2013; Ryan et al., 2009). Im Bereich des computergestützten Lernens wurden die Unterschiede bisher kaum erprobt.

### **3.5 Forschungsfragen**

Aus Gründen der Übersicht wurden die Forschungsfragen nicht unmittelbar bei der Darstellung der Forschungslücken beschrieben. Sie werden daher in diesem Teilkapitel formuliert und ergeben sich aus dem vorherigen Kapitel. Einerseits soll mit HyperDocs ein digitales Lernwerkzeug evaluiert werden, das die Binnendifferenzierung über digitale Lernhilfen ermöglicht. Andererseits bestehen in der Literatur verschiedene offene Forschungsfragen, die mithilfe von HyperDocs beantwortet werden sollen. Neben den im Folgenden aufgeführten Forschungsfragen werden weitere Detailfragen aufgeworfen, die sich im Laufe der Erhebung oder der Auswertung der vorliegenden Studie ergaben. Diese Fragen werden im Ergebnisteil berücksichtigt und im Diskussteil erörtert. In Zusammenhang mit den nachfolgend aufgeführten Forschungsschwerpunkten der intrinsischen Motivation, Usability und Lernhilfen stehen weitere Variablen wie die tabletbezogene Selbstwirksamkeit, der Cognitive Load und die Nutzungshäufigkeit von digitalen Medien (vgl. Kapitel 4.2.2). Die Forschungsfragen werden unter Berücksichtigung der Mittel- und Oberstufe sowie des Gymnasiums und der Gesamtschule betrachtet und gehen aus den dargestellten Forschungslücken hervor.

#### **3.5.1 Forschungsfragen im Bereich der Motivations- und Interessensforschung**

In den Veröffentlichungen zur Interessens- und Motivationsforschung konnte teilweise ein motivationsförderndes Potential von digitalen Werkzeugen unter bestimmten

Voraussetzungen nachgewiesen werden (Hillmayr et al., 2020). Dieser Effekt zeigt sich oft jedoch nur temporär beziehungsweise nimmt mit der Zeit ab (Higgins et al., 2019). Eine mögliche Begründung dafür ist das Fehlen einer *hold*-Komponente, die für ein individuelles Interesse am Lerngegenstand verantwortlich ist, nachdem ein anfängliches Interesse über die *catch*-Komponente geschaffen wurde (Mitchell, 1992; Mitchell, 1997). Manche Autor:innen sehen den Grund im Neuheitseffekt, während andere Forscher:innen diese Annahme verwerfen (Jeno et al., 2019a). Es soll daher überprüft werden, ob motivationale Veränderungen beim Lernen mit HyperDocs auftreten und ob sich eine Veränderung des Fachinteresses Chemie zeigt. Daher wurde folgende Forschungsfrage formuliert:

*F\_M1: Inwieweit verändern sich das Fachinteresse Chemie und die intrinsische Motivation beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

Zudem finden sich in der Literatur Hinweise auf Unterschiede in der Motivation und dem Interesse zwischen den Geschlechtern im Chemieunterricht beim Lernen mit digitalen Medien (z. B. Bruckermann et al., 2016; Higgins et al., 2019). Daher wird folgende Frage aufgeworfen:

*F\_M2: Inwieweit unterscheiden sich das Fachinteresse Chemie und die intrinsische Motivation zwischen den Geschlechtern beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

Aufgrund des bisherigen Forschungsstandes (Karapanos et al., 2018) ist außerdem ein Zusammenhang zwischen der Usability und dem Interesse anzunehmen. Dieser Zusammenhang ist nachvollziehbar und wurde in Teilen bereits nachgewiesen. Um die Forschung in diesem Gebiet weiterhin zu stützen, wird die nachfolgende Forschungsfrage formuliert, wobei die Variablen der intrinsischen Motivation nach der Self-Determination Theory das Interesse beinhalten:

*F\_M3: Welcher Zusammenhang besteht zwischen der Usability und der intrinsischen Motivation beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

Im Bereich der Motivationsforschung werden keine weiteren Detailfragen formuliert.

### 3.5.2 Forschungsfragen im Bereich der Usabilityforschung

Eines der Ziele der vorliegenden Arbeit ist die Erprobung von HyperDocSystems als Unterrichtsmittel im Bereich der Binnendifferenzierung. Dazu erfolgt die Bewertung von HyperDocSystems durch Schüler:innen hinsichtlich der Usability. Durch den umfangreichen und weit gestreuten Einsatz von HyperDocSystems im Regelunterricht sollen Fehler und Benutzerschwierigkeiten aufgedeckt werden, die in der Pilotierung nicht aufgedeckt werden konnten. Zudem existieren bisher wenige Usabilitystudien im Bereich des Schulwesens. Bei HyperDocSystems handelt es sich um ein neu entwickeltes Programm, dessen Gebrauchstauglichkeit noch nicht vollumfänglich erprobt wurde. Es wurde daher bewusst, im Gegensatz zu anderen Usabilitystudien, der Feldeinsatz gewählt, um die tatsächliche Praxistauglichkeit abzudecken (summativer Test). Verschiedene Autor:innen (Meiselwitz & Sadera, 2008; Tselios et al., 2001) konnten bereits einen Zusammenhang zwischen der Usability und der Lernwirksamkeit aufzeigen. Eine hohe Usability ist daher für ein Werkzeug, das primär für den Wissenserwerb konzipiert wurde, unerlässlich. Daraus leitet sich folgende Forschungsfrage ab, mit der die Usability der HyperDocs bestimmt werden soll:

*F\_U1: Wie bewerten Schüler:innen die Usability von HyperDocs während des Einsatzes im Regelunterricht über den Zeitraum einer vierstündigen Unterrichtsreihe?*

In Zusammenhang mit der Usability ergeben sich weitere grundlegende Fragen, die in der Literatur zum Teil widersprüchlich beantwortet wurden. So konnten manche Studien Unterschiede zwischen den Geschlechtern (z. B. Karapanos et al., 2018; Orfanou et al., 2015) finden, während andere Studien keine Unterschiede feststellten (z. B. Billestrup et al., 2016). Aus diesem Grund wird folgende Forschungsfrage formuliert:

*F\_U2: Inwieweit bestehen Unterschiede zwischen den Geschlechtern bei der Bewertung der Usability von HyperDocs durch Schüler:innen über den Zeitraum einer vierstündigen Unterrichtsreihe?*

Verschiedene Usabilitystudien konnten außerdem einen Einfluss der Einsatzdauer beziehungsweise des Zeitpunktes der Evaluation feststellen (z. B. Kortum & Johnson, 2013; Lee & Ha, 2019). Daher wurde nachfolgende Forschungsfrage in Bezug auf HyperDocs formuliert:

*F\_U3: Inwieweit tritt eine Veränderung auf in der Bewertung der Usability durch Schüler:innen in Bezug auf HyperDocs über eine Einsatzdauer von vier Unterrichtsstunden?*

Die Detailfragen in Bezug auf die Usability werden im Folgenden formuliert. Sie ergeben sich aus dem Forschungsstand zum Cognitive Load (Davids et al., 2014; Ibili & Billinghamurst, 2019) und den computerbezogenen (hier: tabletbezogenen) Kompetenzen (Karapanos et al., 2018; Venkatesh, 2000) oder sind im Laufe der Untersuchung aufgetreten.

*D\_U1: Welcher Zusammenhang besteht zwischen der Usability und der kognitiven Last beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

*D\_U2: Welcher Zusammenhang besteht zwischen der Usability und den tabletbezogenen Kompetenzen beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

*D\_U3: Welcher Zusammenhang besteht zwischen der Usability und der Nutzungshäufigkeit von digitalen Endgeräten beim Lernen mit HyperDocs?*



### 3.5.3 Forschungsfragen im Bereich der Heterogenitätsforschung und Binnendifferenzierung durch Lernhilfen

Eine zentrale Funktionalität von HyperDocSystems ist das Anreichern des digitalen Arbeitsblattes mit Lernhilfen. Erst damit wird ein Document zu einem HyperDoc. Durch die digitale Umsetzung der Binnendifferenzierung durch Lernhilfen und Vertiefungsaufgaben können einerseits vielfältigere Lernmaterialien erstellt werden, die unter anderem auch Audio- und Videoelemente beinhalten. Andererseits lässt sich das Nutzungsverhalten der Schüler:innen nachvollziehen, wodurch sich verschiedene Forschungsfragen ableiten lassen, die vorher nur unter teilweise künstlichen Bedingungen untersucht werden konnten, indem zum Beispiel die Lernhilfen in versiegelten Briefumschlägen verteilt wurden und dadurch deren Nutzung nachvollzogen werden konnte (Arnold et al., 2016). Die nachfolgenden Forschungsfragen haben aufgrund der neuen Verfügbarkeit der Nutzungsdaten teilweise einen explorativen Charakter, die unter anderem dazu dienen, neue Desiderata aufzuwerfen. Im Bereich der analogen Nutzung von Lernhilfen sind Unterschiede zwischen den Geschlechtern bekannt (Cheong et al., 2004; Kessel & Steinmayr, 2013; Schenke et al., 2015), die insbesondere auf die Rollenbilder der Jugendlichen zurückzuführen sind (Ryan et al., 2009). Außerdem ist bekannt, dass der Kontext der Hilfe-nutzung einen Einfluss auf die Verwendung von Lernhilfen nimmt (Karabenick & Knapp, 1988). Aus diesen Gründen wird folgende Forschungsfrage aufgeworfen:

*F\_L1: Inwieweit bestehen Unterschiede zwischen den Geschlechtern in der Nutzungshäufigkeit von Lernhilfen über den Zeitraum einer vierstündigen Unterrichtsreihe?*

Eine weitere Forschungsfrage zielt auf einen wenig untersuchten Zusammenhang zwischen der Motivation und dem Differenzierungsangebot ab (z. B. Yang & Cao, 2013; Lynch & Dembo, 2004). Es ist denkbar, dass sich die Motivation beim Lernen mit HyperDocs über die Nutzung der Lernhilfen nachweisen lässt. In einer anderen Studie konnte bereits ein Zusammenhang zwischen der Intensität der Nutzung von digitalem Lernmaterial und dem Interesse festgestellt werden (Urhahne et al., 2004). Aus diesem Grund wird folgende Forschungsfrage aufgeworfen:

*F\_L2: Welcher Zusammenhang besteht zwischen der Nutzungshäufigkeit der Lernhilfen durch Schüler:innen und der intrinsischen Motivation beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

Weiterhin soll untersucht werden, inwieweit ein Zusammenhang zwischen dem Leistungstand der Lernenden und der Nutzung des Differenzierungsangebots besteht, da auch hierbei widersprüchliche Ergebnisse vorliegen (z. B. Bartholomé et al., 2006; Ryan & Schin, 2011) und HyperDocSystems eine genaue Quantifizierung der Nutzungshäufigkeit ermöglicht:

*F\_L3: Welcher Zusammenhang besteht zwischen der Nutzungshäufigkeit der Lernhilfen und dem Leistungstand der Schüler:innen beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

Die nachfolgende Forschungsfrage weist einen explorativen Charakter auf, um die genauen Ursachen der Nutzung des Differenzierungsangebots zu bestimmen:

*F\_L4: Aus welchen Gründen nutzen Schüler:innen Lernhilfen und Vertiefungsaufgaben beim Lernen mit HyperDocs während einer vierstündigen Unterrichtsreihe?*

Die Detailfragen in Bezug auf das Differenzierungsmaterial sind nachfolgend aufgeführt. Sie leiten sich aus dem bisherigen Forschungsstand ab oder sind im Laufe der Untersuchung aufgetreten.

*D\_L1: Welcher Zusammenhang besteht zwischen der Nutzung von Lernhilfen durch Schüler:innen und der kognitiven Last beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

*D\_L2: Inwieweit bevorzugen Schüler:innen eine bestimmte Darstellungsform der Hilfen bei der Nutzung von HyperDocs während einer vierstündigen Unterrichtsreihe?*

## 4. Forschungsdesign

Aus den aufgeworfenen Forschungsfragen leitet sich ein spezielles Forschungsdesign ab, das im folgenden Kapitel dargelegt wird. Dazu folgt zunächst eine begründete Einführung in die verwendeten Methoden. Anschließend werden die Datenbasis sowie die verwendeten Hilfsmittel (Instrumente und Technik) beschrieben. Daraufhin werden die drei Pilotierungen und die daraus abgeleiteten Anpassungen der Studie vorgestellt. Das letzte Teilkapitel beinhaltet die eigentliche Durchführung der Hauptstudie mit den entsprechenden finalen Unterrichtsreihen.

### 4.1 Methoden

In diesem Unterkapitel werden die wichtigsten Methoden dieser Arbeit im Zusammenhang mit der Datenerhebung und Datenaufbereitung zusammengefasst. Das umfasst die Methode der Erhebung der Nutzungshäufigkeit der Hilfen und Vertiefungsaufgaben, die für die Datenaufbereitung und Datenauswertung verwendeten Programme und den Umgang mit fehlenden Werten.

#### 4.1.1 Monitoring und Nutzungshäufigkeit des Differenzierungsangebots

Wie die Nutzung der Hilfen und Vertiefungsaufgaben in HyperDocSystems gespeichert werden, wurde bereits in Kapitel 2.4.4 beschrieben. Es wird nun dargelegt, wie die Rohdaten aufbereitet und ausgewertet wurden.

##### **Erfassung und Aufbereitung der Nutzungshäufigkeiten**

Zunächst mussten die Rohdaten aus der Datenbank von HyperDocSystems exportiert werden. Dazu wurden die Monitoring-Daten mittels eines SQL-Befehls gesammelt und anschließend über die Downloadfunktion von phpMyAdmin über das Webpanel als csv-Datei gespeichert. Anschließend wurden die klassen- und stundenspezifischen Datensätze manuell per Hand extrahiert und in neue xlsx-Dateien gespeichert, sodass für jede Klasse vier neue Datensätze bestanden. Zum Schluss wurden diese vier Datensätze in einem Datensatz mit unterschiedlichen Mappen (Stunde 1, ..., Stunde 4) vereint. Zusätzlich wurde die Differenz zwischen dem Schließ- und Öffnungszeitpunkt einer Hilfeseite<sup>5</sup> gebildet

---

<sup>5</sup> In diesem Kapitel sind damit fortlaufend auch die Vertiefungsaufgaben gemeint, sofern nicht anders beschrieben.

und in einer neuen Spalte gespeichert. Es konnte vorkommen, dass der Schließzeitpunkt einer Hilfe nicht mehr korrekt erfasst wurde, wenn eine Hilfe geöffnet und der *Browser-tab* anschließend geschlossen wurde. In solchen Fällen waren kein Datum und keine Uhrzeit für die Hilfe in der Datenbank hinterlegt, weshalb zum Beispiel Excel bei der Bildung der Differenz den Wert „#WERT“ einträgt. Die Fälle wurden manuell korrigiert, indem die Dauer „00:00:00“ eingetragen wurde. Zusätzlich wurden Hilfen entfernt, die nicht der jeweiligen Stunde zuzuordnen waren, da es sich um Hilfen aus den vergangenen Stunden handelte, die durch die Schüler:innen zum Beispiel aus Langeweile nochmal geöffnet wurden.

#### **Bestimmung der theoretischen Verweildauer der Lernhilfen**

Um die Nutzung des Differenzierungsangebots differenzierter betrachten zu können, werden die Hilfen als "gelesen" und „nicht (vollständig) gelesen“ eingestuft. Das Lesen einer Hilfe oder Vertiefungsaufgabe benötigt eine gewisse Zeit. Diese Zeit ist abhängig von der Komplexität und dem Umfang des Inhalts. Während beispielsweise für ein Video oder eine Audiodatei die Spieldauer als kleinste Zeiteinheit interpretiert werden kann, die ein Individuum aufbringen muss, um das Medium vollständig zu erschließen, müssen bei Texten und Bildern andere Kriterien herangezogen werden. Für das Lesen von Texten kann eine Lesegeschwindigkeit definiert werden, über die wiederum die theoretische Leszeit berechnet werden kann. Diese Lesegeschwindigkeit wurde für die vorliegende Arbeit nach dem „Lesegeschwindigkeits- und -verständnisstest für die Klassenstufen 6 – 12“ (LGVT6-12; Schneider et al., 2007, zitiert nach Mézes, 2016) auf 5.221 Wörter pro Sekunde festgelegt, da diese Geschwindigkeit von einem 10. Klässler nicht erreicht werden kann. Nerlich (2013) untersuchte in einer Studie unter anderem die Lesegeschwindigkeit von bilingual unterrichteten 10. Klassen an Gymnasien. Dort erreichten manche Schüler:innen eine Lesegeschwindigkeit von circa 4.9 Wörtern pro Sekunde. Der Durchschnitt lag im Bestfall bei 3.28 Wörtern pro Sekunde (Nerlich, 2013). Damit zeigt sich die breite Spanne der Lesegeschwindigkeit.

Bei der Erfassungsdauer eines Bildes konnten keine klaren Zahlen gewonnen werden. Kinder ab 10 Jahren erreichen beispielsweise bei der Identifizierung von Gesichtern eine Antwortzeit von unter einer Sekunde (Croker & Maratos, 2011). Bei einer Studie zum Erkennen und Kategorisieren von Tier- und Naturbildern konnten Reaktionszeiten von 337 – 400 Millisekunden von jungen Erwachsenen erreicht werden (Delorme et al., 2004). Eine andere Studie sieht die Verarbeitungszeit von Bildern bei 150 Millisekunden

(Thorpe et al., 1996). Für die vorliegende Arbeit wurde eine Verarbeitungszeit von einer Sekunde festgelegt. Es handelt sich um eine sehr optimistische Schätzung, wonach Jugendliche womöglich mehr Zeit benötigen, um ein Bild zu erfassen. Auf der anderen Seite ist dieser Prozess auch von der Komplexität des Bildes abhängig. Besteht eine Hilfe aus mehreren Bildern, wurde für jedes weitere Bild eine zusätzliche Sekunde addiert. Bei zusätzlichen Sätzen wurde die Anzahl der Wörter über die Lesegeschwindigkeit berücksichtigt. Die genauen Zeiten für jeder Hilfe sind in Anhang 9.1.1 festgehalten.

### **Auswertung der Nutzungshäufigkeiten**

Nach der Aufbereitung der Rohdaten wurden die verschiedenen Datensätze der Klassen, die nun wiederum die Daten der vier Stunden beinhalteten, durch ein Python-Skript ausgewertet. Dabei kamen folgende Bibliotheken zum Einsatz: *Pandas*, welches zum Erzeugen und Verarbeiten von komplexeren Datensätzen dient (NumFOCUS, 2021), und *xlrd*, das zum Einlesen von *xlsx*-Dateien dient (Withers, 2021). Das Skript fasste die einzelnen Hilfeaufrufe nach Nutzer:innenaccount zusammen, wodurch die Gesamtzahl der Aufrufe pro Stunde und Account bestimmt werden konnte. Dabei wurden zwei Datensätze gebildet. Ein Datensatz beinhaltet die Rohdaten, bei denen die Lesedauer nicht berücksichtigt wurde und ein weiterer Datensatz beinhaltet die nach Lesedauer bereinigten Daten. Für die spätere Auswertung der Hilfenutzung über die Zeit (vgl. Kapitel 5.4.3 & 5.4.4) wurde der Datensatz verwendet, der die „gelesenen“ Hilfen pro Account aufführt. Diese Daten wurden anschließend durch die pro Stunde zur Verfügung stehende absolute Anzahl an Hilfen und die zur Verfügung stehende Bearbeitungszeit der Stunde geteilt, um individuelle Unterschiede zwischen den Lerngruppen zu berücksichtigen. Bei den Auswertungen zur Nutzungshäufigkeit wurden Hilfen, die zum Versuch gehörten oder zur Wiederholung dienten (Mittelstufe: 3. Stunde; Oberstufe: 2. & 4. Stunde), nicht berücksichtigt, da sie nicht Teil einer Aufgabenstellung waren und damit zum Beispiel bei den Zusammenhängen zum Cognitive Load oder den Fähigkeiten der Lernenden eine geringe beziehungsweise keine Rolle spielen (vgl. Anhang 9.1.1). Gleiches gilt für Hilfen, die als „Sonstige“ eingestuft wurden. Hierbei handelt es sich um allgemeinere Informationen. Bei den Vertiefungsaufgaben wurde die zur Verfügung stehende Unterrichtszeit ebenfalls berücksichtigt.

### 4.1.2 Analyse der Daten und quantitative Datenauswertung

Die Rohdaten wurden zunächst in SPSS (Version 26) importiert und aufbereitet, indem Variablen umbenannt und fehlende Werte umkodiert wurden. Zusätzlich wurde auf fehlende Eingaben geprüft, die beispielsweise bei der Übertragung der Ergebnisse aus den Befragungen mittels Paper-Pencil-Prinzip auftreten konnten. Alle Analysen erfolgten in R. Das betrifft die deskriptive Beschreibung, die Erstellung von Diagrammen und die eigentliche statistische Auswertung. In Tabelle 1 sind die verschiedenen R-Pakete und deren Funktionen aufgeführt. Eine Ausnahme bildet die Hauptkomponentenanalyse in der letzten Pilotierung, die mittels SPSS durchgeführt wurde.

Tabelle 1: Verwendete R-Pakete in der Datenauswertung.

<i>R-Paket</i>	<i>Funktionsumfang</i>
afex	Durchführung von ANOVA und Prüfung verschiedener Voraussetzungen.
dplyr	Manipulation von Datensätzen, z.B. Filter.
ggplot	Erstellung von Diagrammen.
haven	Importierung von SPSS-Dateien.
lavaan	Faktorenmodelle und Strukturgleichungsmodellierung.
norm2	Imputation fehlender Werte.
rstatix	Statistische Tests und Prüfung von deren Voraussetzungen.
tidyverse	Sammlung verschiedenere Pakete zum Umgang mit Datensätzen.
welchADF	Robuste ANOVA auf Basis der Welch-James Statistik.
WRS	Robuste Testverfahren auf Grundlage von Wilcox (2021).
WRS2	Robuste Testverfahren auf Grundlage von Wilcox (2021). Anmerkung: Geringerer Funktionsumfang als <i>WRS</i> , jedoch teilweise benutzerfreundlicher.

Bei der Datenauswertung wurden die Zusammenhänge aufgrund von Ausreißern mittels Spearmans rho ( $\rho$ ) berechnet. Gruppenunterschiede wurden mittels einer robusten gemischten Varianzanalyse gerechnet, da die Voraussetzungen einer normalen Varianzanalyse nicht immer erfüllt waren. Die Auswertung der gemischten Varianzanalyse selbst erfolgte mittels des R-Pakets „WRS“ (Stevens Institute, 2022). Die Grundlage des Pakets bildet Wilcox (2021). Die Berechnung erfolgt unter Anwendung von getrimmten Mittelwerten (engl.: *trimming*) und *Winsorizing*. Als Trimmung wurde der Faktor 0.20 gewählt. Aufgrund der kleinen Teilstichproben wurde für die Berechnung des Signifikanzniveaus

zusätzlich Bootstrapping angewandt. Robuste Effektstärken stehen für diese Methode nicht zur Verfügung. Die anschließenden Analysen der Zwischensubjektfaktoren erfolgten mittels robuster t-Tests des R-Pakets „WRS2“ (Mair & Wilcox, 2020). Es wurden erneut getrimmte Mittelwerte verwendet.  $\xi$  ist in diesem Zusammenhang das erklärende Maß der Effektgröße. Werte von 0.10, 0.30 und 0.50 gelten als kleine, mittlere und großen Effektstärken (Mair & Wilcox, 2020). Die Untersuchung des Innersubjektfaktors erfolgte mittels einer robusten Variante der Varianzanalyse mit Messwertwiederholung, ebenfalls auf Basis von (Mair & Wilcox, 2020). Die Effektstärken ( $\delta_i$ ) werden nach einer modifizierten Variante von Algina et al. (2005) angegeben. 0.3, 0.5 und 0.7 gelten hierbei als kleine, mittlere und große Effektstärken.

Die Prüfung der Konstruktvalidität erfolgte mittels einer Faktorenanalyse, während die Reliabilitäten über Cronbachs  $\alpha$  bestimmt wurden. Bei der Ergebnisdarstellung sind weitere Anmerkungen zur Auswertung geschildert.

#### 4.1.3 Umgang mit fehlenden Werten

Fehlende Werte (*engl.: Missings*) stellen ein gängiges Problem in der praxisorientierten Forschung dar. In der Regel fehlen bei sozialwissenschaftlichen Studien zwischen ein und zehn Prozent aller Datenpunkte bei einer Variablen (Schnell et al., 2018). Die Gründe für das Fehlen eines Wertes können vielfältig sein. So kann es sein, dass Versuchsteilnehmer:innen eine Frage übersehen oder eine spezielle Frage nicht beantworten möchten. Es wird daher zwischen verschiedenen Mustern fehlender Werte und verschiedenen Mechanismen des Fehlens von Messwerten unterschieden. Als besonders relevant hat sich der Mechanismus des Fehlens von Werten herausgestellt. Dabei wird untersucht, ob das Fehlen in Verbindung mit den zugrundeliegenden Werten der Items beziehungsweise Variablen steht (Little & Rubin, 2002). Little und Rubin (2002, S. 12) unterscheiden zwischen „Missing completely at random“ (MCAR), „Missing at random“ (MAR) und „Missing not at random“ (MNAR).

Bei MCAR hängt das Fehlen der Variable nicht von ihrem Wert oder anderer Variablen ab (Little & Rubin, 2002). Allison (2010) verdeutlicht das an folgendem Beispiel: Sind Angaben zum Einkommen zu geben, so hängt das Fehlen dieser Angabe bei MCAR-Mechanismus nicht vom Einkommen selbst ab und auch nicht von anderen Variablen, wie zum Beispiel der Schulbildung. Nach Allison ist die MCAR in der Praxis nur selten komplett erfüllt. Bei MAR hängt das Fehlen der Variable von den beobachteten Werten der

Variable ab und nicht von ihren fehlenden Werten (Little & Rubin, 2002). Allison (2010) ergänzt, dass das Fehlen der Variable von anderen Variablen, die beobachtet wurden, abhängt. Das Fehlen lässt sich durch andere Variablen begründen. Sind sowohl die MCAR als auch die MAR-Bedingung verletzt, liegt ein NMAR-Mechanismus vor. Dabei hängt das Fehlen einer Variablen einzig von sich selbst ab.

Welcher Ausfallmechanismus vorliegt, kann teilweise überprüft werden. Für die MCAR-Bedingung existiert der LITTLE-Test (Little, 1988). Dabei überprüft der Test, ob die Nullhypothese, die Daten fehlen zufällig verteilt, gehalten werden kann. Wird der Test signifikant, muss die Nullhypothese verworfen werden. Es bleibt allerdings offen, ob die Daten bedingt beziehungsweise zufällig fehlen (MAR) oder nicht zufällig fehlen (MNAR). Der Test erfordert weiterhin eine multivariate Normalverteilung (Reisinger et al., 2012, S. 73). Jedoch zeigt sich der Test gegenüber der Verletzung dieser Annahme als robust (Little, 1988). Reisinger und Kollegen (2012, S. 73) merken an, dass der Test die MCAR-Annahme nur stützt, jedoch nicht beweist. Außerdem deuten verschiedene Simulationsstudien auf eine geringe statistische Power des Tests hin.

Die MAR-These ist, außer bei einem gezielten Fehlen der Werte durch die Befragten, generell nicht vollständig prüfbar (Schafer & Graham, 2002). Graham et al. (1997) merken an, dass selbst bei unbekanntem Mechanismus oder Abweichungen von der MAR-Annahme, nur ein geringer Bias entsteht und das Ergebnis der Studien nicht verändert wird. Gestützt wird diese These von Collins et al. (2001), die bei einer falschen Annahme des MAR-Mechanismus nur geringen Einfluss auf Schätzwerte und Standardfehler gefunden haben. Letztlich kann nicht eindeutig belegt werden, ob die Daten zufällig oder nicht zufällig fehlen (Wirtz, 2004). Bei vollständigem MNAR-Mechanismus müssen die fehlenden Werte anhand eines konkreten Modells ersetzt werden (Schafer & Graham, 2002). Dabei bleibt jedoch oft offen, ob das entwickelte Modell das richtige ist (Allison, 2010). Bei einer falschen Wahl des Modells können sogar schlechtere Schätzungen für die fehlenden Werte entstehen als bei Methoden, die die MAR-Annahme voraussetzen (Demirtas & Schafer, 2003). Je nach Mechanismus können verschiedene Verfahren zur Beseitigung der Missings angewandt werden, um einen vollständigen Datensatz zu erhalten. Dabei erzeugen die angewandten Methoden jedoch unter Umständen einen gewissen Bias, insbesondere dann, wenn der Ausfallmechanismus nicht ausreichend geklärt ist (Allison, 2010). Einige der relevanten Verfahren zum Umgang mit fehlenden Werten werden nachfolgend dargestellt.



### **Listenweiser Fallausschluss**

Der listenweise Fallausschluss ist eine in der Praxis häufig angewandte Methode, um vollständige Datensätze zu erhalten, auch „*complete case analysis*“ (Allison, 2010, S. 637) genannt. Dabei werden unvollständige Datenpunkte vollständig aus dem Datensatz entfernt. In der Praxis erweist er sich als eine einfache und praktische Methode, da der listenweise Fallausschluss einfach umzusetzen ist und in jeder statistischen Methode angewandt werden kann. Unter der MCAR-Annahme wird weiterhin kein Bias erzeugt und Standardfehler richtig geschätzt. Selbst eine Verletzung dieser Annahme führt nicht zwangsläufig zu einem Bias in den Daten. Dem gegenüber stehen der Informationsverlust, der durch das Löschen der unvollständigen Datenpunkte entsteht, sowie eine geringere statistische Power durch die kleinere Stichprobe (Allison, 2010).

### **Paarweiser Fallausschluss**

Der paarweise Fallausschluss findet ebenfalls häufig Anwendung, insbesondere bei der Berechnung linearer Regressionsanalysen oder Korrelationen. Dabei werden bei jeder Berechnung alle zur Verfügung Datenpunkte verwendet (Allison, 2010; Wirtz, 2004). Es können jedoch die Datenpunkte einer Berechnung in der anderen wiederum fehlen. Der Stichprobenumfang bleibt in großen Teilen erhalten. Als Nachteil ist jedoch zu nennen, dass die verschiedenen ermittelten Statistiken aus unterschiedlichen Teilstichproben stammen.

### **Imputation**

Bei der Imputation werden fehlende Werte durch möglichst „plausible zuverlässige“ (Reisinger et al., 2012, S. 124) ersetzt. Man unterscheidet dabei zwischen einfachen Imputationen (*Single Imputation*) und multiplen Imputationen (*Multiple Imputation*). Wie der Name bereits nahelegt, wird bei der einfachen Imputation jeder fehlende Wert durch einen imputierten Wert ersetzt. Am Ende wird ein vervollständigter Datensatz gewonnen. Unter Beachtung der jeweiligen Bedingungen können einfache Imputationsverfahren nützliche Ergebnisse erbringen (Reisinger et al., 2012, S. 126). Schaffer & Graham (2002) merken außerdem an, dass es eine effizientere Methode als den listenweisen Fallausschluss darstellt, da der Verlust der statistischen Power verringert wird. Dennoch gilt es zu beachten, dass die Varianz und der Standardfehler teilweise unterschätzt werden, da die Imputation zu einer Homogenisierung des Datensatzes führen kann (Allison, 2010). Graham (2009) schlägt selbst bei einem geringen einstelligen Prozentsatz an fehlenden Werten eine Imputation mit geeigneten Verfahren vor.

Die einfachste Form der Imputation ist die Ersetzung durch den Variablenmittelwert. Dabei werden die fehlenden Werte einer Variablen durch den Mittelwert der beobachteten Datenpunkte dieser Variable ersetzt (Wirtz, 2004). Das Verfahren ist sehr einfach umzusetzen und wird in SPSS standardmäßig angeboten (Reisinger et al., 2012, S. 127). Diese Methode wird jedoch von verschiedenen Autor:innen nicht empfohlen, da die Stichprobe homogenisiert wird (Graham, 2009; Schafer & Graham, 2002; Wirtz, 2004). Das Verfahren wird daher von Schendera (2007) nur bei sehr geringem Prozentsatz (kleiner fünf Prozent) fehlender Werte empfohlen. Regressionsbasierte Ersetzungsmethoden, die die Varianz etwas erhöhen, sind ebenfalls nicht zu empfehlen (Allison, 2010; Graham, 2009; Wirtz, 2004). Weitere Verfahren sind zum Beispiel die Hot-Deck- und Cold-Deck-Imputation oder auch die „conditional mean imputation“ (Schafer & Graham, 2002, S. 159). Mathematisch aufwändigere Verfahren der Imputation sind *Maximum-Likelihood-Verfahren* (ML). Dazu zählt unter anderem der EM-Algorithmus (*Expectation-Maximization*), der in Bezug auf fehlende Werte von Dempster et al. (1977) beschrieben wurde (Schafer & Graham, 2002; Graham, 2009). Der EM-Algorithmus setzt die MAR-Annahme voraus (Schafer & Graham, 2002; Reisinger et al., 2012, S. 141). Bei dieser Methode wird allgemein eine größere Stichprobe von  $n > 100$  empfohlen, wobei die optimale Stichprobengröße auch stark von der Anzahl der fehlenden Werte abhängt (Reisinger et al., 2012, S. 141; Schafer & Graham, 2002; Wirtz, 2004). McNeish (2017) konnte in seiner Simulationsstudie zeigen, dass ML-Methoden selbst bei relativ kleinen Stichproben ( $n = 50$ ) und einem Prozentsatz fehlender Werte von zehn Prozent ein geringer Bias entsteht, selbst unter der MNAR-Annahme. Mit kleineren Stichproben oder größerem Prozentsatz fehlender Werte entsteht allerdings ein wesentlicher Bias (McNeish, 2017). Der EM-Algorithmus besteht im Wesentlichen aus zwei Schritten: Im E-Schritt werden auf Basis der vorhandenen gültigen Variablen unter Berücksichtigung verschiedener Parameter die fehlenden Werte geschätzt und ersetzt (Reisinger et al., 2012, S. 141 - 142; Wirtz, 2004). Im M-Schritt werden die Parameter wie Mittelwert, Standardabweichung und Korrelationen mit den ersetzten Werten aus dem E-Schritt berechnet. Gibt es eine Abweichung zur ursprünglichen „Informationsstruktur“ (Wirtz, 2004, S. 114), wird erneut der E-Schritt durchlaufen. Diese beiden Schritte werden nun iterativ so lange durchlaufen bis sich die Informationsstruktur nicht mehr ändert (Wirtz, 2004). Der EM-Algorithmus ist robust gegen die Verletzung der multivariaten Normalverteilung der betrachteten Variablen. Variablen mit fehlenden Werten sollten normalverteilt sein (Reisinger et al., 2012, S.

141). Der Algorithmus erweist sich jedoch auch als robust gegenüber schiefen Daten (Graham et al., 1996).

Eine Erweiterung der einfachen Imputation stellt die multiple Imputation (MI) dar. Dabei werden fehlende Werte durch plausible Werte, die aus einer Wahrscheinlichkeitsverteilung gezogen werden, geschätzt. Dadurch entstehen mehrere, in der Regel fünf, vervollständigte Datensätze. Jeder der so gewonnenen Datensätze wird dann mit den gewünschten statistischen Verfahren analysiert (Schnell et al., 2018, S. 430). Abschließend werden die Ergebnisse aus den verschiedenen Datensätzen zu einem gemeinsamen Ergebnis zusammengefasst. Dafür gibt es verschiedene Regeln (Rubin, 1987). Die Vorteile der MI sind unter anderem genauere Standardfehler und eine höhere Varianz, was zum Teil zu einem geringeren Bias führt (Allison, 2010). Gleichzeitig ist jedoch durch die mehrfache Analyse der verschiedenen Datensätze und deren Kombination mehr Aufwand bei dieser Methode erforderlich (Schnell et al., 2018, S. 430).

### **Umgang mit fehlenden Werten in den eigenen Datensätzen**

In den gewonnenen Datensätzen existieren fehlende Werte. Bereits vor einer genaueren Datenanalyse können Gründe für Missings genannt werden. Die Fragebögen wurden im Paper-Pencil-Format erhoben. Damit gibt es im Gegensatz zu digitalen Methoden keine Eingabeerinnerung, falls Fragen ausgelassen wurden. Bei der Digitalisierung der analogen Fragebögen fiel auf, dass teilweise bei Fragen zwei Kreuze gesetzt und bei der darauffolgenden oder vorhergehenden Frage kein Kreuz gesetzt wurde. Es liegt der Schluss nahe, dass die Schülerin oder der Schüler bei der Beantwortung der Frage verrutscht ist. Ein weiterer Ausfall charakterisiert sich durch das Fehlen ganzer Fragebatterien, gerade am Ende des Fragebogens. Dazu zählt auch das Fehlen einzelner Seiten, was gerade bei den mehrseitigen Fragebögen POST\_1\_E und POST\_2 zu beobachten war. Die Gründe für das Auslassen dieser Fragebatterien können einerseits im Übersehen liegen, da die Fragen gar nicht gelesen beziehungsweise wahrgenommen wurden. Andererseits lässt das Auslassen der Fragebatterien am Ende des Fragebogens auf fehlende Motivation zum Ausfüllen der Fragen schließen.

Allgemein lassen sich diese Arten von fehlenden Werten als MCAR (Übersehen der Frage / doppeltes Ankreuzen) oder MAR (Abbruch aufgrund der Motivation / Pause) beschreiben. Die Anzahl der gesamten fehlenden Werte liegt pro Fragebogen bei maximal zwei Prozent (vgl. Tabelle 2). Die Angaben beinhalten keine fehlenden Werte von Variablen durch eine bewusste Nichtbeantwortung durch Aufforderung im Fragebogen. Die

Schüler:innen sollten beispielsweise keine Fragen zu Hilfen oder Vertiefungsaufgaben beantworten, wenn sie diese in einer Stunde nicht nutzten. Gleiches gilt für die tabletbezogenen Kompetenzen in der Vorbefragung oder die Usabilitybewertung für Textfelder beziehungsweise die Freihandeingabe.

*Tabelle 2: Anzahl der fehlenden Werte pro Fragebogen in Prozent für die Mittel- und Oberstufe der Interventions- und Vergleichsgruppe.*

Fragebogen Interventionsgruppe						
	PRE	POST_1_E	POST_1_2	POST_1_3	POST_1_4	POST_2
Mittelstufe	1 %	1%	1%	< 1%	1%	1%
Oberstufe	< 1%	1%	< 1%	< 1%	< 1%	< 1%
Fragebogen Vergleichsgruppe						
	PRE_V	POST_1_EV	POST_1_2V	POST_1_3V	POST_1_F	
Mittelstufe	< 1%	< 1%	< 1%	< 1%	< 1%	
Oberstufe	< 1%	1%	1%	< 1%	2%	

### *Imputation*

Die fehlenden Werte wurden durch den EM-Algorithmus, wie zuvor beschrieben, bedingt ersetzt. Ein konsequenter listenweiser oder paarweiser Fallausschluss über alle Fragebögen hinweg hätte einen hohen Ausfall (*engl.: Dropout*) zur Folge. Das liegt unter anderem auch daran, dass für viele Fragestellungen nur Fälle berücksichtigt wurden, die in allen vier Stunden anwesend waren, beziehungsweise solche, für die vollständige Datensätze über alle vier Stunden bestehen. Gleichzeitig hat sich der EM-Algorithmus unter verschiedenen Bedingungen als ein erprobtes Verfahren erwiesen, um Datensätze zu vervollständigen. Da in den vorliegenden Datensätzen die Ausfallwahrscheinlichkeit (vgl. Tabelle 2) gering war, wurde auf eine multiple Imputation verzichtet. Gerade bei einem geringen Prozentsatz an Missings gibt es nur geringe Unterschiede zwischen den Methoden (McNeish, 2017). Dabei dürfte auch die Problematik der fehlenden Standardfehler (Graham, 2009) nur gering sein, da in den meisten Fällen in dieser Arbeit mithilfe weiterer Variablen zunächst eine Skala gebildet wurde und damit nicht allein die fehlenden Werte zur Auswertung herangezogen wurden. Gleichzeitig erhöht sich der Aufwand bei mehreren Datensätzen durch die multiple Imputation, da von diesen Datensätzen wiederum im Zuge des Verfahrens fünf bis zehn importierte Datensätze erstellt werden. Auch

Allison (2012) präferiert Maximum-Likelihood-Methoden gegenüber multipler Imputation, sofern eine geeignete Methode zur Verfügung steht. Die Imputation erfolgt mittels *R* und der Bibliothek *norm2* (Schafer, 2021). Die Funktion „emNorm“ des Pakets „norm2“ beinhaltet neben dem Imputationsmodell ein Dataframe „mean.imp“, das importierte Werte für die fehlenden Werte des ursprünglichen Dataframes beinhaltet. Die importierten Werte liegen im gleichen Wertebereich wie nach Anwendung der Funktion „impNorm“. Die Funktion „impNorm“ gibt allerdings, aufgrund des EM-Algorithmus, bei jeder Ausführung neue Werte für die Imputation an. Bei einer multiplen Imputation können demnach mehrere dieser Datensätze verwendet werden, um die jeweiligen Statistiken zu berechnen. Da jedoch keine multiple Imputation angewandt werden sollte und die Werte von „mean.imp“ im Wertebereich der Ausgabe von „impNorm“ lagen, wurde auf das Dataframe „mean.imp“ zurückgegriffen.

Für keine der angewandten Skalen gibt es wissentlich eine Mindestanzahl an Items, die zur Bildung dieser herangezogen werden müssen. Daher wurde eine bedingte Imputation aller fehlenden Werte einer Skala gewählt. In seltenen Fällen, zum Beispiel bei einem kompletten Dropout einer Fragebatterie, sind keine Werte für die Skalenbildung vor der Imputation vorhanden. Um dennoch die Nachteile einer Imputation gering zu halten, wurden bei (Sub-)Skalen lediglich Werte importiert, wenn fünfzig Prozent der Variablen vorhanden waren. Bei einer höheren Anzahl an Missings wurde der Fall aus der jeweiligen Berechnung ausgeschlossen.

## 4.2 Datenbasis

Die Daten aus den Erhebungen der Kontroll- und Vergleichsgruppe bilden die Datenbasis dieser Arbeit. Die Datenbasis der Pilotierungen wird in einem gesonderten Kapitel dargestellt. Im Folgenden erfolgt eine genaue Beschreibung der gewonnenen Daten.

### 4.2.1 Stichprobe

Die Stichprobe setzt sich aus zwei Teilstichproben zusammen, die zu verschiedenen Zeitpunkten erhoben wurden. Der gesamte Erhebungszeitraum für die Hauptstudie erstreckt sich von August 2020 bis Juli 2021. Während die Daten der Interventionsgruppe zwischen August 2020 und Dezember 2020 erhoben wurden, folgte die Durchführung mit der Vergleichsgruppe aufgrund der Corona-Pandemie ab Mai 2021. Aufgrund der unsicheren Situation im Zuge der Corona-Pandemie sollte zunächst eine ausreichend große

Stichprobe für die Forschungsfragen in Bezug auf die Usability und Lernhilfen gewonnen werden, um unter anderem HyperDocSystems als Lernwerkzeug zu evaluieren.

### **Interventionsgruppe**

An der Interventionsstudie haben drei Gymnasien und zwei Gesamtschulen in Rheinland-Pfalz teilgenommen. Aus Gründen der Anonymität werden die Schulen nicht weiter benannt. Die Verteilung der Klassen beziehungsweise Kurse ist in Tabelle 3 gezeigt.

*Tabelle 3: Anzahl der Klassen und Kurse, aufgeteilt nach Jahrgangsstufen und Schulform in der Interventionsgruppe.*

Jahrgangsstufe	Anzahl der Klassen und Kurse	
	Gymnasium	Gesamtschule
9	6	2
10	6	5
11 GK	1	1
11 LK	2	-
12 GK	2	-
12 LK	3	-
13 GK	2	-
13 LK	-	-

Insgesamt haben mehr als 550 Schüler:innen an der Interventionsstudie teilgenommen. Aufgrund verschiedener Fallausschlüsse reduziert sich die Stichprobe deutlich. Die Hauptstichprobe umfasst in der Mittelstufe 331 Lernende und in der Oberstufe 110. Die genaue Aufschlüsselung der Stichprobe ist Kapitel 5.1 zu entnehmen. Manche Teilnehmer:innen waren zum Beispiel nur in der letzten Stunde anwesend oder haben nur den Fragebogen der Vorerhebung ausgefüllt. Die Schulen wurden nicht unmittelbar randomisiert ausgewählt. Der Kontakt erfolgte über vorhandene Kooperationen mit den Schulen. Wie aus Tabelle 3 hervorgeht, haben aus verschiedenen Gründen nicht alle Jahrgangsstufen der Schulformen des Gymnasiums und der Gesamtschule an der Studie teilgenommen. Insbesondere die Leistungskurse der 13. Jahrgangsstufe befanden sich zum Zeitpunkt der Erhebung in der Vorbereitung für das Abitur. Zusätzlich sorgte die Coronapandemie allgemein für viele Ausfälle, sodass manche Lehrer:innen nicht noch weitere Unterrichtsversäumnisse durch die Studie verantworten wollten. Weiterhin hat sich

gezeigt, dass in den Gesamtschulen weniger Oberstufenkurse im Fach Chemie anzutreffen waren, weshalb nur ein Oberstufenkurs einer Gesamtschule an der Erhebung teilnahm. Keine der Klassen oder Kurse hatte zuvor mit HyperDocs gearbeitet. Für nahezu alle beteiligten Lerngruppen bildete die Unterrichtsreihe ein neues Fachthema. Lediglich eine ehemalige 9. Klasse, die im Zuge einer Zusammenlegung aufgelöst wurde, und ein 12. Grundkurs hatten die Messmethode, die als Lehrer:innenexperiment durchgeführt wurde, bereits einmal ein halbes Jahr zuvor gesehen. Das Thema wurde allerdings nicht vertiefend behandelt und es wurden keine digitalen Arbeitsblätter eingesetzt, sodass der Wiederholungsaspekt zu vernachlässigen ist.

### **Vergleichsgruppe**

In der Vergleichsgruppe haben insgesamt fünf Gymnasien und eine Gesamtschule teilgenommen. Keine der Schulen war zuvor in der Interventionsgruppe vertreten, sodass die Klassen und Kurse nichts von einer Interventionsgruppe beziehungsweise der Durchführung der Unterrichtsreihe im digitalen Format wussten. Die Verteilung der Lerngruppen zeigt sich in Tabelle 4.

*Tabelle 4: Anzahl der Klassen und Kurse, aufgeteilt nach Jahrgangsstufen und Schulform in der Vergleichsgruppe.*

Jahrgangsstufe	Anzahl der Klassen und Kurse	
	Gymnasium	Gesamtschule
9*	3	-
10*	5	-
11 GK	1	1
11 LK*	3	-
12 GK	-	-
12 LK	1	-
13 GK	-	-
13 LK	-	-

Die Stichprobe der Vergleichsgruppe bemisst 151 Schüler:innen in der Mittelstufe und 53 in der Oberstufe. Die Anzahl der Teilnehmer:innen variiert jedoch aufgrund von Fallausschlüssen, bei denen unter anderem nicht alle fehlende Werte imputiert werden konnten.

Bei den Stufen, die mit einem (\*) gekennzeichnet sind, liegt eine Besonderheit vor. Es handelte sich dabei jeweils um eine Lerngruppe einer G8-Schule. Formal handelte es sich bei zwei der drei neunten Klassen um achte Klassen. Diese werden allerdings dreistündig anstatt zweistündig unterrichtet. Nach Rücksprache mit der Lehrkraft, kann von einem Niveau auf Stufe der neunten Klasse ausgegangen werden. Bei einer 10. Klasse handelte es sich ebenfalls um eine Klasse einer G8-Schule. Formal wurde die Lerngruppe nach der G8-Regelung einem Grundkurs zugeordnet. Allerdings mussten alle Schüler:innen das Fach Chemie weiterhin wählen, um Ihren mittleren Schulabschluss zu erhalten. Der Umfang des Kurses ist mit dem einer regulären 10. Klasse gleichzusetzen, weshalb er den 10. Klassen zugeordnet wurde. Zuletzt ist auch ein Leistungskurs der elften Jahrgangsstufe von dieser Regelung betroffen. Es handelt sich dabei nach der G8-Rechnung noch um eine 10. Klasse. Bei dem Leistungskurs fand allerdings bereits eine Wahl als Leistungskurs statt. Außerdem wird der Kurs ebenfalls vierstündig unterrichtet, sodass hierbei von einem Leistungskursniveau ausgegangen werden kann. Auch bei der Vergleichsgruppe unterscheiden sich die absoluten Stichprobengrößen nach Art und Betrachtung der Auswertung beziehungsweise angewandten Methode. So wurden Schüler:innen aus dem Datensatz entfernt, wenn sie nicht an allen vier Unterrichtsstunden teilnahmen. Die Auswahl der Klassen erfolgte erneut nicht-randomisiert auf Anfrage über die jeweiligen Lehrkräfte. Es gab dabei keine besonderen Auswahlkriterien. Die Lerngruppen hatten auch bei der Vergleichsstudie kein besonderes Vorwissen zum behandelten Thema. Lediglich der Leistungskurs 12 hatte die Methode der Messung zu Beginn der 11. Schuljahres im Zuge einer Lehrer:innenfortbildung kurz kennengelernt. Aufgrund der verstrichenen Zeit von circa eineinhalb Jahren wurde die Gruppe dennoch in die Studie mitaufgenommen.

#### **4.2.2 Erhobene Variablen**

An dieser Stelle soll in Ergänzung zu dem Kapitel 4.3.2 auf die erhobenen Variablen der Studie eingegangen werden. Die Variablen leiten sich aus den Forschungsfragen, den Forschungslücken und dem Forschungsstand ab. Zu den jeweiligen Variablen wird eine kurze Begründung für deren Erhebung gegeben.

##### **- Interesse und Motivation**

Ziel dieser Arbeit ist unter anderem die Untersuchung des interesse- und motivationsfördernden Potentials von HyperDocs. Das Interesse oder auch Vergnügen gilt nach



der Self-Determination Theory bei Ausführung einer bestimmten Tätigkeit als ein Hauptindikator für die intrinsische Motivation<sup>6</sup>. Gleichzeitig bilden das Kompetenzerleben und die Autonomie wichtige Aspekte einer motivationsbezogenen Handlung. Die Autonomie wurde über die wahrgenommene Wahlfreiheit erfasst. Die Wahlfreiheit zeigt starke Zusammenhänge zur Autonomie (Patall et al., 2008, Schneider et al., 2018). Beeinflusst wird die intrinsische Motivation außerdem durch den Druck. Zusätzlich wurde die Anstrengung bestimmt, mit der die Schüler:innen lernen. Die soziale Eingebundenheit wurde nicht erhoben, da kein Einfluss der HyperDocs erwartet wurde. Neben den Facetten der intrinsischen Motivation beim Lernen mit HyperDocs wurde außerdem das Sachinteresse an den Naturwissenschaften und das Fachinteresse an Chemie miterhoben, um daraus Rückschlüsse auf das stabilere Interessenskonstrukt zu ziehen.

#### - **Usability**

Um die Forschungsfragen in Bezug auf die Usability beantworten zu können, muss diese nach der in der Arbeit verwendeten Definition erhoben werden. Dazu wurden, in Anlehnung an die ISO-Norm 9241-11, die Variablen „Zufriedenheit“, „Nützlichkeit“, „Einfachheit des Erlernens der Nutzung“ und „Einfachheit der Nutzung“ erhoben. Zusätzlich findet eine Evaluation der Eingabewerkzeuge (Tastatur, Stift) nach den eben genannten Attributen statt. Als weiterer Aspekt der Benutzerfreundlichkeit wurde die Bewertung der Darstellungsgröße der HyperDocs und Hilfen erhoben.

#### - **Lernhilfen und Vertiefungsaufgabe**

Bei den Lernhilfen und Vertiefungsaufgaben wurde die absolute Anzahl an Aufrufen sowie die Öffnungsdauer in der Interventionsgruppe ermittelt. Im weiteren Verlauf der Auswertung wurde die realistische Nutzungshäufigkeit, ermittelt aus der Öffnungsdauer und der Dauer zum vollständigen Lesen einer Hilfe, als Variable gebildet (vgl. Kapitel 4.1.1). Die Nutzungshäufigkeit der Hilfen und deren bevorzugte Modalität (visuell / verbal) wurde in der Vergleichsgruppe durch zwei Selbstauskünfte erhoben – direkt auf der Hilfe und auf dem Fragebogen. Eine zusätzliche Variable bildete die multimediale Repräsentation der Hilfen in der Interventionsgruppe. Außerdem wurden die Gründe für die Nutzung einer Hilfe und Vertiefungsaufgabe erfragt.

---

<sup>6</sup> Im Folgenden wird im Zusammenhang mit der SDT lediglich von „Interesse“ gesprochen, wobei das auch das Vergnügen miteinschließt.

Ebenfalls wurde die Nützlichkeit der Hilfen, die bevorzugte Art der Hilfenutzung (digital, analog) und die Häufigkeit des Hilfsangebots in der Schule erhoben, um ein vollständigeres Gesamtbild der Hilfenutzung zu erhalten.

- **Tabletbezogene Kompetenzen**

Die tabletbezogenen Kompetenzen umfassen die Selbstwirksamkeit im Umgang mit Tablets, die Angst vor den Tablets und die Nützlichkeit dieser zum Lernen. Dadurch werden Zusammenhänge zur Usability hergestellt. Gleichzeitig können diese Variablen als Erklärungsansatz für die Einordnung der Ergebnisse herangezogen werden.

- **Nutzung und Nutzungshäufigkeit von digitalen Endgeräten**

Aufgrund der bisherigen Literaturrecherche war zudem ein Einfluss der Nutzungshäufigkeit von digitalen Geräten auf die Usability zu erwarten. Daher wurde die Nutzungshäufigkeit von Smartphones, Tablets und PCs beim Lernen in der Schule und zuhause erhoben. Außerdem wurden physische Einschränkungen, beispielsweise die Ermüdung der Augen, durch die Verwendung des Tablets während der Untersuchung erfragt. Mögliche Zusammenhänge zur Usability wurden angenommen.

- **Leistungsniveau**

Das allgemeine, chemische und naturwissenschaftliche Leistungsniveau der Schüler:innen wurde durch eine subjektive Selbsteinschätzung erfragt. Auf eine objektive Erhebung mittels eines Leistungstests wurde verzichtet, um die umfangreiche Studie organisatorisch und zeitlich nicht weiter auszuweiten. Auf eine Auskunft der Noten wurde aus datenschutzrechtlicher Sicht verzichtet. Es wurde angenommen, dass das Leistungsniveau in Zusammenhang mit dem Interesse, der Nutzung von Lernhilfen und dem Cognitive Load steht.

- **Cognitive Load**

Beim Lernen kann die Gestaltung der Lernmaterialien oder der Lerngegenstand selbst das Arbeitsgedächtnis mehr oder weniger stark beanspruchen. Insbesondere der Extraneous Cognitive Load wird durch die eingesetzten Medien und deren Beschaffenheit beeinflusst. Zudem wurde der Mental Effort erhoben, der ebenfalls in das Konstrukt des Cognitive Load einzuordnen ist. Durch die Erhebung aller drei Loadtypen wird ein Vergleich zwischen den analogen und digitalen Arbeitsblättern erfolgen. Zusätzlich soll ein Einfluss auf die Nutzungshäufigkeit des Differenzierungsmaterials untersucht werden.

## 4.3 Hilfsmittel

Die eben genannten Variablen müssen durch geeignete Hilfsmittel erhoben werden. In diesem Unterkapitel werden die verwendeten Geräte (Technik) und Erhebungsinstrumente der Hauptstudie präsentiert. Dabei werden die Gütekriterien, bestehend aus der Reliabilität und Validität, dargestellt.

### 4.3.1 Technik

Die Hauptstudie wurde mit iPads des Modelljahres 2018 ( $n = 18$ ) und 2020 ( $n = 12$ ) durchgeführt. Die Modelle unterscheiden sich in geringem Maße. Das Display bei dem neueren Modell ist 0.5 Zoll (circa 0,89 cm) größer. Auf allen Endgeräten liefen iPad OS 13 und der Webbrowser Safari. Der Zugang zu HyperDocSystems wurde als Favorit gespeichert, sodass ein Eingeben der Webadresse oder das Scannen eines QR-Codes entfiel. Die iPads waren mit einem Apple Pencil der 1. Generation ausgestattet. Die Geräte und Stifte wurden vor jeder Nutzung auf Ihre Funktion hin geprüft und gegebenenfalls, bei geringem Akkustand, geladen, sodass ein Ausfall während den Stunden vermieden wurde. Aufgrund der Corona-Pandemie wurden außerdem alle iPads und Pencils vor jedem Gebrauch desinfiziert. Die Geräte erwiesen sich als robust gegenüber Ausfällen. Bezüglich HyperDocSystems kam ein Notebook der Marke HP (Modell: EliteBook 820 G3) mit Windows 10 und ein Netgearrouter (Modell: AC1750) zum Einsatz. Auf dem Laptop lief ein XAMPP-Server (Apache Friends, 2019). Es handelte sich dabei um einen Apache Webserver mit MariaDB als Datenbanksystem und PHP sowie Perl, wobei letztes beim Betrieb von HyperDocSystems keine Rolle spielte. Auch der Webserver erwies sich robust gegenüber Ausfällen. Für die Durchführung des Lehrer:innenversuchs wurden zwei Android-Tablets der Marke Samsung (Modell: Galaxy Tab A T580, Android-Version: 8) mit den Apps Screen Flashlight (Soriano, 2019) und Color Grab (Loomatix, 2019) verwendet. Diese beiden Geräte fanden auch als einzige in der Vergleichsgruppe zur Durchführung des Lehrer:innenexperiments Verwendung.

### 4.3.2 Erhebungsinstrumente

Um die verschiedenen Forschungsfragen beantworten zu können, mussten die Variablen durch geeignete Instrumente erhoben werden. In diesem Zusammenhang spielen die Gütekriterien der Fragebögen eine wichtige Rolle, damit die Daten valide, reliabel und objektiv erhoben werden können (Bortz & Döring, 2006, S. 195). Aus diesem Grund wurde

auf bereits evaluierte und erprobte Testinstrumente aus der Literatur zurückgegriffen, die die Variablen adäquat erfassen können. In den meisten Fällen mussten die Items, passend auf das eingesetzte Medium, beispielsweise „Digitale Arbeitsblätter“, umformuliert oder übersetzt werden. Für die erneute Prüfung der Reliabilität und Konstruktvalidität wurden alle Gruppen und Klassenstufen in einen Datensatz vereint. Die Bestätigung der Konstruktvalidität erfolgte über eine konfirmatorische Faktorenanalyse. Aufgrund der fehlenden multivariaten Normalverteilung vieler Items wurde ein *MLM*-Schätzer verwendet, der robuste Kennzahlen nennt (Rosseel, 2012). Die Daten stammten aus der Hauptstudie, da die Stichproben der Pilotierungen klein waren und die technische sowie didaktische Testung der Unterrichtsreihe dort im Vordergrund stand.

### **Cognitive Load**

Zur Erhebung des Cognitive Load (CL) wurde ein Instrument von Klepsch et al. (2017) eingesetzt. Es handelt sich dabei um acht verschiedene Fragen, die die drei Typen von Load (Intrinsic CL, Extraneous CL und Germane CL) erfassen. Dabei kam im ursprünglichen Instrument eine 7-stufige Likert-Skala von „völlig falsch“ bis „vollkommen richtig“ zum Einsatz. Dieses Instrument wurde in zwei Studien gewonnen. In der ersten Studie wurden 97 Student:innen der Informatik und Psychologie im ersten und dritten Semester in zwei Gruppen zugeteilt. Eine Gruppe erhielt einen Überblick über die Cognitive-Load-Theorie, während die andere Gruppe keine solche Einführung erhielt. Die Studierenden mussten 24 Aufgaben aus den fünf verschiedenen Bereichen Biologie, Sprache, Mathematik, Technologie und Didaktik lösen. Der Schwierigkeitsgrad der Aufgaben wurde dabei variiert. Von Interesse war hierbei insbesondere die uninformierte naive Gruppe, da Schüler:innen ebenfalls keine Einführung in diese Theorie nach Sweller erhalten. Für alle drei Skalen konnten hohe Reliabilitäten zwischen Cronbachs  $\alpha = 0.80 - 0.86$  erreicht werden (Klepsch et al., 2017). Weiterhin konnte eine hohe Reliabilität mit dem mentalen Cognitive-Load-Rating nach Paas (1992) gemessen werden. Zudem wurde geprüft, ob eine leichtere Aufgabe zu einer geringen Bewertung der kognitiven Last führte. Das war für die Sub-Skalen des Intrinsic und Extraneous Load der Fall, nicht jedoch für den Germane Load. Es wurde daher eine zweite Studie durchgeführt. Es hatten insgesamt 65 von 95 Studierenden der Psychologie oder Informatik im ersten und zweiten Semester alle vorgegebenen Aufgaben vervollständigt. Als Ergänzung zur ersten Studie wurden die Items für den Germane Cognitive Load neu formuliert. Dabei konnten folgende Reliabilitäten erreicht werden:  $\alpha_{ICL} = 0.81$ ,  $\alpha_{ECL} = 0.86$ ,

$\alpha_{GCL} = 0.67$ . Außerdem wurde eine konfirmatorische Faktorenanalyse durchgeführt, wobei ein Modell gewählt wurde, bei dem ein Zusammenhang zwischen allen drei Loadtypen besteht (Klepsch et al., 2017). Der Modell-Fit kann mit  $CFI = 0.970$  und  $RMSEA = 0.021$  als gut erachtet werden. Die Autoren schließen aus ihrer Studie, dass es sich bei dem Instrument um eine einfache und effiziente Art der Erhebung der verschiedenen Loadtypen handelt. Den Vorteil gegenüber anderen Instrumenten (z. B. Leppink et al., 2013) sehen sie, wenn keine spezifischen Themen oder Konzepte für die Formulierung der Items benannt werden können. Zusätzlich zu dem hier eingesetzten Instrument wurde auch das Instrument angelehnt an Paas (1992), formuliert nach Pouw et al. (2016), übersetzt und eingesetzt, welches generell die mentale Anstrengung (*engl. Mental Effort*) erhebt. Im Gegensatz zu den anderen Items wurde bei dieser Frage eine 7-polige Skala mit den Endpunkten „sehr gering“ bis „sehr hoch“ verwendet. Auf eine Benennung der Zwischenkategorien wurde verzichtet.

In der eigenen Studie wurde für das Instrument nach Klepsch et al. (2017) eine 5-stufige Likert-Skala mit den Polen „stimme nicht zu“ bis „stimme zu“ verwendet, um eine homogene Anzahl an Antwortmöglichkeiten und Polbeschriftungen zu den anderen Skalen zu erreichen. Zusätzlich wurden einige Items umformuliert, die sich lediglich auf eine Aufgabe bezogen, die eigene Erhebung jedoch mehrere Aufgaben beinhaltet: „Bei diesen Aufgaben musste man viele Dinge gleichzeitig im Kopf bearbeiten“. Die Formulierung und Reihenfolge der Fragen wurde über die Erhebungsinstrumente nicht geändert. Die Skalen zum Cognitive Load fanden somit bei den Fragebögen *POST\_1\_E*, *POST\_1\_2*, *POST\_1\_3* und *POST\_1\_4* (Interventionsgruppe) sowie *POST\_1\_EV*, *POST\_1\_2V*, *POST\_1\_3V* und *POST\_1\_F* (Vergleichsgruppe) Anwendung. Kursiv geschriebene Sätze wurden aufgrund der Reliabilität oder Faktorenstruktur entfernt. Die Nummern geben die Reihenfolge im Fragebogen an und sind im Faktorenmodell hinterlegt:

*Intrinsic Load:*

- Diese Aufgaben waren sehr schwierig. (CL1)
- Bei den Aufgaben musste man viele Dinge gleichzeitig im Kopf bearbeiten. (CL4)

*Extraneous Load:*

- Bei diesen Aufgaben ist es schwer, die zentralen Inhalte miteinander in Verbindung zu bringen. (CL2)

- Die Darstellung bei diesen Aufgaben ist ungünstig, um wirklich etwas zu lernen. (CL5)
- Bei diesen Aufgaben ist es mühsam, die wichtigsten Informationen zu erkennen. (CL8)

*Germane Load:*

- Es ging mir beim Bearbeiten der Lerneinheit darum, alles richtig zu verstehen. (CL3)
- Ich habe mich angestrengt, mir nicht nur einzelne Dinge zu merken, sondern auch den Gesamtzusammenhang zu verstehen. (CL6)
- Die Lerneinheit enthielt Elemente, die mich unterstützten, den Lernstoff besser zu verstehen. (CL7)

*Mental Effort:*

- Wie war deine mentale geistige Anstrengung während der Bearbeitung des (digitalen) Arbeitsblattes?

Zur Prüfung der Güte des Instruments wurden die Reliabilitäten nach Cronbachs  $\alpha$  bestimmt und eine konfirmatorische Faktorenanalyse durchgeführt. Dazu wurde jeweils der gesamte Datensatz des ersten Post-Tests (*POST\_1\_E* und *POST\_1\_EV*) verwendet. Die Reliabilitäten sind Tabelle 5 zu entnehmen.

*Tabelle 5: Reliabilitäten der Skalen des Cognitive Load (n = 629).*

<i>Skala</i>	<i>Cronbachs <math>\alpha</math></i>	<i>Anzahl Items</i>
Intrinsic Load	0.39	2
Extraneous Load	0.54	2
Germane Load	0.66	3

Die Reliabilitäten für den Extraneous CL und den Germane CL liegen damit etwas unter dem akzeptablen Bereich. Nach Klinke (1999) sollten Werte niedriger als 0.70 vermieden werden. Insbesondere der Subtyp Intrinsic CL weist eine unzureichende Reliabilität von  $\alpha = 0.39$  auf. Die Gründe hierfür können vielseitig sein. Einerseits wurde dieser Loadtyp mit nur zwei Items bestimmt, was zu einer niedrigeren Reliabilität führen kann, wenn sich die Items stärker unterscheiden beziehungsweise die Korrelation zwischen den Items geringer ist (Cortina, 1993). Dennoch sind die gemessenen Werte für den intrinsischen Load womöglich fehlerbehaftet und die Items messen womöglich nicht denselben Faktor. Gleichzeitig können Items jedoch auch zwischenkorreliert sein, was einem hohen Wert

von Alpha entspricht, und dabei multidimensional sein, wodurch ein hoher Alpha-Wert nicht zwangsläufig ein ausreichend reliables Instrument zeigt (Cortina, 1993). Ein weiterer Grund könnte in der Formulierung des ersten Items liegen, wonach die Aufgaben „sehr schwierig“ waren. Gegebenenfalls sorgte dieser Superlativ ebenfalls für eine größere Abweichung in der Beantwortung der beiden Fragen zum Intrinsic Load. Insgesamt ist anzumerken, dass die subjektive Erfassung und die Unterteilung in verschiedene Loadtypen durch Fragebögen als nicht optimal gilt (de Jong, 2010). Es handelt sich jedoch, im Gegensatz zu anderen Verfahren, um eine sehr kosten- und zeitgünstige sowie praxistaugliche Variante für den Einsatz im Regelunterricht. Beim Extraneous Load ist ebenfalls eine niedrige Reliabilität von  $\alpha = 0.54$  zu beobachten. Die Reliabilität unter Beachtung der ursprünglichen Items betrug  $\alpha = 0.61$ . Aufgrund der deutlichen Verbesserung des Fits des Faktorenmodells wurde das Item „Die Darstellung bei diesen Aufgaben ist ungünstig, um wirklich etwas zu lernen“ entfernt. Die Modelle werden in Tabelle 6 verglichen. Modell 1 bildet das ursprüngliche, nicht verschachtelte, Modell. Dabei zeigt sich ein schlechter Fit bei allen Indices. Das Modell (2) konnte wesentlich verbessert werden, indem das oben beschriebene Item aus dem Modell entfernt wurde. Grund dafür könnte sein, dass die Darstellung bei den Aufgaben nicht richtig beurteilt werden konnte. Es ist anzumerken, dass der  $\chi^2$ -Test ein signifikantes Ergebnis zeigt und damit das Modell verworfen werden sollte (vgl. Tabelle 6). Der Test ist allerdings sensitiv gegenüber der Stichprobengröße, wodurch die Summe aller Fitindices zur Beurteilung des Modells herangezogen werden sollten (Schermelleh-Engel et al., 2003; Wheaton et al., 1977). Nach Hu und Bentler (1999) liegen die betrachteten Indices auf einem guten Niveau. Die Einordnung der nachfolgenden Faktorenmodelle erfolgt nach den gleichen Kriterien. Die Modelle sind in Anhang 9.1.2 aufgeführt.

*Tabelle 6: Kennzahlen der verschiedenen Faktorenmodelle (n = 629). Die Kennzahlen werden als robuste Werte angegeben.*

<i>Modell</i>	$\chi^2$	<i>df</i>	<i>p</i>	<i>CFI</i>	<i>TLI</i>	<i>RMSEA</i>	<i>SRMR</i>
1	81.20	17	< 0.001	0.906	0.845	0.084	0.060
2	32.80	11	0.001	0.964	0.932	0.057	0.038

### **Tabletbezogene Kompetenzen**

Zur Erhebung der tabletbezogenen Kompetenzen wurde ein Selbsteinschätzungsinstrument nach Bescherer und Spannagel (2011) angepasst, indem der Begriff „Computer“ durch „Tablet“ ersetzt wurde. Es handelt sich dabei um eine Kurzversion des CUSE-D,

der von Spannagel und Bescherer (2009) aus der ursprünglichen Version des CUSE (Computer User Self-Efficacy Scale; Cassidy und Eachus, 2002) übersetzt und validiert wurde. Die Autor:innen beschreiben dabei das „Zutrauen in die eigene Kompetenz im Umgang mit Computern“ (Spannagel & Bescherer, 2009, S. 24) als „computerbezogene Selbstwirksamkeit“ (Spannagel & Bescherer, 2009, S. 24). Die zugrundeliegende Theorie geht auf das Selbstwirksamkeitskonzept von Bandura (1977, 1997) zurück (Spannagel & Bescherer, 2009). Diese computerbezogene Selbstwirksamkeit nimmt demnach einen Einfluss auf die Nutzung von Computern und auf den Umgang mit Problemen, die bei dieser Nutzung auftreten. Ferner wird beschrieben, dass sich Personen, die sich kompetent im Umgang mit Computern fühlen, auch eher in neue Software einarbeiten können (Spannagel & Bescherer, 2009). Die Autor:innen hatten für ihre Studie zum Erfassen dieser computerbezogenen Selbstwirksamkeit folgende Anforderungen an das Instrument gestellt: Aktualität, Programmunabhängigkeit und Bezug zur Lernsituation. Es sollten damit Fragen formuliert werden, die keine alten Fachwörter (Aktualität) beinhalten, die unabhängig vom genutzten Programm (Programmunabhängigkeit) und die insbesondere in Lernsituationen eingesetzt werden können (Bezug zur Lernsituation). Das übersetzte Instrument (CUSE-D) wurde in einer Studie mit  $n = 407$  Lehramtstudierenden (weiblich: 312, männlich: 95) zweier Hochschulen evaluiert. Weiterhin wurde der gewonnene Fragebogen an 172 Realschüler:innen (weiblich: 92, männlich: 80) der 8. Klasse ausprobiert. Die evaluierte Form des CUSE-D wurde anschließend von Bescherer und Spannagel (2011) mithilfe einer Hauptkomponentenanalyse in eine Kurzversion (CUSE-D-r) mit insgesamt zwölf Fragen überführt. Es bildeten sich dabei drei Subskalen mit folgenden Reliabilitäten ab: Die Computerwirksamkeit selbst ( $\alpha = 0.88$ ), die Computerangst ( $\alpha = 0.60$ ) und Computer beim Lernen ( $\alpha = 0.86$ ). Wachsmuth (2019) konnte ebenfalls gute Reliabilitäten zwischen  $\alpha = 0.75 - 0.81$  erzielen. Allerdings wurde dort die Subskala Computerangst aufgrund der sehr geringen Reliabilität von  $\alpha = 0.22$  entfernt.

Im ursprünglichen Instrument wurde eine 6-polige Likert-Skala von „trifft überhaupt nicht zu“ bis „trifft völlig zu“ eingesetzt (Bescherer & Spannagel, 2011). Für die eigene Studie wurde, wie auch schon bei den anderen Instrumenten, eine 5-polige Likert-Skala mit den Endpunkten „stimme nicht zu“ bis „stimme zu“ eingesetzt, um eine homogene Fragebogenstruktur zu erreichen. Nachfolgend sind die Aussagen des Erhebungsinstruments gezeigt. Kursiv geschriebene Sätze wurden aufgrund der Reliabilität oder Faktorenstruktur entfernt. Die Nummern geben die Itemreihenfolge wieder und sind im



Faktorenmodell hinterlegt. Kursiv geschriebene Fragen wurden aufgrund der Reliabilität oder Faktorenanalyse entfernt.

*Tabletselbstwirksamkeit:*

- Ich kann normalerweise mit den meisten Schwierigkeiten umgehen, auf die ich während der Benutzung eines Tablets stoße. (TS1)
- Ich bin sehr unsicher über meine Fähigkeiten im Umgang mit Tablets. (TS2)
- Ich finde es schwierig, Tablets dazu zu bringen, das zu tun, was ich von ihnen will. (TS4)
- Ich halte mich selbst für einen geschickten Tabletnutzer. (TS6)
- *Tabletbegriffe verwirren mich.* (TS7)
- Wenn ich ein Tablet verwende, dann scheinen manchmal Dinge einfach so zu passieren, und ich weiß nicht warum. (TS9)
- Was Tablets anbelangt, halte ich mich selbst für nicht sehr geübt. (TS12)

*Tabletangst:*

- *Tablets machen mir Angst.* (TS3)
- *Wenn ich ein Tablet benutze, dann befürchte ich, dass ich eine falsche Taste drücken und es beschädigen könnte.* (TS11)

*Tablets beim Lernen:*

- Das Verwenden von Tablets macht Lernen interessanter. (TS5)
- Tablets sind gute Hilfsmittel beim Lernen. (TS8)
- Einige Tabletprogramme (Apps) machen Lernen eindeutig einfacher. (TS10)

Die Analyse der eigenen Daten ergibt die nachfolgenden Reliabilitäten (vgl. Tabelle 7).

*Tabelle 7: Reliabilitäten der Skalen der tabletbezogenen Kompetenzen mit dem angepassten Erhebungsinstrument (n = 377).*

<i>Skala</i>	<i>Cronbachs <math>\alpha</math></i>	<i>Anzahl Items</i>
Tabletselbstwirksamkeit	0.76	6
Tabletangst	0.23	2
Tablets beim Lernen	0.84	3

Aufgrund der geringen Reliabilität wurde die Skala *Tabletangst* verworfen. Ein Grund für den niedrigen Cronbachs  $\alpha$ -Wert könnte in dem Item „Tablets machen mir Angst“ liegen, wonach die wenigsten Schüler:innen wirklich Angst vor der Verwendung der Geräte haben dürften. Die originalen Items stammen aus 2002 beziehungsweise 2009. In

dieser Zeit hatten Computer einen anderen Stellenwert als heute. Das Faktorenmodell (Model 1) mit dem Faktor „Tabletangst“ zeigt zudem lediglich einen akzeptablen Fit (vgl. Tabelle 8). Das Modell verbesserte sich, nachdem der Faktor entfernt wurde (Modell 2). Dabei liegen die Fitindices jedoch immer noch nicht in einem guten Bereich. Aus diesem Grund wurden die Items genauer analysiert und das Item „Tabletbegriffe verwirren mich“ aufgrund seiner unspezifischen Formulierung entfernt. Dadurch konnte das Modell nochmals wesentlich verbessert werden (Modell 3). Die Modelle sind in Anhang 9.1.2 aufgeführt.

*Tabelle 8: Kennzahlen der verschiedenen Faktorenmodelle (n = 377). Die Kennzahlen werden als robuste Werte angegeben.*

<i>Modell</i>	$\chi^2$	<i>df</i>	<i>p</i>	<i>CFI</i>	<i>TLI</i>	<i>RMSEA</i>	<i>SRMR</i>
1	104.36	51	< 0.001	0.949	0.935	0.055	0.050
2	58.36	34	0.006	0.973	0.965	0.046	0.045
3	41.64	26	0.027	0.981	0.974	0.042	0.041

Die tabletbezogenen Kompetenzen wurde in den Fragebögen PRE und POST\_2 erhoben. Die Fragen sollten im PRE-Fragebogen lediglich beantwortet werden, wenn die Schüler:innen bereits einmal ein Tablets genutzt hatten.

### **Usability**

Zur Erhebung der technischen Usability wurde der USE verwendet (Lund, 2001). Das Instrument besitzt vier Subskalen (*Usefulness*, *Satisfaction*, *Ease of Use* und *Ease of Learning*) und besteht aus 30 Fragen. Nach Lund lassen sich die beiden Skalen „Ease of Use“ und „Ease of Learning“ auch zu einer Skala zusammenfassen. Der Fragebogen wurde für verschiedene Domänen entwickelt. Dazu wählte der Autor mögliche Fragen aus verschiedenen Studien aus und entwickelte selbst weitere Items. Dabei legte er besonderen Wert auf generelle Formulierungen und eine einfache Sprache. Nach mehreren Validierungen entstand das endgültige Instrument mit hohen Werten von Cronbachs  $\alpha$ , die allerdings in der originalen Publikation nicht beschrieben sind. Lediglich Items mit niedrigen Faktorladungen werden durch den Autor gekennzeichnet. Als Antwortskala kam eine 7-punktige Likert-Skala mit den Polen „strongly disagree“ bis „strongly agree“ zum Einsatz. Der USE wurde seitdem in mehreren Studien eingesetzt, um die Usability von Anwendungen oder Systemen zu testen. Bei einer Studie mit 25 Lehrkräften zur Prüfung der Usability einer Webanwendung erreichten Hendra et al. (2018) eine Reliabilität

aller Fragen von  $\alpha = 0.915$ . Hariyanto et al. (2020) nutzten den USE in einer Studie mit 62 High-School Schüler:innen zur Evaluation eines adaptiven E-Learningsystems und erreichten folgende Reliabilitäten:  $\alpha_{Usefulness} = 0.770$ ,  $\alpha_{Ease\ of\ Use} = 0.792$ ,  $\alpha_{Ease\ of\ Learning} = 0.717$  und  $\alpha_{Satisfaction} = 0.703$ . Etwas bessere Reliabilitäten erreichten Huang et al. (2013) bei einer Studie zu einem interaktiven E-Book mit 166 Schüler:innen der Elementary School:  $\alpha_{Usefulness} = 0.88$ ,  $\alpha_{Ease\ of\ Use} = 0.86$ ,  $\alpha_{Ease\ of\ Learning} = 0.88$  und  $\alpha_{Satisfaction} = 0.87$ . Die Reliabilität des gesamten Instruments liegt bei  $\alpha = 0.96$ . Gao et al. (2018) konnten in einer vergleichenden Studie mit 151 Personen ebenfalls hohe Reliabilitäten von  $\alpha = 0.88 - 0.95$  erreichen. Gleichzeitig zeigten sich hohe Korrelationen zum SUS-Fragebogen, der ebenfalls in der Studie Anwendung fand. In einer konfirmatorischen Faktorenanalyse konnte allerdings nur ein mittelmäßiges Modell mit der ursprünglichen Struktur nach Lund (2001) errechnet werden ( $\chi^2(399) = 1198.7$ ,  $p < 0.001$ ). Wallace und Yu (2009) konnten ebenfalls sehr hohe Reliabilitäten von  $\alpha = 0.86 - 0.96$  für alle Subskalen erzielen.

Aufgrund der einfachen und kurzen Formulierungen sowie der Unterteilung in die vier Subskalen wurde der USE für die Erhebung der Usability ausgewählt. Dazu wurden je Subskala drei bis vier Items mit hohen Faktorenladungen nach Lund (2001) gewählt, um ein kürzeres Instrument zu schaffen. Die Items wurden wörtlich übersetzt. Als Gegenstand der Items wurde die Formulierung „Digitale Arbeitsblätter“ gewählt, da der Begriff „HyperDocs“ den Schüler:innen gegebenenfalls nicht geläufig sein könnte. Wie auch bei den anderen Instrumenten wurde eine 5-polige Likert-Skala mit den Polen „stimme nicht zu“ bis „stimme zu“ gewählt. Für die Skalen wurden folgende Reliabilitäten bestimmt (vgl. Tabelle 9).

*Tabelle 9: Reliabilitäten der Skalen der Usability mit dem angepassten Erhebungsinstrument (n = 407).*

<i>Skala</i>	<i>Cronbachs <math>\alpha</math></i>	<i>Anzahl Items</i>
Nützlichkeit (Usefulness)	0.89	3
Zufriedenheit (Satisfaction)	0.70	2
Einfachheit der Nutzung (Ease of Use)	0.74	3
Einfachheit des Erlernens der Nutzung (Ease of Learning)	0.71	3

Die Konstruktvalidität konnte über eine Faktorenanalyse bestätigt werden. Das erste Modell, bestehend aus allen Items und vier Faktoren, zeigte den schlechtesten Fit. Daraufhin wurden Items entfernt, die eine niedrige Faktorenladung aufwiesen und nur einen geringen Einfluss auf die Reliabilität ausübten. Ein Item zeigte zudem im Nachhinein eine für Schüler:innen unter Umständen schwer verständliche Formulierung, wie beispielsweise „Die digitalen Arbeitsblätter sind bedienungsfreundlich“ (Ease of Use). Das betraf ein Item bei „Satisfaction“, „Usefulness“ und „Ease of Use“. Dadurch konnte ein gutes Modell (Modell 5) gewonnen werden.

Daraufhin wurden die Anmerkungen von Lund (2001) überprüft, wodurch auch die Faktoren „Ease of Use“ und „Ease of Learning“ als gemeinsamer Faktor verstanden werden können. In den eigenen Daten zeigte sich dieser Zusammenhang unter anderem an der hohen Korrelation der beiden Faktoren zueinander. Eine Zusammenlegung (Modell 4) führte allerdings zu einem schlechteren Modell. Danach wurden zwei Modelle mit jeweils einem Faktor (Modell 2: Ease of Learning; Modell 3: Ease of Use) gebildet. Die Modelle zeigten einen sehr guten Fit. Da in der vorliegenden Stunde allerdings beide Faktoren untersucht werden sollten und der Fit des Modells als zufriedenstellend eingestuft wurde, wurde das Modell 5 gewählt (vgl. Tabelle 10). Die Modelle sind in Anhang 9.1.2 aufgeführt.

*Tabelle 10: Kennzahlen der verschiedenen Faktorenmodelle (n = 407). Die Kennzahlen werden als robuste Werte angegeben.*

<b>Modell</b>	$\chi^2$	<b>df</b>	<b>p</b>	<b>CFI</b>	<b>TLI</b>	<b>RMSEA</b>	<b>SRMR</b>
1	166.50	71	< 0.001	0.946	0.931	0.073	0.050
2	24.78	17	0.100	0.992	0.987	0.041	0.027
3	28.24	17	0.042	0.988	0.980	0.051	0.031
4	68.67	41	0.004	0.978	0.971	0.052	0.042
5	55.26	38	0.035	0.987	0.981	0.042	0.035

Nachfolgend sind die Items des Erhebungsinstruments gezeigt.

*Usefulness:*

- Die digitalen Arbeitsblätter helfen mir im Vergleich zu normalen Arbeitsblättern, Aufgaben effektiver zu bearbeiten. (US4)
- Die digitalen Arbeitsblätter erleichtern mir im Vergleich zu normalen Arbeitsblättern, Aufgaben zu bearbeiten. (US7)

- Die digitalen Arbeitsblätter helfen mir im Vergleich zu normalen Arbeitsblättern, mehr Aufgaben zu bearbeiten. (US9)
- *Die digitalen Arbeitsblätter sind nützlich für mich, um Aufgaben zu bearbeiten.* (US14)

*Satisfaction:*

- Die digitalen Arbeitsblätter sind großartig. (US2)
- *Ich bin zufrieden mit den digitalen Arbeitsblättern.* (US5)
- Es hat Spaß gemacht, die digitalen Arbeitsblätter zu benutzen. (US10)

*Ease of Use:*

- *Die digitalen Arbeitsblätter sind bedienungsfreundlich.* (US1)
- Die digitalen Arbeitsblätter sind einfach für mich zu verwenden. (US6)
- Ich kann mir einfach merken, wie ich die digitalen Arbeitsblätter benutze. (US11)
- Ich kann mit den digitalen Arbeitsblättern in wenigen Schritten Aufgaben bearbeiten. (US13)

*Ease of Learning:*

- Ich habe schnell gelernt, die digitalen Arbeitsblätter zu nutzen. (US3)
- Es ist einfach zu erlernen, wie ich die digitalen Arbeitsblätter nutze. (US8)
- Ich kann mir einfach merken, wie ich die digitalen Arbeitsblätter benutze. (US12)

Zusätzlich zu dem Instrument nach Lund (2001), wurden noch zwei weitere Fragebatterien zu der Eingabe mittels Tastatur und mittels Apple Pencil erhoben. Die Blöcke enthielten je eine Frage aus den Subskalen des USE, sodass pro Eingabemethode vier Aussagen gemacht wurden. Die Fragen sollten lediglich beantwortet werden, wenn die Schüler:innen die jeweilige Texteingabe genutzt hatten.

*Usability Texteingabe:*

- Die handschriftliche Texteingabe mit dem Stift ist nützlich.
- Die handschriftliche Texteingabe mit dem Stift ist einfach zu verwenden.
- Ich habe schnell gelernt, die handschriftliche Texteingabe mit dem Stift zu verwenden.
- Ich bin zufrieden mit der handschriftlichen Texteingabe über den Stift.

Die Formulierung für die Texteingabe mit der Tastatur war analog. Je zwei zusätzliche Aussagen wurden für die Lesbarkeit der Arbeitsblätter und Hilfen formuliert.

*Lesbarkeit HyperDocs:*

- Die Darstellung der Inhalte auf den digitalen Arbeitsblättern ist ausreichend groß.
- Die Inhalte auf den digitalen Arbeitsblättern sind gut erkennbar.

*Lesbarkeit Hilfen:*

- Die Darstellung der Inhalte auf den Hilfen ist ausreichend groß.
- Die Inhalte auf den Hilfen waren gut erkennbar.

Die Aussagen zur Benutzerfreundlichkeit wurden in den Fragebogen POST\_1\_E und POST\_2 erhoben.

### **Intrinsische Motivation und Interesse**

Die intrinsische Motivation wurde über das sogenannte *Intrinsic Motivation Inventory (IMI)* erhoben. Das Instrument wurde in Studien der Autoren der Selbstbestimmungstheorie entworfen. Insgesamt besteht es aus sechs Subskalen, wovon jedoch für diese Arbeit lediglich die Subskalen Interesse, wahrgenommene Kompetenz, wahrgenommene Wahlfreiheit (Autonomie), Druck und Anstrengung genutzt wurden. Die wahrgenommene Kompetenz wird im Folgenden als Kompetenzerleben bezeichnet. Die Konstruktvalidität sollte nach den Autoren dadurch nicht verringert werden (CSDT, o. D.). Das Interesse bildet dabei den Hauptindikator für die intrinsische Motivation. Jede Skala ist durch fünf bis acht Items im originalen Instrument abgedeckt. Ab vier Items pro Skala wird von keiner merklichen Verbesserung der Reliabilität berichtet (CSDT, o. D.; McAuley et al., 1989). Es wird außerdem vorgeschlagen, den Begriff „activity“ durch einen, für die Studie passenden, Begriff zu ersetzen (CSDT, o. D.). Aus den oben genannten Gründen wurden daher pro Subskala vier Items identifiziert, die eine einfache Formulierung aufwiesen. Damit sollte das Instrument nicht zu viele Fragen umfassen, die die Schüler:innen ermüden könnten. Die Auswahl der Items wurde auch Lehrkräften vorgelegt, um die Items auf Verständnisschwierigkeiten in Bezug auf die Schüler:innen zu prüfen.

Eine der ersten Validitätsstudien untersuchte das Instrument bestehend aus dem Interesse, dem Kompetenzerleben, der Anstrengung und dem Druck. 116 Sportstudierende (weiblich: 36, männlich: 80; Durchschnittsalter: 21.35) nahmen an einem Seminar teil. Dabei zeigte sich ein zufriedenstellendes Faktorenmodell, das durch Entfernen zweier Items weiter verbessert werden konnte ( $\chi^2(101) = 252.36, p < 0.05, RMSR = 0.136$ ; McAuley et al., 1989). Die Reliabilität wurde nicht bestimmt. Andere Studien mit Schüler:innen und Student:innen bestätigten das Faktorenmodell und hohe Reliabilitäten (Ostrow & Heffernan, 2018; Monteiro et al., 2015; Markland & Hardy, 1997; Gibbens, 2019). In

einer Studie von Tsigilis und Theodosiou (2003) oder auch Bätz et al. (2009) wurde eine geringere Reliabilität ( $\alpha = 0.60 - 0.66$ ) der Subskala „Druck“ gefunden. In einer anderen Studie wurde eine höhere Reliabilität ( $\alpha = 0.817$ ) erzielt (Markland & Hardy, 1997).

In der eigenen Studie wurden gute bis akzeptable Reliabilitäten gefunden (vgl. Tabelle 11). Die Subskalen „Druck“ und „Anstrengung“ weisen eine geringe Reliabilität auf.

*Tabelle 11: Reliabilitäten der Subskalen der intrinsischen Motivation mit dem angepassten Erhebungsinstrument (n = 634).*

<i>Skala</i>	<i>Cronbachs <math>\alpha</math></i>	<i>Anzahl Items</i>
Interesse / Spaß (Interest / Enjoyment)	0.91	4
Kompetenzerleben (Perceived Competence)	0.65	4
Wahrgenommene Wahlfreiheit (Perceived Choice)	0.74	4
Druck (Pressure / Tension)	0.52	4
Anstrengung (Effort / Importance)	0.47	4

Aufgrund der aktuellen Nutzungsbestimmungen des Intrinsic Motivation Inventory dürfen die gesamten Fragen an dieser Stelle nicht veröffentlicht werden. Aus diesem Grund wird auf das Originaldokument verwiesen (CSDT, o. D.) und nachfolgend die Itemnummer pro Subskala genannt. Für jede Skala beginnt die Nummerierung erneut. Invertierte Items sind mit einem (I) gekennzeichnet (vgl. Tabelle 12).

*Tabelle 12: Verwendete Items zur Erfassung der intrinsischen Motivation über das Intrinsic Motivation Inventory.*

<i>Subskala</i>	<i>Itemnummer</i>
Interesse / Spaß (Interest / Enjoyment)	2, 3 (I), 5, 7
Kompetenzerleben (Perceived Competence)	1, 3, 4, 6
Wahrgenommene Wahlfreiheit (Perceived Choice)	1, 2 (I), 3 (I), 7 (I)
Druck (Pressure / Tension)	1, 2 (I), 3, 5 (I)
Anstrengung (Effort / Importance)	2 (I), 3, 4, 5

Die Items wurden ins Deutsche übersetzt und der Begriff „activity“ wurde beispielsweise durch „...beim Bearbeiten des digitalen Arbeitsblattes...“ (Interventionsgruppe) oder „...beim Bearbeiten des Arbeitsblattes...“ (Vergleichsgruppe) ersetzt. Es wurde eine 5-polige Likert-Skala mit den Polen „stimme nicht zu“ bis „stimme zu“ gewählt. Korner et al. (2012) machen darauf aufmerksam, dass die Itempolarität bei Verneinungen Schüler:innen vor Probleme stellt. Diese Problematik konnte auch in den Pilotierungen festgestellt werden. Aufgrund der dort genannten Gründe wurde allerdings die Verneinung beibehalten. Die Ergebnisse der Faktorenanalyse sind in Tabelle 13 dargestellt. Die Modelle sind Anhang 9.1.2 zu entnehmen. Das erste Modell beinhaltet alle fünf Faktoren und zeigt einen schlechten Fit. Das Modell konnte wesentlich verbessert werden, indem der Faktor „Anstrengung“ daraus entfernt wurde (Modell 2).

*Tabelle 13: Kennzahlen der verschiedenen Faktorenmodelle (n = 634). Die Kennzahlen werden als robuste Werte angegeben.*

<b>Modell</b>	$\chi^2$	<b>df</b>	<b>p</b>	<b>CFI</b>	<b>TLI</b>	<b>RMSEA</b>	<b>SRMR</b>
1	396.47	160	< 0.001	0.931	0.918	0.052	0.063
2	216.30	98	< 0.001	0.960	0.950	0.049	0.055

Zusätzlich zur intrinsischen Motivation wurde außerdem das Fachinteresse Chemie und das Sachinteresse an Naturwissenschaften erhoben. Die Items stammen von Engeln (2004, Sachinteresse - gekürzt) beziehungsweise Pawek (2009, Fachinteresse - umformuliert), der die Items von Engeln anpasste. Dabei zeigten sich hohe Reliabilitäten von  $\alpha_{\text{Fachinteresse}} = 0.86 - 0.88$  und  $\alpha_{\text{Sachinteresse}} = 0.86 - 0.88$  (Pawek, 2009). Diese Reliabilitäten konnten für die eigene Studie bestätigt werden (vgl. Tabelle 14).

*Tabelle 14: Reliabilitäten der Skala Fachinteresse Chemie und Sachinteresse Naturwissenschaften (n = 588).*

<b>Skala</b>	<b>Cronbachs <math>\alpha</math></b>	<b>Anzahl Items</b>
Fachinteresse	0.81	2
Sachinteresse	0.83	2

Die Aussagen für das Fachinteresse Chemie und das Sachinteresse Naturwissenschaften lauten wie folgt:

*Fachinteresse Chemie:*

- Der Unterricht im Fach Chemie bringt mir Spaß.
- Ich komme im Unterricht im Fach Chemie gut mit.



*Sachinteresse Naturwissenschaften:*

- Naturwissenschaften bringen mir Spaß.
- Naturwissenschaften gehören für mich persönlich zu den wichtigen Dingen.

Das Fachinteresse an Chemie und das Sachinteresse an Naturwissenschaften wurden in den Fragebogen PRE, PRE\_V, POST\_2 und POST\_F erhoben. Das IMI bei den Fragebogen POST\_1\_E, POST\_1\_V, POST\_1\_2, POST\_1\_2V, POST\_1\_3, POST\_1\_3V, POST\_1\_4 und POST\_1\_F.

**Hilfenutzung und Vertiefungsaufgaben**

Die Erfassung der Nutzungshäufigkeiten der Hilfen und Vertiefungsaufgaben erfolgte in der Interventionsgruppe vollautomatisch über HyperDocSystems. In der Vergleichsgruppe erfolgte die Erfassung über eine Selbstauskunft auf den jeweiligen Kärtchen und auf dem Fragebogen. Es wurde beobachtet, dass die Selbstauskunft auf den Kärtchen nicht zuverlässig war, da die Schüler:innen häufig daran erinnert werden mussten und manche Lernenden keine Kreuze setzten. Für die weitere Auswertung wird daher die Angabe auf dem Fragebogen verwendet. In diesem Zusammenhang sollten die Schüler:innen auch angeben, ob sie eher die verbale oder eher die visuelle Hilfe verwendeten. Die Nutzung der Vertiefungsaufgaben wurde nicht erfasst. In beiden Gruppen wurde zudem erfragt, ob die Hilfen als nützlich erachtet wurden und aus welchen Gründen sie die Hilfen und Vertiefungsaufgaben nutzten. Die Fragen dazu wurden im Vorfeld der Erhebung entworfen und leiten sich aus der Arbeit von Fitting (2018) ab. Zu Beginn der Erhebung wurde außerdem gefragt, wie oft die Lernenden Lernhilfen im Unterricht nutzen konnten. Daraus leiten sich die nachfolgenden Items für das Erhebungsinstrument ab. Die Fragen sollten lediglich beantwortet werden, wenn die Schüler:innen das Differenzierungsangebot auch in den Stunden genutzt hatten.

Verwendung in POST\_1\_1, POST\_1\_1V, POST\_1\_2V, POST\_1\_2, POST\_1\_3V, POST\_1\_3, POST\_1\_F, POST\_1\_4:

*Gründe Hilfenutzung:*

- Ich habe die Hilfen aus Neugier aufgerufen.
- Ich habe die Hilfen aufgerufen, weil ich Hilfe zum Bearbeiten der Aufgaben brauchte.
- Ich habe die Hilfen aus Langeweile aufgerufen.
- Ich habe die Hilfen aus Interesse aufgerufen.

*Gründe Vertiefungsnutzung:*

- Ich habe die Vertiefungsaufgabe aus Neugier aufgerufen.
- Ich habe die Vertiefungsaufgabe aus Interesse aufgerufen.
- Ich habe die Vertiefungsaufgabe aus Langeweile aufgerufen.
- Ich habe die Vertiefungsaufgabe aufgerufen, um ein besseres Fachwissen zu erlangen.
- Ich habe die Vertiefungsaufgabe aufgerufen, da ich noch Zeit hatte.

Gemischte Verwendung:

*Hilfenutzung*

- Wie oft kannst du Lernhilfen, wie z.B. Hilfekarten oder QR-Codes, im Unterricht nutzen? (Verwendung in PRR und PRE\_V)
- Ich empfand die Hilfen als nützlich. (Verwendung in POST\_1\_1, POST\_1\_1V, POST\_1\_2V, POST\_1\_2, POST\_1\_3V, POST\_1\_3, POST\_1\_F, POST\_1\_4)
- Ich finde es besser Hilfen am Tablet aufzurufen, als Hilfekärtchen ausgeteilt zu bekommen. (Verwendung in POST\_1\_E)
- Gib an, ob du dir bei der entsprechenden Aufgabe eher die visuelle (Bild-) Hilfe oder eher die verbale (Text-) Hilfe betrachtet hast. Gib zusätzlich an, wie oft du dir ungefähr die Hilfen der jeweiligen Aufgabe betrachtet hast. (Verwendung in POST\_1\_V, POST\_2\_V, POST\_3\_V und POST\_1\_F)

**Sonstige**

Nachfolgend werden Fragen aufgeführt, die nicht genau einer bestimmten Skala zugeordnet werden können. Im Pre-Test wurden unter anderem durch eine Selbsteinschätzung die Fähigkeiten in allen Schulfächern, insbesondere diejenigen im Fach Chemie und den Naturwissenschaften, erhoben. Auf eine genaue Nennung von Noten wurde aus Datenschutzgründen verzichtet. Zusätzlich könnte die Erhebung von konkreten Noten zu einer Ablehnung gegenüber dem Erhebungsinstrument und damit zu fehlenden Werten oder falschen Angaben führen, wie es zum Beispiel bei Fragen zum Einkommen oder zur Berufstätigkeit aus Panelbefragungen bekannt ist (Möhring & Schlütz, 2019, S. 60). Stattdessen wurde eine 7-stufige Skala mit den Polen „sehr schlecht“ und „sehr gut“ verwendet. Die Zwischenschritte wurden nicht benannt. Weiterhin wurden Daten zur Smartphone-, Tablet- und Computernutzung im schulischen und privaten Umfeld erhoben, um Zusammenhänge zur Usability und zur tabletbezogenen Selbstwirksamkeit herstellen zu können. Hierbei wurde ebenfalls eine 7-stufige Skala verwendet, bei der lediglich die

Pole mit „nie“ und „immer“ beschriftet wurden. Zuvor sollten die Lernenden angeben, ob sie das Gerät prinzipiell nutzen dürfen. Eine Frage diente außerdem zur Erhebung der Nutzungsfrequenz von Lernhilfen im Unterricht.

Es wurden außerdem Fragen zur zeitlichen Gestaltung der Stunde erhoben, die erfassen sollten, ob die Schüler:innen genügend Zeit hatten, alle Aufgaben zu bearbeiten. Weitere Fragen zielten auf eine erhöhte physische Belastung durch den Tabletgebrauch ab. Dadurch könnten womöglich auch Zusammenhänge zum Cognitive Load hergestellt werden. Zudem war es wichtig zu erheben, ob durch den mehrmaligen Tabletgebrauch diese physischen Belastungen zunehmen und ob andere Apps die Lernenden ablenkten. Bei allen Fragebogen wurden die soziodemografischen Daten „Alter“ und „Geschlecht“ abgefragt.

Verwendung in PRE und PRE\_V:

*Fähigkeiten im Fach Chemie, Naturwissenschaften und in der Schule:*

- Wie schätzt du deine Fähigkeiten im Fach Chemie ein?
- Wie schätzt du deine Fähigkeiten in den anderen Naturwissenschaften ein?
- Wie schätzt du deine Fähigkeiten in der Schule generell ein?

Verwendung in PRE:

*Nutzungshäufigkeiten Digitale Medien zuhause:*

- Welches der folgenden Geräte kannst du zuhause zum Lernen nutzen? (Smartphone, Tablet, Computer (PC) oder Laptop)
- Wie oft nutzt du zuhause ein Smartphone zum Lernen?
- Wie oft nutzt du zuhause ein Tablet zum Lernen?
- Wie oft nutzt du zuhause einen PC oder Laptop zum Lernen?

*Nutzungshäufigkeiten Digitale Medien Schule:*

- Welches der folgenden Geräte kannst du in der Schule zum Lernen nutzen? (Smartphone, Tablet, Computer (PC) oder Laptop)
- Wie oft nutzt du in der Schule ein Smartphone zum Lernen?
- Wie oft nutzt du in der Schule ein Tablet zum Lernen?
- Wie oft nutzt du in der Schule einen PC oder Laptop zum Lernen?

*Physiologische und psychische Einschränkungen durch Tabletnutzung:*

- Ich finde es anstrengend für meine Augen, die ganze Zeit auf das Tablet zu schauen.

- Ich finde die Nutzung des Tablets ermüdend.
- Ich wurde durch andere Apps auf dem Tablet vom Lernen abgelenkt.

Verwendung in POST\_1\_E, POST\_1\_V, POST\_1\_2, POST\_1\_2V, POST\_1\_3, POST\_1\_3V, POST\_1\_4 und POST\_1\_F:

*Zeitlicher Umfang der Stunden:*

- Ich hatte genügend Zeit, um das digitale Arbeitsblatt zu bearbeiten.
- Ich hätte mir noch mehr Zeit zum Bearbeiten des digitalen Arbeitsblattes gewünscht. (I)

### 4.3.3 HyperDocs und Differenzierungsangebot

Durch die Beschaffenheit der HyperDocs können diese an dieser Stelle nicht vollumfänglich im schriftlichen Teil dieser Arbeit gezeigt werden. Sie sind digital verfügbar. Im Anhang sind Screenshots der HyperDocs gezeigt (vgl. Anhang 9.1.3). Die Lernhilfen in Form von Texten und Bildern entsprechen denen der Vergleichsgruppe und können dem Anhang 9.1.6 entnommen werden. Die gesamte Unterrichtsreihe und deren Inhalt sind in Kapitel 4.5.3 geschildert.

Mit jedem HyperDoc standen den Lernenden Lernhilfen und pro HyperDoc eine Vertiefungsaufgabe zur Verfügung. Sofern möglich, wurden den Schüler:innen der Interventionsgruppe Hilfen in Form eines kurzen Textes, eines Bildes mit gegebenenfalls einem kurzen Text, einer gesprochenen Audioaufnahme oder eines Videos angeboten. Es handelt sich dabei um ein grundlegendes Prinzip von HyperDocs (Fitting, 2018). Für einen Überblick der Verteilung der Lernhilfen nach Darstellungsform vergleiche Anhang 9.1.1. Im Falle der Text- und Audiohilfen waren die genauen Inhalte und Formulierungen identisch. Bei Bild- und Videohilfen wurde versucht, den Inhalt über alle Darstellungsformen äquivalent zu halten. Aus diesem Grund konnten nicht zu allen Aufgaben Hilfen in allen Darstellungsformen angeboten werden, da in einigen Fällen zum Beispiel eine audiovisuelle Hilfe nicht didaktisch sinnvoll eingebunden werden konnte. Die Videohilfen wurden mithilfe von *Microsoft PowerPoint* erstellt.

Die Vertiefungsaufgaben wurden ebenfalls in Form eines Informationskärtchen zur Verfügung gestellt, um deren Öffnen durch Lernende in HyperDocSystems sichern zu können. Unterhalb des Symbols, das eine solche Aufgabe kennzeichnet, wurde ein freies Textfeld zum Bearbeiten der Aufgabe bereitgestellt. Die neuen HyperDocs wurden den Schüler:innen zu Beginn der Stunde, in der Regel nach dem Einstieg, freigeschaltet. Auf

die HyperDocs der vorherigen Stunde konnten die Lernenden zurückgreifen, um gegebenenfalls erarbeitete Inhalte erneut betrachten zu können. An dieser Stelle sei an die automatische Speicherfunktion der HyperDocs verwiesen, sodass bei versehentlichem Schließen oder bei einem Programmfehler der Bearbeitungsstand der letzten Minute wiederhergestellt werden konnte. Der Vergleichsgruppe standen die gleichen Hilfen in Form von Text und Bild sowie die Vertiefungsaufgaben in Briefumschlägen zur Verfügung (vgl. Kapitel 4.5.4).

## 4.4 Pilotierung

In diesem Unterkapitel werden die Anlage und Durchführung der drei Pilotierungen beschrieben und die daraus gewonnenen Erkenntnisse für die Hauptstudie diskutiert. Dabei ist anzumerken, dass die Pilotierung lediglich für die Interventionsgruppe durchgeführt wurde. Aufgrund der gleichen Unterrichtsreihen, Arbeitsmaterialien und Testinstrumente wurde auf eine gesonderte Pilotierung für die Vergleichsgruppe verzichtet. Insgesamt wurden drei Pilotierungen durchgeführt, in denen insbesondere die Passung der Unterrichtsinhalte, der Materialien und die Zuverlässigkeit sowie der Funktionsumfang von HyperDocSystems getestet wurden. Auf eine ausführliche Testung der Erhebungsinstrumente wurde zunächst verzichtet, da die zugrundeliegenden Stichproben klein waren und es sich um bereits erprobte Testinstrumente handelte. Insgesamt wurden drei Pilotierungen durchgeführt, die in einem iterativen Design schrittweise optimiert wurden. Die erste Pilotierung fand in einer achten Klasse an einer Gesamtschule Ende 2019 statt und stand nicht in Zusammenhang mit dem Thema der endgültigen Unterrichtsreihe. Die zweite Erhebung fand an einem Gymnasium in zwei achten Klassen im Februar 2020 statt. Dabei handelt es sich bereits um das endgültige Unterrichtsthema, das jedoch weiterhin optimiert wurde. Die letzte Pilotierung fand im Juni und Juli 2020 an einer Gesamtschule mit zwei Kursen der 10. Klasse und einem Leistungskurs der 12. Jahrgangsstufe eines Gymnasiums statt. Bei dieser Pilotierung wurden unter anderem das neue Design von HyperDocSystems und die endgültige Unterrichtsreihe getestet.

### 4.4.1 Erste Pilotierung: Erprobung von HyperDocSystems

Die erste Pilotierung wurde mit einer achten Klasse einer Gesamtschule im November / Dezember 2019 im Fach Chemie durchgeführt. Im März 2020 folgte ein Follow-Up. Im Vordergrund der Erprobung stand die Systemstabilität von HyperDocSystems, da die

Anwendung damit zum ersten Mal seit der Neuentwicklung ab 2018 mit einer Lerngruppe im „offline-Modus“ getestet wurde. Bei diesem Modus wird HyperDocSystems über ein lokales WLAN-Netzwerk zur Verfügung gestellt (vgl. Kapitel 4.3.1). Zuvor erfolgte die Entwicklung und Testung auf lokalen Testgeräten über XAMPP oder über den Universitätsserver der ehemaligen TU Kaiserslautern, jetzt Rheinland-Pfälzische Technische Universität. Zwar wurde die erste Version der HyperDocs bereits in einer Masterarbeit (Fitting, 2018) erprobt, allerdings besitzt die neue Anwendung ein anderes Backend sowie eine andere Benutzeroberfläche, die mitunter längere Ladezeiten verursacht. Zum damaligen Zeitpunkt bestand außerdem die Überlegung, neben den bereits genannten Forschungsfeldern, eine weitere Forschungsfrage zu untersuchen. Während die ursprünglichen HyperDocs ein hybrides Format aus einem digitalen Arbeitsblatt mit Hilfen und einer analogen Sicherung auf dem Papier bildeten, bietet die neue Version auch die voll-digitale Sicherung über die Tastatur oder den Apple Pencil. Die wenigsten Schulen besaßen jedoch aufgrund der fehlenden Apple Pencils die technische Ausstattung für die voll-digitale Variante. Daher sollte zunächst ein Vergleich zwischen der ursprünglichen halbdigitalen und der volldigitalen Variante unter Betrachtung der Motivation, Usability und des Cognitive Load erfolgen. Dieser Ansatz wurde jedoch für die Hauptstudie verworfen (vgl. Kapitel 4.4.2) Daher wird in den folgenden Teilkapiteln kein Vergleich zwischen den beiden Nutzungsvarianten vollzogen. Insgesamt erstreckte sich die erste Pilotierung über drei Messzeitpunkte im November und Dezember 2019. Ein Follow-Up mit zwei weiteren Unterrichtsstunden folgte im März 2020. Die Stundeninhalte gliederten sich wie in Tabelle 15 gezeigt. Das verwendete Material wird nicht aufgeführt, da es im weiteren Verlauf nicht weiterverwendet wurde.

*Tabelle 15: Stundeninhalte der ersten Pilotierung in einer Gesamtschule.*

<i>Stunde</i>	<i>Zeitpunkt</i>	<i>Thema</i>
1	November 2019	Einführung: Safety First!
2	Dezember 2019	Fortführung Safety First!
3		Verbrennung – Die wohl bekannteste chemische Reaktion
4	März 2020	Follow-Up: Wasser – Vom alltäglichen Löschmittel zum Brennstoff der Zukunft.
5		Follow-Up: Chemische Reaktion – Auf Teilchenebene erklärbar

Aufgrund eines technischen Problems wurde der erste Unterrichtsinhalt auf zwei Stunden aufgeteilt (s. nächster Absatz). Die Klasse wurde zur Hälfte mit iPads und Apple Pencils ausgestattet sowie mit Samsung-Tablets und der analogen Arbeitsblattvariante. Der Unterricht erfolgte als selbstständige Erarbeitungsphase mit einem kurzen Einstieg und einer kurzen Sicherung.

### **HyperDocSystems und Technik**

Über die fünf Unterrichtsstunden erwies sich HyperDocSystems als stabil und schnell. Eine Ausnahme bildet die erste Stunde. Hierbei führten ein fehlendes JavaScript und eine fehlende CSS-Datei bei den Samsung-Tablets zu verzögerten Ladezeiten von 20 Sekunden. Die iPads öffneten die Seite nicht. Als Ursache des Problems konnte die fehlende Internetverbindung identifiziert werden. Da HyperDocSystems über einen lokalen Hotspot ohne Internetverbindung zur Verfügung gestellt wurde, konnten die Skripte aus dem Internet nicht geladen werden. Der Fehler wurde behoben, indem die Skripte ebenfalls auf dem lokalen Server gespeichert wurden. In der ersten Stunde erfolgten daher lediglich die Einführung in die Studie und die Verbindung mit dem WLAN. In der zweiten Stunde wurde dann der eigentliche Inhalt bearbeitet. Weitere Probleme in Bezug auf die Nutzung von HyperDocs und der Tablets bezogen sich auf die Bedienung durch die Nutzer:innen selbst. Den Schüler:innen war beispielsweise nicht verständlich, dass sie die HyperDocs nutzen können, obwohl ihnen kein Internet zur Verfügung stand. Ein weiteres Problem betraf die falsche Eingabe des WLAN-Passworts oder der Nutzungsdaten von HyperDocSystems. Andere Schüler:innen wiederum wussten nicht, wie sie das Tablet entsperren oder wie das HyperDoc über den QR-Code zu öffnen ist, wobei sich die letzteren Probleme für den Zeitpunkt des Follow-Up zeigten. Weiterhin war zu beobachten, dass manche Schüler:innen das Tablet hochkant nutzten und damit weniger Platz für die Nutzung der Hilfen hatten, da diese am rechten Bildschirmrand angezeigt wurden und die Gesamtfläche in der Breite kleiner wurde.

### **Testinstrumente**

An der ersten Pilotierung nahmen insgesamt 17 Lernende teil (weiblich: 11, männlich: 6; Durchschnittsalter: 13.47). Fehlende Werte wurden mittels des EM-Algorithmus imputiert. Dazu zählen auch fehlende Werte aus der Antwort „Weiß nicht“, die bei der Pilotierung als zusätzliche Antwortmöglichkeit angeboten wurde. Zusätzlich wurden die Skalen „Furcht vor Misserfolg“ und „Hoffnung auf Erfolg“ erhoben, um mögliche Zusammenhänge zur Hilfenutzung aufzudecken. Dieser Forschungsansatz wurde allerdings für die

Hauptstudie aufgrund des umfassenden Studiendesigns verworfen, sodass die beiden Skalen nicht weiter untersucht wurden. Nach der ersten Pilotierung zeigten einige der Testinstrumente zu geringe Reliabilitäten (vgl. Tabelle 16). Während manche Subskalen, insbesondere die des Cognitive Load, niedrige Reliabilitäten zeigten, wiesen andere ausreichende bis gute Cronbachs  $\alpha$  Werte auf. Dazu zählen beispielsweise drei Subskalen der Usability, das Interesse und die Selbstwirksamkeit bei der Tabletnutzung.

*Tabelle 16: Reliabilitäten der einzelnen Skalen in der ersten Pilotierung.*

<i>Subskala</i>	<i>Cronbachs <math>\alpha</math> (Mittelstufe, n = 17)</i>
<i>Intrinsic Motivation Inventory</i>	
Interesse	0.75
Kompetenzerleben	0.63
Anstrengung	0.54
Druck	0.47
<i>Cognitive Load</i>	
Intrinsic Load	0.43
Germane Load	0.22
Extraneous Load	0.34
<i>Tabletbezogenen Kompetenzen</i>	
Selbstwirksamkeit	0.89
Tablets beim Lernen	< 0.1
Angst vor Tablets	<i>Nicht feststellbar aufgrund fehlender Varianz eines Items</i>
<i>Usability</i>	
Nützlichkeit	0.84
Zufriedenheit	0.77
Einfachheit des Lernens	0.89
Einfachheit der Nutzung	< 0.1

Die Gründe für die geringen Werte konnten nicht genau benannt werden. Beispielsweise war die Reliabilität der Subskala „Einfachheit der Nutzung“ im Vergleich zu den anderen drei Faktoren der Usability ungewöhnlich niedrig. Gleiches gilt bei der Subskala „Tablets beim Lernen“. Ein möglicher Grund für die niedrigen Reliabilitäten könnte im Alter der Lernenden liegen, sodass diese die Fragen nicht vollständig verstanden oder die



Aufmerksamkeitsspanne zur Beantwortung des Fragebogens am Ende der Stunde bereits erschöpft war. In diesem Zusammenhang fiel auf, dass oftmals auch die invertierten Items zu geringeren Cronbachs  $\alpha$  Werten führten. Auf der anderen Seite war die Stichprobe mit 17 Schüler:innen auch gering, zumal fehlende Werte imputiert wurden, was zu einer zusätzlichen Verzerrung bei dieser niedrigen Anzahl an Teilnehmenden führen könnte.

### **Fazit und Anpassungen**

Die erste Pilotierung war ein wichtiger Schritt in Richtung der Hauptstudie. So konnte ein großes technisches Problem früh entdeckt und behoben werden. Außerdem zeigte sich, dass die Schüler:innen zum Teil geringere technische Kompetenzen, als erwartet, aufwiesen. Aus diesem Grund wurde die Anleitung zur Inbetriebnahme ausführlicher und mit anschaulichen Symbolen dargestellt. Die Schüler:innen wurden von nun an außerdem darauf hingewiesen, das Tablet im Querformat zu nutzen, um eine größere Bedienungsfläche zu sehen. Die hier dargestellte Unterrichtsreihe eignete sich aufgrund ihrer umfangreichen selbstständigen Bearbeitungsphase nur bedingt für eine repräsentative Untersuchung von HyperDocSystems im Regelunterricht. Im Schulalltag sind demnach überwiegend Stunden anzutreffen, die auch lehrerzentrierte beziehungsweise plenumszentrierte Unterrichtsphasen aufweisen. Außerdem handelte es sich bei den Inhalten der Stunden um zentrale Themen des neuen Lehrplans, die in vielen Klassen behandelt werden. Da in der Hauptstudie allerdings von einem gleichen Vorwissensstand ausgegangen werden sollte, schränkte das die spätere Akquirierung weiterer Schulgruppen stark ein. Aus diesem Grund wurde auf eine andere Unterrichtsreihe zurückgegriffen, die im Lehrplan verankert ist, jedoch eher ein Randthema darstellt und damit in der Form bisher von wenigen Klassen durchgeführt wurde. Die Unterrichtsreihe wird in Kapitel 4.5.3 dargestellt und ab der zweiten Pilotierung näher untersucht.

Nach der ersten Pilotierung zeigten die Testinstrumente zum Teil unzureichende Reliabilitäten. Die Gründe hierfür können vielfältig sein. Die Lernenden waren womöglich nach der Unterrichtsstunde unkonzentriert und die Stichprobe ist insgesamt sehr klein. Nichtsdestotrotz wurden die Instrumente zur zweiten Pilotierung beibehalten und mit einer größeren Stichprobe und einer anderen Schulform erneut getestet.

#### **4.4.2 Zweite Pilotierung: Erprobung der Unterrichtsreihe**

Die zweite Pilotierung fand im Januar und März (Mittelstufe) sowie im Februar (Oberstufe) 2020 statt. Die Erprobung in der Mittelstufe erfolgte mit zwei achten Klassen im

Chemieunterricht, die der Oberstufe mit einer Schüler:innengruppe an der Universität. In beide Stichproben erfolgte erstmals der Einsatz der neuen Unterrichtsreihen, die von Czubatinski et al. (2020) und Nieß et al. (2020) ursprünglich entwickelt wurden. Die Reihen waren in klassischen Lehr-/Lernszenarien erprobt und wurden demnach für den Einsatz mit HyperDocs durch beispielsweise die Integration von Lernhilfen erweitert. Wie bereits in der ersten Pilotierung, wurden die Mittelstufenklassen geteilt und zur Hälfte mit iPads und zur anderen Hälfte mit Samsung-Tablets sowie analogen Arbeitsblättern unterrichtet. Die Erprobungszeiträume sind in der Mittelstufe in nachfolgender Tabelle zusammengefasst (vgl. Tabelle 17).

*Tabelle 17: Stundeninhalte der zweiten Pilotierung im Gymnasium (Mittelstufe).*

<i>Stunde</i>	<i>Zeitpunkt</i>	<i>Thema</i>	<i>Material (Anhang)</i>
1		Einführung: Safety First! (Gefahrstoffe u. Gefahrensymbole)	
2	Januar 2020	Farbstoffe in Lebensmittel Farbstoffanalyse mit Schüler:innenexperiment	9.2.1
3		Farbstoffanalyse und Konzentrationsbegriff	
4		Konzentrationsbegriff	
5	März 2020	Follow-Up: Zucker in Lebensmitteln <sup>7</sup>	/

In der Mittelstufe wurde ein Follow-Up durchgeführt. In einer Klasse kam es aufgrund eines Fehlers beim Starten von XAMPP bei der Einführung zu einem erneuten Ausfall der Technik, sodass diese Gruppe keine unmittelbare Einführung in die HyperDocs erhielt. Aufgrund der engen Taktung konnte die Stunde nicht nachgeholt werden. In der Oberstufe erfolgte die Erhebung an einem Nachmittag (vgl. Tabelle 18). Dort stand die inhaltliche Validierung der Unterrichtsreihe im Vordergrund. Dadurch wurde lediglich ein Fragebogen am Ende der Erhebung eingesetzt. Alle Schüler:innen arbeiteten mit iPads und führten die Versuche ebenfalls selbst durch.

<sup>7</sup> Material nicht aufgeführt, da für Hauptstudie nicht verwendet.

Tabelle 18: Stundeninhalte der zweiten Pilotierung in der Oberstufe.

<i>Stunde</i>	<i>Zeitpunkt</i>	<i>Thema</i>	<i>Material (Anhang)</i>
1		Einführung: Safety First! (Gefahrstoffe u. Gefahrsymbole)	
2	Februar 2020	Farbstoffanalyse mit Schüler:innenexperiment	9.2.1
3		Spektralphotometer	
4		Optimierte Farbstoffanalyse	

### **HyperDocSystems und Technik**

Das erwähnte Problem beim Starten des lokalen Servers ließ sich auf eine fehlerhafte Datei im Zusammenhang mit der Datenbank und MariaDB zurückführen. Durch den Ersatz mit einer funktionierenden Datei konnte der Fehler behoben werden. Der Bug trat in diesem Zusammenhang zum ersten Mal auf und blieb bei der ersten Pilotierung unentdeckt. Ein weiteres Problem bezog sich auf das Schreibwerkzeug von HyperDocSystems mittels des Apple Pencil. Die Radierfunktion wurde in Form eines Weißstiftes implementiert, sodass bei regulären weißen Textfeldern mit einem weißen Stift über den Text geschrieben wurde. Bei beschreibbaren Abbildungen, wie beispielsweise einem Koordinatensystem, überdeckt die Farbe das Bild und sie ließ sich nicht mehr entfernen. Daher wurde der Radierer überarbeitet, sodass auch auf Bildern geschrieben und radiert werden konnte. Ein weiteres Problem lag in der kurzen Dauer der Session, die in Cookies gespeichert wurde. Dadurch war es manchen Lernenden nicht mehr möglich, ihre Lösungen zu speichern. Dieses Problem trat sowohl in der Mittel- als auch in der Oberstufe auf. Ein weiteres Problem trat auf, wenn sehr lange Texte gespeichert wurden. Hierbei war in der Datenbank der Dateityp für die Lösungen auf „text“ eingestellt. Der Typ wurde auf „longtext“ geändert, womit das Problem gelöst werden konnte.

### **Testinstrumente**

Die Stichprobe der Mittelstufe bestand aus 36 Schüler:innen (weiblich: 16, männlich:19, keine Angabe: 1; Durchschnittsalter: 13.60) aus zwei Klassen. Die der Oberstufe aus 14 Lernenden (weiblich: 10, männlich: 4; Durchschnittsalter: 16.21). Fehlende Werte wurden, falls nicht anders angegeben, vollständig mittels des EM-Algorithmus imputiert. Die nachfolgende Tabelle (vgl. Tabelle 19) zeigt die Reliabilitäten der Subskalen der verschiedenen Testinstrumente (vgl. auch Kapitel 4.3.2.). Dabei wurden die Mittel- und

Oberstufe getrennt analysiert, da bei der Oberstufe bereits angepasste Testinstrumente eingesetzt wurden. Bei den Aussagen zur Motivation wurde die Verneinung mancher Items aufgehoben, die durch ein „nicht“ invertiert wurden, da der Verdacht bestand, die Schüler:innen würden die Fragen nicht genau lesen. Auch bei der Usability wurde bei einer Frage die Invertierung aufgehoben und ein Item ergänzt. Beim Cognitive Load wurden Fragen aus anderen Instrumenten, angelehnt an Eysink et al. (2009) und übersetzt nach Leppink et al. (2013) ergänzt. Das neue Instrument bestand aus elf Fragen. Außerdem wurde in der Oberstufe die Option „weiß nicht“ entfernt.

*Tabelle 19: Reliabilitäten der einzelnen Skalen in der zweiten Pilotierung.*

<i>Subskala</i>	<i>Cronbachs <math>\alpha</math> (Mittelstufe, <math>n = 36</math>)</i>	<i>Cronbachs <math>\alpha</math> (Oberstufe, <math>n = 14</math>)</i>
<i>Intrinsic Motivation Inventory</i>		
Interesse	0.88	0.92
Kompetenzerleben	0.68	0.94
Anstrengung	0.24	0.89
Druck	0.17	0.89
<i>Cognitive Load</i>		
Intrinsic Load	0.42	0.90
Germane Load	0.61	0.80
Extraneous Load	0.73	0.85
<i>Tabletbezogene Kompetenzen</i>		
Selbstwirksamkeit	0.85	0.90
Tablets beim Lernen	0.85	0.92
Angst vor Tablets	0.39	<i>Nicht feststellbar aufgrund fehlender Varianz eines Items</i>
<i>Usability</i>		
Nützlichkeit	0.90	0.89
Zufriedenheit	0.93	0.92
Einfachheit des Lernens	0.73	0.97
Einfachheit der Nutzung	0.56	0.96

Bei beiden Gruppen wurde jeweils der Datensatz für die Reliabilitätsanalyse herangezogen, der den ersten Kontakt mit dem Testinstrument widerspiegelte. Bei einer Klasse wurde daher aufgrund des Ausfalls in der ersten Stunde der Datensatz der zweiten Stunde (Motivation und Cognitive Load) und der vierten Stunde (Usability) herangezogen. In der Mittelstufe fielen die Ergebnisse der Reliabilitätsüberprüfung erneut gemischt aus. Während bei den Subskalen der intrinsischen Motivation zwei Skalen eine bessere Reliabilität zeigten, wiesen die Skalen „Anstrengung“ und „Druck“ trotz größerer Stichprobe eine geringere Reliabilität auf. Auf der anderen Seite konnten die Cronbachs  $\alpha$  Werte der Subskala „Germane Load“ und „Extraneous Load“ merklich verbessert werden. Das Gleiche gilt für die Skalen „Tabletbezogene Selbstwirksamkeit“ und „Tablets beim Lernen“ bei den tabletbezogenen Kompetenzen. Eine Faktorenanalyse des neuen Instruments zur Erfassung des Cognitive Load wurde jedoch aufgrund der kleinen Stichprobe zu diesem Zeitpunkt noch nicht durchgeführt. Auch die Skalen der Usability konnten, bis auf „Einfachheit des Lernens“ verbessert werden. In der Oberstufe hingegen zeigen alle Skalen eine hervorragende Reliabilität, wobei hier die Reliabilität der „Tablet Angst“ nicht feststellbar war. Die Unterschiede zwischen Mittel- und Oberstufe sind womöglich auf die Änderungen an den Instrumenten zurückzuführen. Insbesondere beim Cognitive Load zeigten sich deutlich bessere Reliabilitäten. Ein weiterer Grund für die besseren Werte von Cronbachs  $\alpha$  lag gegebenenfalls darin, dass die Lernenden der Oberstufe in Ruhe die Fragen durchlesen konnten und wahrscheinlich auch eine höhere Auffassungsgabe als die Mittelstufenschüler:innen aufwiesen. Dadurch wurden die Fragen sorgfältiger gelesen und beantwortet.

### **Fazit und Anpassungen**

Auch die zweite Pilotierung erwies sich als gewinnbringend in Bezug auf die Behebung technischer Probleme und die Planung der Unterrichtsreihe. Bei der technischen Komponente konnten gleich zwei kritische Fehler beseitigt werden, die ansonsten zu einem Abbruch der Unterrichtsstunden führen würden. Auf der Seite der didaktischen und methodischen Planung erwies sich die Unterrichtsreihe für eine achte Klasse als zu komplex. Daher wurde für die folgenden Studien die Stufe bei der Mittelstufe auf die 9. Klasse angehoben. Zusätzlich wurden weitere Hilfen bei den Aufgaben ergänzt und die eigentliche Unterrichtsreihe, abzüglich der Einführung, auf volle drei Stunden ausgedehnt. Wie in Tabelle 17 ersichtlich, gab es teilweise Überschneidungen des Stundeninhalts zwischen einzelnen Stunden, die darauf zurückzuführen sind, dass die Unterrichtsreihe

ursprünglich zwei Stunden umfassen sollte. Zuletzt wurde aus zwei Gründen das Schüler:innenexperiment als Lehrer:innenexperiment durchgeführt. Zum einen erfordert die Durchführung durch die Lernenden mehr Zeit und zum anderen interagiert womöglich das Schüler:innenexperiment mit der Motivation und der Nutzung der HyperDocs, weshalb diese Störvariable aus dem Untersuchungsdesign entfernt wurde. Da Schüler:innen zudem äußerten, lieber mit iPads und Apple Pencils zu arbeiten und die Kombination aus digitalem und analogem Arbeitsblatt teilweise als redundant ansahen, insbesondere, wenn die Hilfen nicht genutzt wurden, wurden ab diesem Zeitpunkt die folgenden Erhebungen lediglich mit der volldigitalen Variante durchgeführt. In diesem Sinne erfolgte eine abschließende dritte Pilotierung unter einer angepassten Unterrichtsreihe. Zum Follow-Up der letzten Pilotierung wurde außerdem eine überarbeitete Version von HyperDocSystems, die unter anderem auch ein neues Layout und eine längere Laufzeit der Cookies enthält, eingesetzt. Weiterhin wurde nach der zweiten Pilotierung die Anleitung zum Einloggen und zum Erstkontakt mit den HyperDocs weiter überarbeitet. Zusätzlich wurden alle iPads in das WLAN eingewählt, sodass die Schüler:innen dies nicht mehr selbst durchführen mussten. Der Link zu HyperDocSystems wurde direkt in Safari als Favorit hinterlegt, womit die Einwahl mittels QR-Codes entfiel. Zudem wurde ab diesem Zeitpunkt bei jedem Tablet der Browsercache nach jeder Stunde manuell geleert, damit sich die Schüler:innen jedes Mal erneut mit ihrem eigenen Account einwählen mussten und nicht aus Versehen einen anderen Account nutzten.

Aufgrund der teilweise erneut schlechten Reliabilitäten wurde für die Oberstufe der Fragebogen angepasst. Die Anpassungen zeigten Wirkung, wobei die besseren Werte auch auf die generellen Umstände zurückzuführen waren. So erfolgte die Befragung innerhalb der Universität unter geringerem Zeitdruck und die Lernenden hatten eine bessere Auffassungsgabe. Aus diesem Grund wurde für die dritte und letzte Pilotierung das neue Instrument erneut eingesetzt. Dazu wurde noch ein Item der Subskala „Zufriedenheit“ der Usability „Ich würde an den digitalen Arbeitsblättern nichts ändern“ durch „Ich finde die digitalen Arbeitsblätter großartig“ getauscht, da das erste Item nach erneuter Prüfung als sehr allgemein wahrgenommen wurde. Außerdem wurde die Skala „Wahrgenommene Wahlfreiheit“ des IMI ergänzt, da ein Einfluss der HyperDocs auf das Autonomieempfinden, nach erneuter Prüfung der Literatur, vermutet wurde.

#### 4.4.3 Dritte Pilotierung: Erprobung der finalen Version der Unterrichtsreihe und von HyperDocSystems

Die dritte Pilotierung fand im Mai und Juni 2020 mit zwei Mittelstufenkursen (10. Klasse) der Gesamtschule und einem Chemieleistungskurs 12 des Gymnasiums statt. Aufgrund der Beschränkungen durch Corona fand der Unterricht in der Mittelstufe mit Teilgruppen im wöchentlichen Wechsel statt. Eine Erhebung in der Mittelstufe fand mit einem Kurs des Wahlpflichtfaches Ökologie statt. Er war daher entsprechend kleiner als eine reguläre Klasse. Die Durchführung erfolgte mit der angepassten Unterrichtsreihe (vgl. Tabelle 20).

*Tabelle 20: Stundeninhalte der dritten Pilotierung an der Gesamtschule (Mittelstufe).*

<i>Stunde</i>	<i>Zeitpunkt</i>	<i>Thema</i>	<i>Material (Anhang)</i>
1	Mai / Juni 2020	Einführung: Safety First! (Gefahrstoffe u. Gefahrsymbole)	
2		Farbstoffe in Lebensmittel	9.2.2
3	Juni 2020	Farbstoffanalyse mit Lehrer:innenexperiment	
4		Konzentrationsbegriff	
5		Follow-Up: Zucker – das weiße Gold? <sup>*8</sup>	/

*\*Am Follow-Up hat lediglich eine Lerngruppe teilgenommen.*

Der Chemieleistungskurs wies keine Besonderheiten auf. Aufgrund der kleinen Gruppengröße fand der Unterricht mit der gesamten Lerngruppe statt. Die Unterrichtsreihe wurde in einer Woche durchgeführt, das Follow-Up circa vier Wochen später (vgl. Tabelle 21).

*Tabelle 21: Stundeninhalte der dritten Pilotierung an dem Gymnasium (Oberstufe).*

<i>Stunde</i>	<i>Zeitpunkt</i>	<i>Thema</i>	<i>Material (Anhang)</i>
1	Mai 2020	Einführung: Safety First! (Gefahrstoffe u. Gefahrsymbole)	
2		Farbstoffanalyse mit Lehrer:innenexperiment	9.2.2
3		Spektralphotometer	
4	Juni 2020	Optimierte Farbstoffanalyse	
5		Follow-Up: Zucker – das weiße Gold? (Oberstufe) <sup>8</sup>	/

<sup>8</sup> Material nicht enthalten, da für Hauptstudie nicht verwendet.

## **HyperDocSystems und Technik**

Im Rahmen der Erprobung traten keine technischen Schwierigkeiten auf. Die aktualisierte Version von HyperDocs konnte erfolgreich im Follow-Up eingesetzt werden. Die Schüler:innen bewerteten das neue Layout besser als das alte (vgl. Abbildung 16). Dabei wurden folgende drei Fragen gestellt:

- NEU1: „Ich finde die neue Version der digitalen Arbeitsblätter nützlicher als die vorherige.“
- NEU2: „Ich finde die neue Version der digitalen Arbeitsblätter moderner als die vorherige.“
- NEU3: „Ich finde die neue Version der digitalen Arbeitsblätter übersichtlicher als die vorherige.“

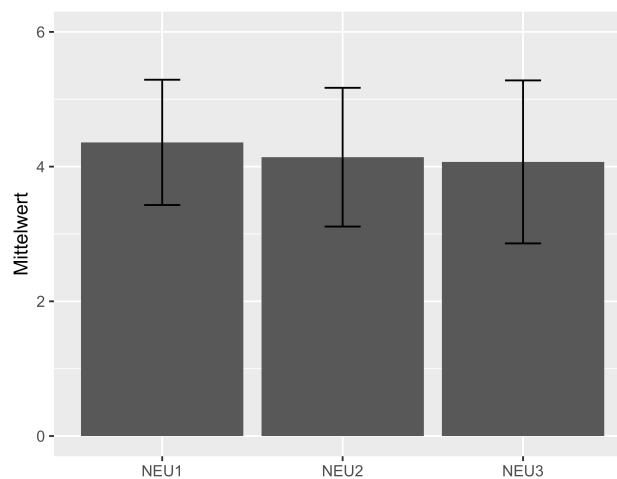


Abbildung 16: Mittelwerte mit Standardabweichung der Befragung zur aktualisierten Version von HyperDocSystems. 5-stufige Skala vom Typ Likert mit den Polen „stimme nicht zu“ (1) bis „stimme zu“ (5).

## **Testinstrumente**

Die Stichproben teilten sich in eine Gruppe für die Mittelstufe (weiblich: 8, männlich: 16; Durchschnittsalter: 15.91) und eine für die Oberstufe auf (weiblich: 3, männlich: 3; Durchschnittsalter: 18.33). Manche Lernende hatten zu der allerersten Version von HyperDocs Kontakt, allerdings lag dieser zu diesem Zeitpunkt zwei Jahre zurück. Fehlende Werte wurden durch den EM-Algorithmus ersetzt. Die Reliabilitäten der Mittel- und Oberstufe wiesen in Anbetracht der Stichprobengrößen zufriedenstellende Werte auf (vgl. Tabelle 22). Eine Ausnahme bildeten hierbei die Subskalen „Interesse“, „Kompetenzerleben“, „Zufriedenheit“ und „Einfachheit der Nutzung“ in der Oberstufe. Die niedrigen Werte sind wahrscheinlich auf die sehr kleine Stichprobe zurückzuführen. In der Seite | 142



Mittelstufe waren die Reliabilitäten ausreichend. Eine Ausnahme bildete die Skala zum Germane Load. Zusätzlich wurde für das Instrument des Cognitive Load eine Hauptkomponentenanalyse durchgeführt, da es sich um ein neu zusammengestelltes Instrument handelte. Dabei ist jedoch anzumerken, dass die Stichprobenzahl gering war, um eine valide Analyse durchführen zu können. Field und Kolleg:innen (2012) empfehlen in diesem Zusammenhang mindestens zehn Proband:innen pro Variable. Daher wurden die Daten der Mittel- und Oberstufe vereint und mittels SPSS die Analyse durchgeführt.

*Tabelle 22: Reliabilitäten der einzelnen Subskalen in der dritten Pilotierung.*

<i>Subskala</i>	<i>Cronbachs <math>\alpha</math> (Mittelstufe, <math>n = 20</math>)</i>	<i>Cronbachs <math>\alpha</math> (Oberstufe, <math>n = 6</math>)</i>
<i>Intrinsic Motivation Inventory</i>		
Interesse	0.85	0.45
Kompetenzerleben	0.88	< 0.1
Anstrengung	0.67	0.76
Druck	0.66	0.95
Wahrgenommene Wahlfreiheit	0.72	0.94
<i>Cognitive Load</i>		
Intrinsic Load	0.70	0.99
Germane Load	< 0.1	0.73
Extraneous Load	0.71	0.72
<i>Tabletbezogenen Kompetenzen</i>		
Selbstwirksamkeit	0.84	0.95
Tablets beim Lernen	0.78	0.94
Angst vor Tablets	0.91	<i>Nicht feststellbar aufgrund fehlender Varianz eines Items</i>
<i>Usability</i>		
Nützlichkeit	0.94	0.87
Zufriedenheit	0.79	0.44
Einfachheit des Lernens	0.96	<i>Nicht feststellbar aufgrund fehlender Varianz zweier Items</i>
Einfachheit der Nutzung	0.83	< 0.1

Die Stichprobe umfasste 26 Schüler:innen. Der Kaiser-Mayer-Olkin-Test ergab einen Wert von 0.567, der Barlett-Test auf Sphärizität erreichte das Signifikanzniveau ( $\chi^2(55) = 172.84, p < 0.001$ ). Damit waren die prinzipiellen Voraussetzungen der weiteren Analyse erfüllt (Field et al., 2012). Anschließend wurde die Korrelationsmatrix analysiert. Hierbei zeigten zwei Variablen keine Korrelation größer 0.3 zu anderen Variablen. Auch die Anti-Image-Korrelationen der beiden Variablen zeigten einen Wert von unter 0.2. Es konnten auf Basis der Eigenwerte vier Komponenten extrahiert werden. Der Screeplot blieb uneindeutig (vgl. Anhang 9.2.3). Hierbei waren eher zwei Komponenten anzunehmen. Die erklärte Varianz der letzten beiden Komponenten lag zwischen neun und dreizehn Prozent, wodurch sich das Ergebnis des Screeplots erklären ließ. Die verschiedenen Ladungen der Items zeigten, dass eine vollständige Mischung der ursprünglichen Faktorzugehörigkeit vorlag (vgl. Abbildung 17). Die Items CL1 bis CL3 waren ursprünglich dem Intrinsic Load zuzuordnen, die Items CL4 bis CL8 dem Extraneous Load und CL9 bis CL11 dem Germane Load. Nach Ausschluss der Items CL9 und CL11, die wie oben beschrieben eine geringe Korrelation und Anti-Image-Korrelation zu anderen Variablen aufwiesen, konnte keine Verbesserung der rotierten Komponentenmatrix festgestellt werden. Das Instrument wurde daher wieder verworfen.

Rotierte Komponentenmatrix<sup>a</sup>

	Komponente			
	1	2	3	4
CL4	,946			
CL7	,897			
CL2	,889			
CL8		,921		
CL6		,877		
CL10INV		-,778		
CL1		,738		
CL5			,807	
CL3	,503		,718	
CL11			-,593	,316
CL9				,957

Extraktionsmethode: Hauptkomponentenanalyse.  
Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 5 Iterationen konvergiert.

*Abbildung 17: Rotierte Komponentenmatrix der Hauptkomponentenmatrix des Instruments für den Cognitive Load.*

### **Fazit und Anpassungen**

Die dritte Pilotierung verlief trotz des Wechselunterrichts, bedingt durch Corona, wie geplant. Es gab keine technischen Probleme. Die neue Version von HyperDocSystems

erwies sich im Follow-Up als gewinnbringend und zuverlässig. Auch die Unterrichtsreihe konnte sowohl in der Mittel- als auch in der Oberstufe innerhalb der vier Unterrichtsstunden durchgeführt werden. Die Follow-Up-Erhebung wurde jedoch für die Haupterhebung gestrichen, da bereits vier Unterrichtsstunden im Regelunterricht Chemie einen großen Stundenumfang bilden und nach Rücksprache mit Lehrkräften nicht mehr Stunden für ein solches Projekt eingeräumt werden konnten. Für die Aufgaben wurde außerdem eine Lösungswahrscheinlichkeit bestimmt (vgl. Anhang 9.2.4). Aufgrund der ermittelten Aufgabenschwierigkeiten wurden Aufgaben daraufhin umformuliert, indem beispielsweise eine Aufgabe in zwei Aufgaben aufgeteilt oder der Operator geändert wurde, und Hilfen angepasst, beispielsweise durch eine stärkere Unterstützung.

Zur Hauptstudie wurde eine Funktion implementiert, die das Wiederherstellen einer Lösung ermöglichte, falls das Arbeitsblatt beispielsweise aus Versehen geschlossen wurde. Die überarbeitete Version von HyperDocSystems, die im Follow-Up der dritten Pilotierung eingesetzt wurde, blieb für die Hauptstudie bestehen. In Bezug auf die Testinstrumente fanden kleine Änderungen statt. So wurden die Items der Fragebatterien randomisiert und die Verneinungen der Items der Motivation wieder eingeführt, da beobachtet wurde, dass Schüler:innen Fragen durchgängig ankreuzen, ohne diese ausreichend zu lesen. Dadurch stieg zwar die Reliabilität des Instruments, allerdings entsprachen vermutlich die beobachteten Werte nicht mehr den realen Werten der zugrundeliegenden Persönlichkeitskonstrukte. Das Instrument zur Erfassung des Cognitive Load erschien zwar reliabel, jedoch nicht valide. Die Zuordnung der Items zu den einzelnen Arten des Cognitive Load konnte in einer Hauptkomponentenanalyse nicht mehr gefunden werden. Aus diesem Grund wurde für die Hauptstudie wieder das Instrument von Klepsch et al. (2017) eingesetzt, da es sich um ein validiertes Instrument handelt.

## 4.5 Durchführung der Hauptstudie

Dieses Unterkapitel beschreibt die Durchführung der Studie. Der generelle Ablauf zwischen Interventions- und Kontrollgruppe war in großen Teilen identisch wie unten beschrieben. Es gab geringe Abweichungen, da zum Beispiel keine Tablets ausgeteilt werden mussten, wobei dieser Vorgang generell nur wenig Zeit beanspruchte. Ein weiterer Unterschied lag bei der inhaltlichen Gestaltung der Fragebögen. Da bei der Kontrollgruppe keine Fragen bezüglich der Nutzung von digitalen Geräten oder der Usability erhoben wurden, waren die Fragebögen zum Teil kürzer, woraus sich eine geänderte

Erhebungsstruktur ergab. Insgesamt entfielen bei der Kontrollgruppe alle Tätigkeiten, die im Zusammenhang mit dem digitalen Medium standen. Nachfolgend wird daher der Ablauf aus Sicht der Interventionsgruppe geschildert, wobei Abweichungen zu der Kontrollgruppe vermerkt sind.

### **4.5.1 Vorbefragung**

Im Vorfeld der Studie wurden die beteiligten Klassen über die Studie und die Erhebung der Daten informiert. Das Einverständnis der Eltern wurde eingeholt. In wenigen Fällen wurde kein Einverständnis gegeben, sodass die Schüler:innen mit regulären analogen Arbeitsblättern am Unterrichtsgeschehen teilnahmen. Es erfolgte anschließend, in den meisten Fällen circa eine Woche vor Beginn der eigentlichen Studie, die Erhebung des Pre-Fragebogens als Paper-Pencil-Test. Die Fragebögen wurden den Lehrkräften mit einer Instruktion in Umschlägen überreicht. In manchen Fällen, bedingt durch die durch Corona erschwerte Organisation in den Schulen, druckten die Lehrkräfte die Fragebögen selbst aus. Über alle Fragebögen hinweg wurde ein anonymisierter Code zur Zuordnung über mehrere Messzeitpunkte eingesetzt. In der Vergleichsgruppe wurde dieser Fragebogen in veränderter Form (vgl. Kapitel 4.3.2) unmittelbar vor der ersten Stunde erhoben. Die dafür benötigte Zeit ergab sich durch das Wegfallen der Einführung in die iPads und durch die kürzere Struktur des Fragebogens, indem dort beispielsweise das Instrument zur Usability entfiel.

### **4.5.2 Ablauf der Hauptstudie**

Die Studie fand im Regelunterricht statt. In der Regel mussten Schüler:innen den Saal wechseln und die Technik musste kurzfristig aufgebaut werden. In manchen Fällen fanden auch Stunden in verschiedenen Klassen direkt hintereinander statt, sodass ein Wechsel des Saales oder ein erneutes Aufbauen des Laptops und Routers oder des Versuchs entfiel. Da jedoch nach jeder Stunde eine fünfminütige oder eine große Pause stattfand, konnte die Technik oder das Experiment ohne größere Verzögerungen aufgebaut werden. Die Schulstunden begannen daher pünktlich. Ein Überblick über die Erhebung bietet Abbildung 18. Die viergliedrigen Unterrichtsstreihen für beide Stufen sind das Resultat der Pilotierungen. Den Beginn und das Ende bildete jeweils ein beziehungsweise zwei Fragebögen. Nach jeder Stunde wurde ebenfalls ein Fragebogen eingesetzt. In der ersten Stunde wurden die Schüler:innen kurz in den Ablauf der kommenden Stunden eingeführt. Danach wurde der Lerngegenstand der Stunde vorgestellt. Anschließend kamen die

Lernenden reihenweise nach vorne und holten sich ein Tablet und die Einföhrungserklarung (vgl. Anhang 9.1.3). Diese erklarte schrittweise die Nutzung der iPads und das Öffnen von HyperDocSystems.

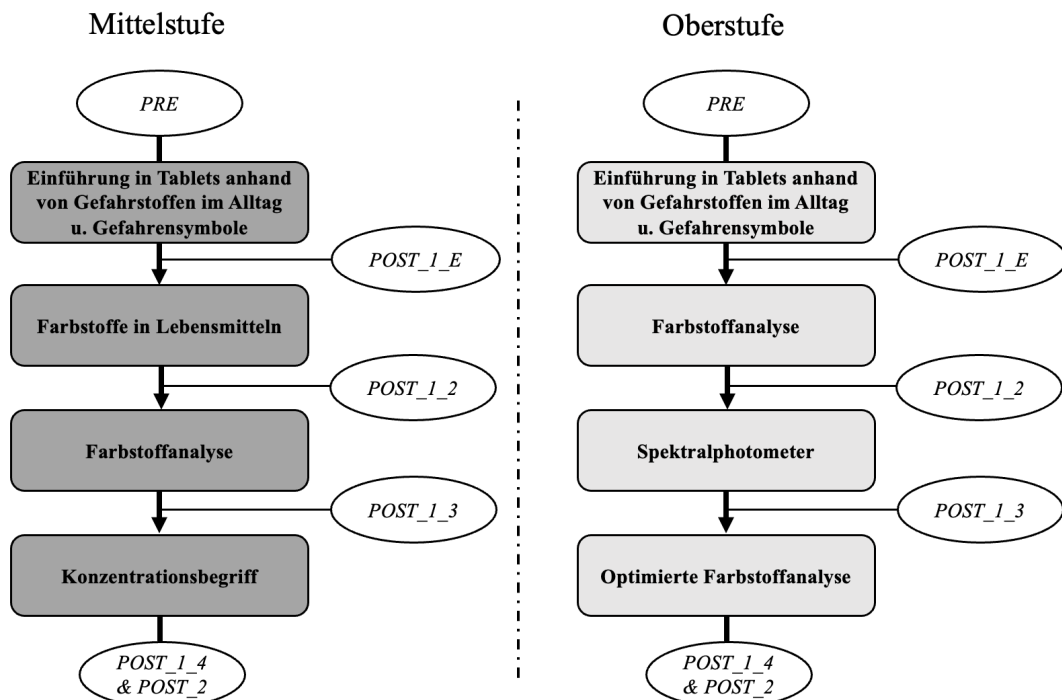


Abbildung 18: Ablauf der Studie mit Erhebungsinstrumenten in der Interventionsgruppe der Mittel- und Oberstufe. Die Intervention in der Kontrollgruppe verlief analog mit anderen Fragebögen.

Zusatzlich wurde den Schöler:innen mit der Erklarung auch der Logincode ausgehandigt. Die Zuteilung der Codes erfolgte randomisiert. Die Codes wurden im Vorfeld der Erhebung durch ein PHP-Skript nach folgendem Muster erstellt: Die Codes begannen mit einem Kurzel, das die beteiligte Schule reprasentiert, wie zum Beispiel „igs“. Der zweite Teil des Codes war eine einzigartige, zufallig generierte ganze Zahl zwischen eins und hundert. Wahrend in den Pilotierungen das Skript fortlaufende Zahlen zwischen eins und dreißig generierte, wurde das PHP-Programm fur die Hauptstudie um eine Randomisierung erganzt, da in der Pilotierung beobachtet wurde, dass Schöler:innen teilweise die gleichen Accounts verwendeten. Das konnte zum Beispiel passieren, indem die Nutzer:innen die Accounts errieten oder sich bei der Eingabe vertippten.

Diese Reihenfolge wurde gewahlt, da die Gefahr bestand, dass Schöler:innen nach Aushandigung der iPads nicht mehr richtig dem Unterrichtsgeschehen folgten und stattdessen die Tablets begutachteten. Zum Schluss wurden die Aufgaben besprochen und es erfolgte die Ausföllung des Fragebogens POST\_1\_E. Dazu hatte die Lerngruppe circa 15 Minuten

Zeit. Die Tablets wurden am Ende der Stunde zusammen mit dem Fragebogen zum Pult zurückgebracht. Die weiteren Stunden verliefen nach dem gleichen Schema. Die genaue Konzeption der einzelnen Stunden kann Kapitel 4.5.3 entnommen werden. Der Fragebogen POST\_2 wurde in der Regel unmittelbar nach der letzten Unterrichtsstunde ausgefüllt.

Der Ablauf in der Kontrollgruppe verlief analog, wobei die technische Einführung in die Unterrichtsmethodik entfiel und durch eine Erklärung zur Nutzung der Lernhilfen ersetzt wurde. Zudem fanden, aufgrund der reduzierten Forschungsfragen in der Vergleichsgruppe, kürzere Fragebögen Anwendung. Die genaue Verteilung wurde bereits in Kapitel 4.3.2 dargelegt.

### **4.5.3 Fachdidaktische Erläuterungen zur Unterrichtsreihe**

In diesem Kapitel erfolgt eine fachdidaktische Einordnung der Unterrichtsreihen. Diese umfasst neben der Relevanz des Lerngegenstandes für den Chemieunterricht auch eine didaktische und methodische Analyse der Stunden, um eine bessere Diskussionsgrundlage für die Auswertung der Ergebnisse zu schaffen. Für die Mittel- und Oberstufe fanden zwei verschiedene Unterrichtsreihen Anwendung, die jedoch thematisch miteinander verwandt sind und teilweise die gleichen Inhalte auf einem unterschiedlichen Niveau behandeln. Die nachfolgenden Ausführungen gelten sowohl für die Interventions- als auch für die Vergleichsgruppe. Eine Ausnahme bildet natürlicherweise der Einsatz der HyperDocs, die nur in der Interventionsgruppe verwendet wurden.

Die erste Stunde der Reihe wurde neu entwickelt und diente zum Kennenlernen der HyperDocs. Im Lehrplan ist die Stunde dem Themenfeld 10 „Gefährliche Stoffe“ (MBWWK, 2014, S. 84) zuzuordnen, da sie sich intensiver mit den Gefahrensymbolen auseinandersetzt. Ausgehend von den Gefahrstoffen in der Chemie wird eine Überleitung zu Stoffen geschaffen, die nicht unmittelbar als Gefahrstoffe gekennzeichnet sind – im vorliegenden Fall Azofarbstoffe. Bei der eigentlichen Unterrichtsreihe ab der zweiten Stunde handelt es sich um adaptierte Versionen von Czubatinski et al. (2020) für die Mittelstufe und Nieß et al. (2020) für die Oberstufe. Zentraler Gegenstand der Einheiten ist die gesundheitliche Einordnung eines Azofarbstoffes (Azorubin – E122), verbunden mit dessen photometrischer Konzentrationsbestimmung dessen. Der Alltagsbezug erfolgt über einen Grenadine-Sirup, der unter anderem den eben genannten Farbstoff enthält und für die Zubereitung von Mischgetränken eingesetzt wird. Die Themen haben eine

fachliche Relevanz und sind daher explizit in dem Lehrplan der Mittel- und Oberstufe verankert. Das Thema ist in der Mittelstufe nach dem rheinland-pfälzischen Lehrplan in Themenfeld 9 „Den Stoffen auf der Spur“ (MBWWK, 2014, S. 82) zugeordnet. Im Themenfeld steht die analytische Chemie im Vordergrund. Damit verbunden sind die Einführung des Konzentrationsbegriffs und die Kolorimetrie. Als mögliche Kontexte schlägt der Lehrplan „Tägliches Leben und Gesundheit“, ferner „Das Etikett – Was sagen mir die Informationen?“ (MBWWK, 2014, S. 83) vor. Als Vertiefung wird beispielsweise das Prinzip der Kalibrierung vorgeschlagen. Ferner bietet das Thema der Unterrichtsreihe auch Bezüge zu Themenfeld 10 „Gefährliche Stoffe“ (MBWWK, 2014, S. 84), indem beispielsweise der ADI-Wert als ein Grenzwert im gesundheitlichen Kontext eingeführt wird. Auch in der Oberstufe finden sich Bezüge zum Lehrplan. In der Einführungsphase der Oberstufe wird unter anderem der Konzentrationsbegriff wiederholt (MBWWK, 1998, S. 93). Der eigentliche Schwerpunkt lässt sich jedoch im Wahlpflichtbaustein „Analytik II – Chromatographie und Photometrie“ (MBWWK, 1998, S. 95) einordnen. Dort werden der Aufbau und die Funktion des Photometers, das Lambert-Beersche-Gesetz und der Begriff der Extinktion eingeführt. Außerdem finden sich Bezüge zur Lebensmittelchemie (E-Nummern) und zu Azofarbstoffen, wobei letztere nicht auf fachlicher Ebene vertieft werden. Eine genaue Beschreibung der einzelnen Inhalte der Unterrichtsreihe folgt weiter unten.

Aus den eben genannten Bezügen zu den rheinland-pfälzischen Lehrplänen wird die Relevanz der Unterrichtsreihe aus fachlicher Sicht ersichtlich. Die Reihe weist allerdings ebenso eine alltagsbezogene Relevanz für die Schüler:innen auf, wie auch Czubatinski et al. (2020) und Nieß et al. (2020) herausstellen. Die Lernenden sind mit einem realen Problem konfrontiert, das sie selbst betrifft. So befinden sich Azofarbstoffe in vielen Lebensmitteln aus unserem Alltag, obwohl gewisse gesundheitliche Bedenken bestehen, wenn eine hohe Menge der Farbstoffklasse zugeführt wird. Damit stellt sich generell die Frage, welche Stoffe in welchem Maße konsumiert werden können. Die gesundheitliche Thematik könnte besonders auch Schülerinnen interessieren, da humanzentrierte Themen von ihnen als interessanter wahrgenommen werden (Holstermann & Bögeholz, 2007).

Aufgrund der dargestellten fachlichen und alltäglichen Relevanz eignen sich die Unterrichtsstreihen für einen Einsatz im regulären Schulunterricht. Hinzu kommt, dass die Reihen von den Autor:innen in der Praxis erprobt wurden. Die Unterrichtsreihe für die Mittelstufe wurde mit vier achten und einer neunten Klasse durchgeführt. In der Erprobung

konnte eine hohe Motivation des Inhaltes und der Methode (Konzentrationsbestimmung) festgestellt werden. Auch die bereits erwähnte gesundheitliche Relevanz des Themas stieß bei den Schüler:innen auf Interesse (Czubatinski et al., 2020; Nieß et al., 2020). Da kein gesondertes Vorwissen für die Unterrichtsreihe notwendig ist (Czubatinski et al., 2020), eignet sich der Einsatz außerdem für eine Studie, die über das Schuljahr verteilt stattfindet und aus organisatorischen Gründen nicht auf ein bestimmtes Vorwissen aufbauen kann. Zudem handelt es sich bei der Photometrie beziehungsweise der Auswertung photometrischer Messungen mittels einer Kalibriergeraden trotz der hohen fachlichen Relevanz um ein Thema, das erfahrungsgemäß nur von wenigen Lehrkräften in vollem Umfang behandelt wird. Das ist unter anderem darin begründet, dass viele Schulen nur ein oder kein Photometer besitzen. Aus diesem Grund konnte mit einer größeren potenziellen Stichprobe gerechnet werden.

#### **Änderungen gegenüber der ursprünglichen Unterrichtsreihe**

Die Entwürfe der Autor:innen (Czubatinski et al., 2020; Nieß et al., 2020) sind als Doppelstunden geplant. Aus den Pilotierungen wurde ersichtlich, dass zwei Schulstunden nicht ausreichen, um die Inhalte vollständig abzuhandeln, unter anderem auch, weil zum Beispiel ein gewisser zeitlicher Puffer von circa 5 Minuten nach jeder Einzelstunde für die Beantwortung der Fragebögen eingeplant werden musste. Daher wurde für beide Stufen die Reihe auf drei Unterrichtsstunden ausgeweitet. Diese Änderung erwies sich als zielführend (vgl. Kapitel 4.4.2 & 4.4.3). Die konzipierten Arbeitsblätter wurden neu strukturiert und in HyperDocSystems übertragen. In der Mittelstufe wurden die Bilder der beiden Menschen bei der Berechnung des ADI-Wertes (Stunde 2) durch Bilder in besserer Qualität ersetzt. Um die Arbeitsblätter in HyperDocs umzuwandeln, mussten Lernhilfen und eine Vertiefungsaufgabe pro HyperDoc ergänzt werden. Bei den Vertiefungsaufgaben handelt es sich um ergänzende Aufgaben oder um Aufgaben, die auf dem ursprünglichen Arbeitsblatt vorhanden waren, jedoch aus Zeitgründen zur Vertiefungsaufgabe deklariert wurden. Die Eignung der Lernhilfen wurde mit Herrn Czubatinski geprüft und im Rahmen der zweiten und dritten Pilotierung getestet und abschließend überarbeitet. Bei den Lernhilfen wurde kein spezielles Konzept, wie beispielsweise gestufte Lernhilfen (Leisen, 1999), eingesetzt, da eine solche Einbindung mit den Schüler:innen geübt werden muss (Hänze et al., 2010) und Änderungen im Nutzungsverhalten daher wahrscheinlich erst über eine längere Zeit sichtbar werden. Bei den hier eingesetzten Lernhilfen handelt es sich um inhaltliche Unterstützungen, die bestimmtes Fachwissen



wiederholen oder erste Lösungsansätze beschreiben. Die Hilfen in Form von Text- und Bildhilfen sind Anhang 9.1.6 zu entnehmen. Es handelt sich um die Lernhilfen der Vergleichsgruppe, die inhaltlich identisch zu denen der Interventionsgruppe sind. Die auditiven Hilfen beinhalten den gleichen Inhalt wie die Texthilfen. Die Videohilfen liegen digital vor und können hier nicht abgebildet werden. Die ursprünglichen Arbeitsblätter und Stundenverlaufspläne der Mittelstufe können Czubatinski et al. (2020) entnommen werden, die der Oberstufe befinden sich in Anhang 9.1.8. Im Folgenden werden die Didaktik und Methodik der Unterrichtsstunden, die im Rahmen dieser Arbeit gehalten wurden, beschrieben. Viele Teile der Didaktik, wie zum Beispiel der Lehrgegenstand, und Teile der Methodik, beispielsweise das Experiment, waren bereits durch die Unterrichtsreihen von Czubatinski und Kolleginnen (2020) sowie Nieß et al. (2020) vorgeben. Sie werden dennoch explizit hier beschrieben, da Anpassungen vorgenommen wurden. Eine Übersicht der verwendeten Inhalte bietet Anhang 9.1.7.

Die Kompetenzen der nachfolgenden Stunden wurden für die Mittelstufe in Anlehnung an die Bildungsstandards für das Fach Chemie (KMK, 2004a) formuliert. Für die Oberstufe wurde sich an den Einheitlichen Prüfungsanforderungen in der Abiturprüfung Chemie (KMK, 2004b) orientiert.

### **Didaktik und Methodik der Unterrichtsstunden der Mittelstufe**

#### *Stunde 1: Safety First!*

Die erste Stunde ist sowohl für die Mittelstufe als auch für die Oberstufe identisch, da eine kurze Vorstellung der Abläufe und die Einführung in die Methode der HyperDocs in beiden Stufen gleichermaßen erforderlich ist. Schwerpunkt der Stunde ist daher der Umgang mit dem HyperDoc.

Didaktik. Für die Stunde wurde ein Thema gewählt, das kein notwendiges Vorwissen für die nachfolgenden Stunden darstellt und dennoch eine gewisse Fach- und Alltagsrelevanz für die Schüler:innen aufweist. Der Lerngegenstand sollte außerdem in circa 35 Minuten zu erarbeiten sein, damit den Lernenden ausreichend Zeit für die Bearbeitung des ersten Fragebogens zur Verfügung stand. Ein geeigneter Inhalt erschien daher das Thema der Gefahrensymbole im Alltag. Es ist dem Themenfeld 10 „Gefährliche Stoffe“ (MBWWK, 2014, S. 84) des rheinland-pfälzischen Lehrplans zuzuordnen, womit eine fachliche Relevanz gegeben ist. Aber auch im Alltag spielt das Thema eine Rolle. So begegnen den Jugendlichen auf nahezu allen Reinigungsmitteln Gefahrensymbole. Dieser Sachverhalt

wird für den Einstieg in die Stunde ausgenutzt, indem den Lernenden Spülmaschinentabs präsentiert werden.

In der eigentlichen Bearbeitung zeichnet die Lerngruppe zunächst ein Gefahrensymbol. Dazu führt das HyperDoc die Schüler:innen durch die einzelnen Funktionen der Methode. Die Zeichnung eines Gefahrensymbols dient der Übung der Verwendung des Schreibwerkzeugs von HyperDocSystems, da viele Lernende zu diesem Zeitpunkt wahrscheinlich noch nie mit einem Apple Pencil geschrieben haben. Die Verwendung des Pencils ist allerdings notwendig, um später in der dritten und vierten Stunde die jeweiligen Abbildungen ordnungsgemäß zu beschriften. Für „einfache“ Aufgaben können die Nutzer:innen zwischen der Eingabe über die Tastatur oder den Pencil wählen. Die zweite Aufgabe bildet eine Anwendung des erarbeiteten Wissens, indem die Lernenden verschiedenen Chemikalien und Alltagsgegenständen Gefahrensymbole zuordnen sollen. Gezeigt werden hierbei drei Bilder: Gasflaschen, Natronlauge und Deodorant Dosen. Durch die Aufgabe soll die Relevanz des Themas für den Unterricht und den Alltag verstärkt werden. Gleichzeitig müssen die Schüler:innen sich erneut mit den Gefahrensymbolen beschäftigen, womit das Fachwissen gefestigt wird.

Der Inhalt der Hilfen und deren didaktische Intention sind Anhang 9.1.5 zu entnehmen. In der Vertiefungsaufgabe sollen die Lernenden, angelehnt an die zweite Aufgabe, weitere Stoffe und deren mögliche Gefahrensymbole nennen. Die Aufgabe motiviert die Schüler:innen über ihren Alltag zu reflektieren und weitere Gefahrstoffe zu identifizieren. Damit wird erneut ein hoher Alltagsbezug geschaffen.

Das Ende der Stunde bildet eine gemeinsame Sicherungsphase, in der die zweite Aufgabe und die Vertiefungsaufgabe besprochen werden. Danach wird der Fragebogen bearbeitet.

#### *Kompetenz:*

- Mittelstufe: Die Schüler:innen beschreiben und begründen die Einordnung verschiedener Stoffe mithilfe ihrer typischen Eigenschaften als Gefahrstoff.
- Oberstufe: Die Schüler:innen erkennen die Zusammenhänge zwischen der Eigenschaft eines Stoffes und dessen Einordnung als Gefahrstoff.

#### *Lernziele:*

- Die Schüler:innen beschreiben das Aussehen verschiedener Gefahrensymbole von unterschiedlichen Gefahrstoffen (Deodorant, Natronlauge, brennbares Gas).
- Die Schüler:innen erklären das Gefahrenpotential verschiedener Gefahrstoffe (Deodorant, Natronlauge, brennbares Gas).

Methodik. Die Einführung in die HyperDocs erfolgt zunächst in Einzelarbeit, da die Schüler:innen ihre eigenen iPads erhalten. Der Einstieg in die eigentliche Unterrichtsreihe erfolgt in einem Unterrichtsgespräch. Dazu wird eine reale Verpackung des Spülmaschinenreinigers gezeigt, wodurch gegebenenfalls bei einigen Lernenden eine Wiedererkennung und damit stärkere Identifizierung des Lerngegenstandes stattfindet. Die Thematik wird durch den Einbezug der Schüler:innen geöffnet. Die Schüler:innen erklären, dass die Symbole auf mögliche Gefahrenquellen hindeuten. Durch einzelne Schüler:innenvorträge wird sich schrittweise dem Lerngegenstand, den Gefahrensymbolen, angenähert. Die Lernenden sollen, wie in einer regulären Unterrichtsstunde, aktiv am Unterricht beteiligt werden und nicht nur die HyperDocs in Stillarbeit bearbeiten. Die anschließende Erarbeitungsphase erfolgt in Einzelarbeit. Diese Vorgabe ergibt sich aus dem Studiendesign, da unter anderem für die Forschungsfragen in Bezug auf die Lernhilfen deren individuelle Nutzung erforderlich ist. Auch die Vertiefungsaufgabe ist Teil des Studiendesigns, wobei sie in diesem Fall auch als zeitlicher Puffer für die schnelleren und leistungstärkeren Schüler:innen dient. Die abschließende Sicherung erfolgt erneut in Form von Schülervorträgen und einem zusammenfassenden Unterrichtsgespräch, um eine reguläre Unterrichtsstunde abzubilden. Durch das Unterrichtsgespräch sollen die Lernenden außerdem motiviert werden, sich am Unterrichtsgeschehen zu beteiligen und ihre Lösung vorzustellen. Die Sicherung erfolgt mündlich, da keine umfassenden Antworten formuliert werden müssen.

### *Stunde 2: Farbstoffe in Lebensmitteln*

Didaktik. Der Stundeneinstieg erfolgt mit einem Grenadine-Sirup als gefärbtes Lebensmittel. Der Kontext „Farbstoffe in Lebensmitteln“ wirkt dabei sehr motivierend auf die Lerngruppe und zeichnet sich durch eine hohe Alltagsrelevanz aus, da der Sirup im Lebensmittelhandel einfach zu erwerben ist und von Jugendlichen konsumiert wird (Czubatinski et al., 2020). Farbstoffe finden sich darüber hinaus in Lebensmitteln wieder. Dabei spielen sowohl natürliche als auch künstliche Farbstoffe eine wichtige Rolle. Einigen Schüler:innen ist womöglich die Problematik der künstlichen Farbstoffe, insbesondere der Azofarbstoffe, bekannt. Ab der 9. und 10. Klasse geraten die Lernenden zunehmend in ein Alter, in dem auch gesundheitliche Fragen und Fragen bezüglich des eigenen Körpers eine übergeordnete Rolle spielen. Das Thema lässt sich aus fachlicher Sicht im Lehrplan verankern. Dazu bestehen einerseits Kontextbezüge zu Themenfeld 9 „Den Stoffen auf der Spur“ (MBWWK, 2014, S. 83) und Themenfeld 10 „Gefährliche Stoffe“

(MBWWK, 2014, S. 83). Bei letzterem spielt insbesondere der ADI-Wert eine größere Rolle.

Nach allgemeinen Aspekten zum Einsatz des Sirups, rückt das Etikett in den Vordergrund. Auf dem Etikett wird bezüglich zweier E-Nummern (E122: Azorubin und E102: Tartrazin) der Hinweis gegeben: „Kann die Aufmerksamkeit von Kindern und Jugendlichen beeinträchtigen“. Es handelt sich dabei um Azofarbstoffe. Nun tritt die individuelle Relevanz in den Vordergrund und die Lernenden stellen sich womöglich die Frage, wie gefährlich der Sirup ist und ob die Schädigung ab einer bestimmten Menge eintritt. Diese Gedanken werden als Problemfrage aufgegriffen und formuliert (Czubatinski et al., 2020).

Die Problemfrage begleitet die gesamte Unterrichtsreihe und wird immer wieder in den einzelnen Stunden aufgegriffen. In der Erarbeitung werden die notwendigen Begrifflichkeiten wie Azofarbstoffe, E-Nummern und auch der ADI-Wert behandelt. Dabei wird der Begriff „Azofarbstoff“ oberflächlich eingeführt und elementarisiert, da eine fachliche Auseinandersetzung mit Azofarbstoffen in der Mittelstufe nicht vorgesehen und in Bezug auf die Problemfrage nicht zielführend ist. Um ein besseres Verständnis für den ADI-Wert zu entwickeln, sollen die Lernenden in der ersten Aufgabe begründen, weshalb die Einheit „Milligramm pro Kilogramm Körpergewicht“ gewählt wurde. Ein gutes Verständnis für die Bedeutung der Einheit ist für den weiteren Verlauf der Unterrichtsreihe wichtig, da die Schüler:innen später ihre eigene maximale Verzehrmenge des Sirups berechnen sollen und eine erste Hinführung zur Konzentrationsangabe „Milligramm pro Liter“ gelegt werden kann (Czubatinski et al., 2020). In der zweiten Aufgabe folgen nun einfache Rechnungen in Bezug auf den ADI-Wert. Um die dritte Stunde der Reihe zu entlasten, wird in der dritten Aufgabe erfragt, welcher Farbstoff in Bezug auf seine gesundheitsschädliche Wirkung untersucht werden sollte. Azorubin besitzt einen niedrigeren ADI-Wert und ist daher schädlicher als Tartrazin (Czubatinski et al., 2020).

Erläuterungen zu den eingesetzten Lernhilfen befinden sich in Anhang 9.1.5. Bei der Vertiefungsaufgabe handelt es sich um eine Textaufgabe, in der die Lernenden nochmals mit dem ADI-Wert, dieses Mal in Bezug auf Cumarin und Zimt, rechnen.

Die Stunde endet mit der Sicherungsphase und, ausgehend von Aufgabe drei, mit dem Ausblick der Konzentrationsbestimmung des Farbstoffs im fertigen Mischgetränk. Die Analyse wird für die kommende Stunde in Aussicht gestellt.

*Kompetenz:*

- Die Schüler:innen nennen und beschreiben Azofarbstoffe mit ihren Eigenschaften im gesundheitlichen Kontext als bedeutende Stoffe.

*Lernziele:*

- Die Schüler:innen erklären die Notwendigkeit eines körperrgewichtbezogenen Grenzwerts (ADI-Wert).
- Die Schüler:innen berechnen die maximale tägliche Verzehrmenge von Azorubin für zwei unterschiedliche Personen.
- Die Schüler:innen erläutern unter Einbezug des ADI-Wertes, welcher Azofarbstoff (Azorubin / Tartrazin) für eine Konzentrationsbestimmung herangezogen werden sollte.

Methodik. Die Stunde beginnt mit einem Realobjekt, der Sirup-Flasche. Dadurch wird unmittelbar ein Alltagsbezug geschaffen. Normalerweise könnte die Flasche herumgereicht und den Lernenden gegeben werden. Aufgrund der Corona-Pandemie werden jedoch das Etikett und die Flasche auf dem HyperDoc auf dem Tablet abgebildet. Dadurch ist das Etikett für jede Schülerin und jeden Schüler gut lesbar. In einem Unterrichtsgespräch sollen die Schüler:innen ihre Erfahrungen und Anwendungskenntnisse bezüglich des Sirups im Allgemeinen wiedergeben. Dabei erfolgen die Einführung des Unterrichtsthemas und die Formulierung der Problemfrage. Die Möglichkeit der Unterrichtsbeiträge der Lernenden wirkt dabei aktivierend. Die Schüler:innen können durch das Gespräch über ihre unmittelbaren Erfahrungen im Umgang mit Getränkessirups berichten, was zur Plateaubildung beiträgt.

Im weiteren Verlauf der Stunde dient das Tablet als zentrales Erarbeitungswerkzeug. Die Lehrkraft übernimmt in der Erarbeitungsphase aufgrund der Lernhilfen eine untergeordnete Rolle, steht jedoch für Fragen zur Verfügung. Bei inhaltlichen Fragen wird weitestgehend auf die Lernhilfen verwiesen. Die Vertiefungsaufgabe kann, neben der Funktion als Forschungsgegenstand, erneut als zeitlicher Puffer und als vertiefende Übung angesehen werden.

Die Sicherung erfolgt in Form von Schüler:innenbeiträgen, wodurch die Lernenden unmittelbar in den Sicherungsprozess einbezogen werden. Zudem besteht die Möglichkeit von Rückfragen, falls Inhalte nicht vollständig verstanden wurden. Die Aufgaben werden exemplarisch an der Tafel gesichert.

### *Stunde 3: Farbstoffanalyse mit dem Smartphone*

Didaktik. Die Stunde beginnt mit einer kurzen Wiederholung der ursprünglichen Problemfrage. Wie bereits angedeutet, durchzieht die Problemfrage alle drei Unterrichtsstunden, da sie sich nicht innerhalb einer Stunde beantworten lässt. Danach wird eine geeignete Untersuchungsmethode vorgestellt, um die Farbstoffmenge zu bestimmen.

Zunächst steht der optische Farbvergleich mit Kalibrierlösungen im Vordergrund. Hierbei muss davon ausgegangen werden, dass keinerlei Vorerfahrungen vorhanden sind. Den Lernenden werden daher die Lösungen in aufsteigender Konzentration präsentiert. In einem anschließenden Gespräch wird herausgearbeitet, dass die Getränkeprobe mittels eines Farbabgleichs zugeordnet werden kann. Bereits Czubatinski et al. (2020) weisen darauf hin, dass den Schüler:innen der Bezug zwischen einer Menge und der stofflichen Erscheinung aus dem Alltag bekannt ist, indem beispielsweise eine Suppe mit zunehmende Salzmenge salziger schmeckt oder die Teefarbe mit längerer Ziehzeit intensiver wird. Die Zuordnung mit dem bloßen Auge auf der stofflichen Ebene ist allerdings ungenau und nicht immer eindeutig. Aus diesem Grund wird die Relevanz für eine genauere Messmethode ersichtlich. Daher wird eine Untersuchungsmethode mit dem Smartphone / Tablet präsentiert und angewandt. Das Messprinzip ist in Czubatinski et al. (2020) zu finden und funktioniert durch die Messung von RGB-Werten. An dieser Stelle kann gegebenenfalls auf das Vorwissen der Jugendlichen bezüglich RGB-Werte zurückgegriffen werden, da die additive Farbmischung auch in LED-Displays Anwendung findet (Czubatinski et al., 2020).

In der schülerzentrierten Phase werden nun die gesicherten Messwerte in die HyperDocs übertragen und von den Schüler:innen eine Kalibriergerade erstellt. Anhand dieser erfolgt die Bestimmung der Farbstoffkonzentration im Mischgetränk. Auf eine fachliche Vertiefung der Photometrie wird verzichtet, da sie in der Stunde kein Lerngegenstand, sondern lediglich ein Methodenwerkzeug darstellt. Alternativ hätten die Schüler:innen an dieser Stelle auch eine vollautomatische Auswertung mittels eines Tabellenkalkulationsprogrammes durchführen können, allerdings wäre damit der Zusammenhang zwischen der Konzentration und den Messwerten weniger intensiv herausgearbeitet worden.

Die Inhalte und didaktische Intentionen der Hilfen sind in Anhang 9.1.5 zu finden. Die Vertiefungsaufgabe beschäftigt sich mit der Frage, welche Wertreihe (Rot, Grün oder Blau) am besten für die Auswertung herangezogen wird.

In der Sicherungsphase werden die Ergebnisse im Plenum besprochen. Dabei wird die Frage aus der Vertiefungsaufgabe aufgeworfen, welcher Farbanteil nun zur Konzentrationsbestimmung herangezogen wird. Es ist schnell ersichtlich, dass der Rotwert aufgrund des un stetigen Verlaufs nicht zur Bestimmung herangezogen werden kann. Damit bleibt, den Blau- und Grünwert zu diskutieren. Hierbei müssen die Schüler:innen erkennen, dass der Grünwert mit steigender Konzentration des Farbstoffs stärker fällt als der Blauwert und damit ein genaueres Ablesen einer unbekannt Probe ermöglicht (Czubatinski et al., 2020). Zusätzlich folgen die Blauwerte, je nach Messung, nicht immer einem linearen Verlauf, sodass sie sich zum Zeichnen einer Kalibriergeraden weniger eignen. Eine Beantwortung der ursprünglichen Problemfrage ist in der Stunde aus Zeitgründen nicht mehr möglich. Sie erfolgt in der letzten Stunde.

*Kompetenz:*

- Die Schüler:innen finden in einer Messreihe zur Bestimmung der Farbstoffkonzentration, Trends und Beziehungen, erklären diese und ermitteln daraus die Konzentration von Azorubin in einer unbekannt Probe.

*Lernziele:*

- Die Schüler:innen beschreiben das Vorgehen einer halb-quantitativen Bestimmung der Farbstoffkonzentration mithilfe einer Kalibrierreihe.
- Die Schüler:innen zeichnen drei Kalibrierungsgeraden anhand der Rot-, Grün-, und Blauwerte einer smartphone-photometrischen Messung.
- Die Schüler:innen bestimmen die Konzentration des Farbstoffs Azorubin in der Getränkeprobe mithilfe einer Kalibriergeraden.

Methodik. Die Hinführung zum Problem erfolgt mittels eines Lehrer:innenvortrags, da damit die relevanten Informationen gegeben werden können. Die Materialien für die Durchführung des Versuchs werden in einem kleinen Aluminiumkoffer transportiert. Aus diesem Grund werden die Kalibrierlösungen in kleinen Zellkulturfläschchen aufbewahrt. Einerseits können damit die Lösungen einfach pipettiert werden, andererseits lässt sich damit die optische Einordnung der Getränkeprobe in die Reihe leichter durchführen. Durch ein weißes Blatt, das hinter den Lösungen positioniert wird, kann der optische Farbabgleich genauer erfolgen.

Die Methode wird durch die Lehrkraft vorgestellt und den Schüler:innen der Begriff „RGB-Werte“ beschrieben. Die eigentliche photometrische Messung mittels der mobilen Endgeräte erfolgt in einem Lehrer:innenexperiment. Damit soll eine Interferenz im

Studiendesign bezüglich der HyperDocs vermieden werden. Zusätzlich erfordert die Durchführung als Schüler:innenexperiment eine größere Vorbereitungs- und Durchführungszeit, die im Rahmen der Studie nicht gewährt werden kann. Die Bestimmung der Messreihe wird verbal durch die Lehrkraft begleitet und die Messergebnisse anschließend an der Tafel gesichert, sodass die Schüler:innen diese übernehmen können. Durch einen kurzen Impuls wird das Vorgehen anhand eines Messpunktes verdeutlicht. Danach werden die Messwerte zunächst in das Diagramm übertragen (Aufgabe 2). Anschließend muss aus den jeweiligen Wertereihen eine Kalibriergerade erstellt und die Konzentration anhand der Geraden bestimmt werden (Aufgabe 3). Dazu müssen die Lernenden die Werte mittels einer Linie verbinden. Dabei dürfte die Konzentrationsbestimmung über diese Methode den Lernenden unbekannt sein (Czubatinski et al., 2020). Die Schüler:innen sollten jedoch erkennen, dass der Messwert der Probe zwischen zwei Werten der Kalibrierlösungen und damit auch die Konzentration zwischen der der beiden Messpunkte liegt. Über eine graphische Auswertung kann schließlich der genaue Messwert ermittelt werden.

In der Sicherungsphase werden die Kalibriergeraden von der Lehrkraft auf der Tafel gesichert, sodass die Lernenden ihre Ergebnisse abgleichen können, und die Schüler:innen erklären die Auswertung anhand der Geraden. Auf digitale Lösungen wurde verzichtet, da die Ausstattung der Schulen im Vorfeld nicht bekannt war und für alle Klassen die gleichen Bedingungen herrschen sollten.

#### *Stunde 4: Der Konzentrationsbegriff*

Didaktik. Die Stunde beginnt mit einer kurzen Wiederholung der letzten Unterrichtsstunde, wodurch auf die ursprüngliche Problemfrage verwiesen wird, wie viel von einem Mischgetränk des Sirups getrunken werden darf. Anschließend wird nach der Wiederholung, die rein auf der stofflichen Ebene erfolgt, der Blick auf die Teilchenebene geworfen, indem sich dem Konzentrationsbegriff auf dieser angenähert wird. Dazu sollen die Schüler:innen in der ersten Aufgabe die jeweiligen Kalibrierlösungen im Teilchenmodell darstellen. Die Lösung mit der Konzentration 50 mg / l des Farbstoffes ist bereits vorgegeben. Das Modell besteht aus einem rechteckigen Rahmen, in dem Kreise eingezeichnet sind. Die Lernenden müssen erkennen, dass pro 10 mg des Stoffes ein Teilchen ausgemalt werden muss und der Rahmen das Äquivalent von einem Liter der Lösung darstellt. Die ausgemalten Teilchen stellen damit die Farbstoffteilchen dar, während die anderen Teilchen Wasserteilchen sind. Aus Gründen der Elementarisierung werden alle Teilchen als



Kreise dargestellt. Alternativ hätten die Wasserteilchen auch mit Dreiecken oder ähnlichen Formen präsentiert werden können, allerdings war das Vorwissen der Klassen nicht bekannt, sodass diese Darstellung die Schüler:innen gegebenenfalls verwirrt hätte. Das Teilchenmodell stellt im Chemieunterricht ein zentrales methodisches Werkzeug dar, um chemische Vorgänge darstellen zu können und damit das Verständnis von nicht sichtbaren Vorgängen zu fördern. Es hat damit auch eine hohe Relevanz für die anderen Teilgebiete der Mittel- und Oberstufenchemie.

Nach einer kurzen Sicherungsphase erfolgt die Definition des Konzentrationsbegriffs, die auf dem HyperDoc gesichert ist. In der zweiten Aufgabe sollen die Lernenden nun die Menge an Farbstoff berechnen, die sie konsumieren dürfen. Dazu wird in der Aufgabe erneut der ADI-Wert von Azorubin genannt. Nachdem die Schüler:innen ihre individuelle Verzehrmenge des Mischgetränks berechnet haben, wird der Lösungsweg gesichert und das Ergebnis eingeordnet. In den meisten Fällen wird die Verzehrmenge deutlich über fünf Litern liegen, womit das Gefahrenpotential gering erscheint. Das Ergebnis wird außerdem in der Anzahl an Gläsern ausgedrückt. Einerseits kann damit das Rechnen mit Gehaltsangaben geübt werden und andererseits wird die maximale Trinkmenge nochmals veranschaulicht und die Problemfrage beantwortet. In diesem Zusammenhang wird dann vertiefend die Relevanz des Mischungsverhältnisses aufgeworfen. Die untersuchte Probe wurde nach den Vorgaben des Herstellers angefertigt. Es kann allerdings auch mehr Sirup gewählt werden, womit die Konzentration des Azofarbstoffs sich erhöht. Dieser Sachverhalt kann diskutiert werden und leitet schließlich in eine erneute Betrachtung des Teilchenmodells ein. Die Schüler:innen sollen nun das Teilchenmodell für den unverdünnten Sirup ergänzen und anschließend die Konzentration an Azorubin im unverdünnten Sirup, ausgehend von dem angegebenen Mischungsverhältnis, berechnen.

Die Inhalte der Hilfen und deren didaktische Intention sind in Anhang 9.1.5 aufgeführt. Die Vertiefungsaufgabe bildet eine Anwendung der erarbeiteten Inhalte, indem die Lernenden berechnen sollen, ob ein Kleinkind bei einem Verzehr des Getränks die erlaubte Tagesdosis von Azorubin erreicht. Mit der Vertiefungsaufgabe wird erneut ein Alltagsbezug hergestellt und das Rechnen mit den Gehaltsangaben geübt.

*Kompetenz:*

- Die Schüler:innen nutzen das Teilchenmodell, um die Konzentration des Farbstoffs Azorubin in verschiedenen Lösungen darzustellen.

*Lernziele:*

- Die Schüler:innen stellen die Konzentrationen der einzelnen Kalibrierlösungen und der Probelösung im Teilchenmodell dar.
- Die Schüler:innen berechnen die maximale Tagesdosis eines zubereiteten Getränks anhand des eigenen Körpergewichtes und des ADI-Wertes von Azorubin in Liter und Anzahl der Gläser.
- Die Schüler:innen stellen das Mischungsverhältnis und die Konzentration des Farbstoffs im unverdünnten Sirup im Teilchenmodell dar.

Methodik. Der Einstieg in die Stunde erfolgt in einer Plenumsphase, in der in einzelnen Schüler:innenvorträgen die bisherigen Erarbeitungen zusammengefasst und alle Lernenden auf den gleichen Stand gebracht werden. Danach folgt die erste Erarbeitungsphase mit einer anschließenden Sicherung an der Tafel. Die Definition für den Konzentrationsbegriff wird direkt auf dem HyperDoc präsentiert, da die Definition von den Schüler:innen nicht selbstständig erarbeitet werden kann. Sie wird in einem Schüler:innenvortrag vorgelesen. In diesem Zusammenhang können die Lernenden auch Rückfragen zur Definition stellen. Die zweite Erarbeitungsphase wird ebenfalls an der Tafel gesichert. Die Einordnung der Ergebnisse findet im Unterrichtsgespräch statt, um eine Diskussion der Ergebnisse im Klassenverband zu ermöglichen. Damit sollen möglichst viele Ergebnisse und Meinungen eingeholt werden, um auch ein breites Bewusstsein für die Thematik zu schaffen. So wird zum Beispiel die Spannweite der maximalen Verzehrmenge, abhängig vom Körpergewicht, deutlich. Die letzte Erarbeitungsphase wird erneut an der Tafel und mit Schüler:innenbeiträgen gesichert. Die Vertiefungsaufgabe wird besprochen, falls dazu Zeit in der Stunde zur Verfügung steht.

**Didaktik und Methodik der Unterrichtsstunden der Oberstufe**

Die erste Stunde ist identisch mit der der Mittelstufe. Die weiteren Stunden der Oberstufe werden nun beschrieben.

*Stunde 2: Farbstoffanalyse mit dem Smartphone*

Didaktik. Die zweite Stunde der Oberstufe bildet inhaltlich eine Mischung aus der zweiten und dritten Stunde der Mittelstufe. Den Schüler:innen wird die Sirupflasche präsentiert und nach deren Erfahrungen gegenüber einem solchen Getränk gefragt (Nieß et al., 2020). Damit sollen die Lernenden von Beginn an in die Problematik der Stunde eingebunden werden. Dadurch zeigt sich die Relevanz des Themas, womit mit einer höheren Motivation zu rechnen ist. Anschließend wird auf das Etikett des Sirups verwiesen,

wonach die Farbstoffe E122 und E102 die Aufmerksamkeit von Kindern beeinträchtigen können. In diesem Zusammenhang wird nun die Problemfrage formuliert, wie viel eines Mischgetränks getrunken werden kann, bis eine schädigende Wirkung eintritt. Hierbei sind Fragen zu erwarten, die auf die Wirkungsmechanismen im Körper abzielen oder direkt die erlaubte Trinkmenge fokussieren. Bei ersterem kann darauf verwiesen werden, dass im Chemieunterricht nicht die nötigen Mittel und Methoden vorliegen, um die Frage zu beantworten und sie daher der Biologie zuzuordnen ist. Bei letzterem können die Schüler:innen unmittelbar die Problemfrage formulieren, wie viel eines Getränks mit diesem Sirup bedenkenlos getrunken werden kann. Die Lösung der Problemfrage wird in zwei Schritten erarbeitet: Erstens muss ein Grenzwert festgelegt werden und zweitens muss die Menge des Farbstoffs im Sirup bekannt sein. Die zweiteilige Problematik wird den Schüler:innen über eine kurze Information über die E-Nummern und die Azofarbstoffe, sowie den ADI-Wert verdeutlicht. Diese Informationen erarbeitet sich die Mittelstufe in der zweiten Stunde schriftlich. Auch in der Oberstufe wird auf eine fachliche Vertiefung der Azofarbstoffe verzichtet. Die Inhalte werden nicht selbst erarbeitet, da die Oberstufe zwar leistungsstärker ist, aber in der nächsten Stunde bereits die Photometrie als Methode vertieft werden soll. Insgesamt steht somit weniger Zeit als in der Mittelstufe für diese Thematik zur Verfügung. Der erhöhte Anforderungsbereich in der Oberstufe gegenüber der Mittelstufe ergibt sich nicht zwangsläufig durch den Inhalt, sondern durch dessen Aufbereitung. Das weitere didaktische Vorgehen entspricht ab der Erarbeitungsphase dem der dritten Stunde in der Mittelstufe, da die Bestimmung des Farbstoffs nach dem gleichen Versuch durchgeführt wird. Unterschiede bestehen jedoch in der Aufgabenführung. Die Oberstufe erhält weniger Hinweise in den Aufgaben, wie die Kalibriergeraden zu erstellen sind. Außerdem wird die Begründung, welche Wertereihe für die Auswertung genutzt wird, nicht als Vertiefungsaufgabe formuliert, sondern sie ist explizit von jedem zu bearbeiten.

In der Vertiefungsaufgabe wiederum sollen die Lernenden auf Teilchenebene erklären, wie die Farbigkeit eines Stoffes zustande kommt. Sie müssen dabei erkennen, dass zum Beispiel roterscheinende Stoffe „rotes Licht“ (Licht einer bestimmten Wellenlänge) reflektieren und „grünes Licht“ absorbieren. Die Inhalte und didaktische Intention der Hilfen sind in Anhang 9.1.5 zu finden. Die Stunde endet mit einer Sicherung der Ergebnisse und dem Rückbezug zur Ausgangsfrage.

*Kompetenz:*

- Die Schüler:innen tragen Messwerte in ein Koordinatensystem ein und erstellen eine Kalibriergerade, um daraus die Azorubinkonzentration einer unbekannt Probe zu ermitteln.

*Lernziele:*

- Die Schüler:innen erklären das Vorgehen bei der halb-quantitativen Bestimmung der Farbstoffkonzentration mithilfe einer Kalibrierreihe.
- Die Schüler:innen bestimmen die Konzentration des Farbstoffs Azorubin in der Getränkeprobe mithilfe einer Kalibriergeraden.
- Die Schüler:innen begründen die Auswahl der Grünwerte für die Bestimmung der Konzentration von Azorubin in der Getränkeprobe.

Methodik. Den Schüler:innen wird die Sirupflasche präsentiert, dann erfolgt der Einstieg mittels Schüler:innenvorträgen und einem anknüpfenden Unterrichtsgespräch. Das Unterrichtsgespräch wurde gewählt, um möglichst viele Erfahrungen im Umgang mit dem Sirup zu sammeln. Damit soll die Relevanz des Themas für die Jugendlichen herausgestellt werden. Ein:e Schüler:in liest die Inhaltsstoffe und den Warnhinweis auf dem Etikett vor, um eine hohe Schüler:innenbeteiligung zu erreichen. Die Informationen über die E-Nummern, den ADI-Wert und die Azofarbstoffe werden in einem Lehrer:innenvortrag dargestellt. Dabei kann ausstattungsgründen auf keine Visualisierung zurückgegriffen werden, allerdings handelt es sich um Informationen, die den Schüler:innen zum Teil bereits bekannt sein sollten beziehungsweise, die keiner umfassenderen Darstellung bedürfen. Die ADI-Werte der beiden Farbstoffe werden an der Tafel notiert und für die Hinführung zum Experiment genutzt.

Das Experiment selbst wird aus Organisations- und Zeitgründen als Lehrer:innenexperiment durchgeführt. Das methodische Vorgehen und die Auswertung auf dem HyperDoc sind analog zur Mittelstufe.

Die Sicherung erfolgt in einem Unterrichtsgespräch an der Tafel. Dazu wird ein Diagramm an der Tafel gezeichnet, woran auch die Auswahl der Messreihe (Grünwerte) diskutiert werden kann.

*Stunde 3: Das Spektralphotometer*

Didaktik. Die Stunde behandelt das Thema „Spektralphotometrie“ als grundlegende wissenschaftliche Methode der Konzentrationsbestimmung eines Stoffes. Das Thema besitzt

daher eine hohe fachliche Relevanz, die anhand eines Wahlpflichtbausteins im rheinland-pfälzischen Lehrplan der Oberstufe verdeutlicht wird.

Der Einstieg in die Stunde erfolgt über die Reflexion der letzten Unterrichtsstunde. Dabei wird die Frage aufgeworfen, wie in den Wissenschaften und den Laboren die Konzentration von Stoffen in Lösungen bestimmt wird. In einem Gespräch werden mögliche Nachteile des „Smartphotometers“ angesprochen (z.B. unterschiedliche Smartphones, Lichteinfall), wodurch die Relevanz einer wissenschaftlich genauen Methode ersichtlich wird. Durch die Diskussion kann das wissenschaftspropädeutische Vorgehen geschult werden. Der Einstieg mündet in die Erarbeitungsphase, in der sich die Lernenden über das Spektralphotometer informieren. Sie sollen anschließend den Aufbau eines Spektralphotometers beschriften und den Begriff der Extinktion erklären. Beide Aufgaben werden anschließend gesichert.

In der weiteren Erarbeitungsphase werden Grundlagen der Messmethode, wie das Erstellen eines Absorptionsspektrums und die Relevanz des Absorptionsmaximums, erklärt. In diesem Schritt werden die Absorptionsmaxima der beiden Azofarbstoffe gezeigt. Dadurch wird ersichtlich, dass die Farbstoffe getrennt voneinander bestimmt werden können. Es folgt eine Zusammenfassung, wie eine photometrische Messung durchgeführt wird.

Zum Ende der Stunde sollen die Lernenden die Methode des Spektralphotometers mit der des „Smartphotometers“ vergleichen. Hierbei sind einige Gemeinsamkeiten und Unterschiede zu nennen. Zu den Gemeinsamkeiten zählen die Ermittlung von „Messwerten“ über einen Sensor, wobei sich die genaue Methode der Erfassung unterscheidet, und die Messung mehrerer Kalibrierlösungen. Der Vergleich der beiden Methoden dient als Rückbezug zum Ausgangsproblem. Zudem soll durch die Aufgabe die nächste Stunde entlastet werden.

Inhalte und didaktische Intention der Hilfen befinden sich in Anhang 9.1.5. Die Vertiefungsaufgabe beinhaltet eine Rechnung, die den Bezug zwischen der Konzentration und der Extinktion herstellt. Es handelt sich damit um eine vertiefende Übung.

*Kompetenz:*

- Die Schüler:innen nutzen verschiedene Texte, erkennen darin fachliche Kernaussagen, wählen diese gezielt und kritisch aus und verknüpfen diese mit dem erworbenen Wissen, um sich die Methode einer photometrischen Messung zu erschließen.

*Lernziele:*

- Die Schüler:innen beschreiben den Aufbau eines Spektralphotometers mit Fachbegriffen.
- Die Schüler:innen erklären den Begriff der „Extinktion“.
- Die Schüler:innen erläutern das Vorgehen bei der Durchführung einer photometrischen Messung.
- Die Schüler:innen vergleichen das Smartphotometer mit dem Spektralphotometer anhand selbst gewählter Kriterien.

Methodik. Der Einstieg in die Stunde erfolgt nach der Wiederholung durch Schüler:innenvorträge über einen kurzen Lehrer:innenvortrag, um das Spannungsfeld zwischen der wissenschaftlichen Methode der Spektroskopie und dem „Smartphotometer“ aufzuwerfen.

Die anschließende Erarbeitung anhand eines Textes ermöglicht eine selbstständige Auseinandersetzung mit dem Thema, die den Schüler:innen den Freiraum der individuellen Bearbeitung ermöglicht. Bei dem verwendeten Text wurde sich an einer Schulbuchseite orientiert, sodass die Thematik ausreichend elementarisiert wurde. Aufgrund der Komplexität der Messmethode ist zu erwarten, dass die Lernenden sich in unterschiedlichem Tempo die Lerninhalte erschließen, weshalb für diese erste Phase etwas mehr Zeit eingeplant wurde. Die Beschriftung des Photometers wird anhand einer Zeichnung an der Tafel besprochen. Die Zeichnung kann gut genutzt werden, um den Begriff der Extinktion zu besprechen, indem dort die Intensität des Lichtstrahls ergänzt wird. Beide Inhalte werden durch Schüler:innenvorträge gesichert.

Die zweite Erarbeitungsphase findet erneut mithilfe eines kurzen Textes statt. Die Sicherung wird anhand einer Skizze des Absorptionsspektrums an der Tafel visualisiert. Der Abschluss der Stunde erfolgt, nach einer kurzen schriftlichen Erarbeitung der Aufgabe, mündlich, ohne weitere Visualisierungen. Den Lernenden wird dafür der Versuch erneut vorne aufgebaut, um die beiden Messmethoden besser miteinander vergleichen zu können.

*Stunde 4: Die optimierte Farbstoffanalyse mit dem Smartphone*

Didaktik. Die vierte Unterrichtsstunde dient zum einen der vertiefenden Betrachtung der Smartphotometer-Methode und zum anderen der abschließenden Beantwortung der Frage, wie viele Gläser des zubereiteten Getränks des Sirups getrunken werden dürfen.

Durch die Beantwortung der abschließenden Frage ist eine hohe Motivation auf Seite der Schüler:innen zu erwarten.

Im Einstieg werden die Unterschiede zwischen der Spektralphotometrie und dem Smartphotometer kurz gegenübergestellt. Anschließend wird die Frage aufgeworfen, wie der Aufbau beziehungsweise die Durchführung des Versuchs mit dem Smartphotometer verbessert werden könnte. Mit dieser Frage werden die Lernenden in eine kurze Erarbeitungsphase entlassen.

Die Ergebnisse werden anschließend anhand des Versuchsaufbaus besprochen und umgesetzt. Es handelt sich dabei um den „Weißabgleich“, der durch die App ermöglicht wird. Er ist mit einer Blindprobe gleichzusetzen. Für den Abgleich wird Tartrazin der im Sirup vorkommenden Konzentration gewählt. Die zweite Änderung betrifft die Variation des Durchlichts, für das nun eine grüne Farbe gewählt wird. Damit soll die Idee des Absorptionsmaximums bei einer bestimmten Wellenlänge aufgegriffen werden. Die Durchführung und Sicherung sind, unter Berücksichtigung der Anpassungen, analog zu denen der zweiten Stunde. In der Auswertung entfallen die Blau- und Rotwerte. Die Ergebnisse der Messung werden kurz besprochen und eine Abweichung zur vorherigen Messung bestimmt. Den Schüler:innen wird anschließend der genaue Wert in der Probe, die mit einem Photometer bestimmt wurde, mitgeteilt. Die Abweichung zum Smartphotometer beträgt in der Regel weniger als fünf Prozent, womit die Genauigkeit der Methode bestätigt wird.

Mit dieser Erkenntnis kann eine zuverlässige Berechnung der erlaubten Verzehrmenge pro Tag erfolgen. Wie in der Mittelstufe, berechnen die Schüler:innen für ihr eigenes Körpergewicht die maximale Verzehrmenge. Diese wird ebenfalls einen hohen Wert haben. Die Ergebnisse werden abschließend eingeordnet und diskutiert. Hierbei kann ebenfalls auf das Mischungsverhältnis oder das Körpergewicht, zum Beispiel bei Kleinkindern, eingegangen werden.

Die Inhalte und didaktische Intention der Hilfen sind Anhang 9.1.5 zu entnehmen. Die Vertiefungsaufgabe beinhaltet die Berechnung der Farbstoffkonzentration im unverdünnten Sirup. Es handelt sich um eine Übungsaufgabe zum Rechnen mit Mengenangaben.

*Kompetenz:*

- Die Schüler:innen können selbstständig einen optimierten Versuch zum Smartphotometer planen, beobachten, beschreiben und auswerten.

*Lernziele:*

- Die Schüler:innen erklären Änderungen am Versuchsaufbau zur Optimierung des Smartphotometers.
- Die Schüler:innen berechnen die maximale Tagesdosis eines zubereiteten Getränks anhand dessen Farbstoffkonzentration, des eigenen Körpergewichtes und des ADI-Werts von Azorubin in Liter und Anzahl der Gläser.
- Die Schüler:innen erläutern, wann bei Verzehr die maximale Tagesdosis eines zubereiteten Getränks trotz des hohen ADI-Wertes überschritten werden kann (geringes Körpergewicht, Kleinkinder, ungünstiges Mischungsverhältnis).

Methodik. Der Einstieg in die Stunde erfolgt über ein Unterrichtsgespräch, um wieder möglichst viele Schüler:innen am Unterrichtsgeschehen zu beteiligen. Die eingeschobene Erarbeitungsphase wird mündlich gesichert. Dazu wird den Lernenden nochmal der Aufbau des Smartphotometers gezeigt, da dieser für die erneute Durchführung des Versuchs erforderlich ist.

Der Versuch selbst wird als Lehrer:innenexperiment durchgeführt, um erneut Interferenzen mit dem Untersuchungsdesign der Studie zu vermeiden und um Zeit zu sparen. Die Unterschiede im Versuchsaufbau werden nochmals im Lehrer:innenvortrag während der Durchführung beschrieben, um den Schüler:innen diese Änderungen am Realobjekt (hier: Smartphotometer) zu zeigen. Damit soll das Verständnis erhöht werden. Die Messwerte aus dem Versuch werden an der Tafel präsentiert, von den Lernenden übernommen und schließlich grafisch auf dem HyperDoc beziehungsweise an der Tafel ausgewertet.

Die Diskussion über die Messwerte findet im Plenum statt, um möglichst viele Ergebnisse und damit die Spannweite der Trinkmenge zu erfassen. Dadurch kann eine breitere Diskussionsgrundlage geschaffen werden.

#### **4.5.4 Gestaltung der Vergleichsgruppe**

Bei einer experimentellen Studie stellt sich die Frage einer adäquaten Gestaltung der Vergleichsgruppe, damit die Wirksamkeit der Intervention auch wirklich auf die untersuchte Variable zurückzuführen ist. Um die interne Validität der Studie zu sichern, wurde die Studie der Vergleichsgruppe so gut wie möglich an die der Untersuchungsgruppe angepasst. Dafür wurde bereits die Didaktik und Methodik der Stunden im vorherigen Kapitel beschrieben. Im Folgenden wird auf die interne und externe Validität der Studie eingegangen.



### **Interne Validität**

Die interne Validität bezieht sich auf die Frage, inwieweit andere Faktoren als die unabhängige Variable Einfluss auf die Untersuchungsergebnisse beziehungsweise die abhängige Variable genommen haben (Rost, 2013, S. 123). So genannte „Störfaktoren“ (Schnell et al., 2018, S. 190) können demnach zu falschen Schlussfolgerungen bei der Auswertung der abhängigen Variablen führen. Dabei gibt es eine Reihe dieser Faktoren, auf die nachfolgend nach Schnell et al. (2018, S. 190 - 194) und Rost (2013, S. 124 - 128) eingegangen wird.

Unter dem zwischenzeitlichen Geschehen versteht man eine Veränderung im Verhalten einer Versuchsperson, bedingt durch ein bestimmtes Ereignis, das zum Beispiel während einem Pre-Post-Design auftritt. Da die Studie im Rahmen des Regelunterrichts stattgefunden hat, können Einflüsse, zum Beispiel durch eine Klassenarbeit, die unmittelbar vor oder nach einer Untersuchungsstunde geschrieben wurde, nicht ausgeschlossen werden. Das kann einen Einfluss auf die Motivation genommen haben. Ähnliches gilt für andere Leistungsnachweise oder für den Zeitpunkt der Stunde (erste oder sechste Stunde).

Ein weiterer Faktor ist der Reifungsprozess der Proband:innen. Die Änderung in der abhängigen Variable kann auch durch „intrapersonale Prozesse“ (Schnell et al., 2018, S.191) hervorgerufen werden. Es handelt sich dabei um alle andauernden Veränderungen der Versuchspersonen, die nicht auf die experimentelle Manipulation zurückzuführen sind. Solche Reifungsprozesse, bedingt durch zum Beispiel biologische Entwicklungsprozesse, waren bei dieser Studie eher auszuschließen, da die Intervention ein bis zwei Wochen in Anspruch nahm und keine gesonderten kurzfristigen Ereignisse stattfanden. Auf einen möglichen Einfluss der Corona-Pandemie, auch unter der Betrachtung des Störfaktors „Zwischenzeitliches Geschehen“, wird in Kapitel 4.5.5 eingegangen.

Änderungen am Messinstrument, wie zum Beispiel Umformulierung von Items spielen ebenfalls eine Rolle. Schnell et al. (2018, S. 192) zählen dazu auch noch Änderungen bezüglich der Versuchsleiter. Die Messinstrumente wurden in dem vorliegenden Fall nicht verändert. Das Verhalten des Versuchsleiters wurde ebenfalls nicht verändert. Es wurde stets auf eine offene und freundliche Art geachtet. Der Unterricht erfolgte nach der zuvor ausgearbeiteten Didaktik und Methodik, sodass der Ablauf zwischen den Untersuchungsgruppen identisch war.

Ebenfalls von Relevanz sind eine verzerrte Auswahl und Ausfälle. Bei diesem Punkt geht es um die ungleiche Auswahl der Individuen für die Untersuchungs- und

Vergleichsgruppe oder um systematische Ausfälle von Versuchspersonen im Laufe einer Studie. Auch wenn die einzelnen Schulklassen für die Untersuchungs- und Vergleichsgruppe nicht völlig zufällig gezogen wurden, konnten sich die einzelnen Proband:innen nicht einer bestimmten Gruppe zuordnen. In Teilen fehlten Schüler:innen aufgrund einer Erkrankung an einzelnen Tagen der Erhebung. Ein systematisches Fehlen, um der Studie zu entgehen, ist hier auszuschließen, da die Teilnahme an der Studie freiwillig war und weitere Schulstunden durch das Fehlen versäumt würden. Es gibt keinen erkennbaren Grund, weshalb Schüler:innen die Untersuchungsstunden absichtlich meiden sollten.

Auch, wenn Schnell et al. (2018, S. 190 - 194) und Rost (2013, S. 124 - 128) bei der internen Validität nicht explizit auf die realistische Umsetzung und Gestaltung der Vergleichsgruppe verweisen, soll an dieser Stelle auf die Gestaltung der Materialien für die Vergleichsgruppe eingegangen werden. Die abhängige Variable, hier die Intervention mittels analoger oder digitaler Arbeitsblätter, wurde bewusst manipuliert, um einen vollständigen Vergleich zur analogen Variante der Arbeitsblätter abzubilden. Dennoch sollten auch die Inhalte und die Darstellung beider Formate einander ähneln, sodass nicht schon durch eine mangelnde Gestaltung des analogen Materials ein Nachteil entsteht beziehungsweise tatsächlich eine andere Variable untersucht wird, wie zum Beispiel der Vergleich zwischen Arbeitsblättern mit oder ohne Lernhilfen. Weiterhin können Änderungen im Untersuchungsablauf oder andere studienbezogene Umsetzungen zu einer mangelnden Vergleichbarkeit zwischen der Interventions- und Vergleichsgruppe führen. Bei der vorliegenden Studie wurden jedoch lediglich die Arbeitsmaterialien variiert und der Stundenablauf beibehalten, sodass die Lerngruppen die gleichen Inhalte erarbeiteten. Die Erhebungsinstrumente beziehungsweise deren Fragen, die auf beide Studiengruppen zutrafen, wurden entsprechend umformuliert, indem der Begriff „Digitales Arbeitsblatt“ durch „Arbeitsblatt“ ersetzt wurde. Unpassende Fragen, wie zum Beispiel Items zur Usability, wurden in der Vergleichsgruppe folglich nicht erhoben.

#### **Arbeitsblätter**

Um die Wirksamkeit der Intervention auf die Motivation und die Nutzung der Lernhilfen testen zu können, wurde eine analog lernende Vergleichsgruppe geschaffen. Der Vergleichsgruppe wurden die gleichen Arbeitsblätter und Lernhilfen (Text und Bild) zur Verfügung gestellt, lediglich die auditiven und audiovisuellen Hilfen konnten nicht angeboten werden. Dazu wurden sämtliche Inhalte (Texte und Abbildungen) übernommen und gegebenenfalls an das analoge Format angepasst. Beispielsweise mussten die HyperDocs

auf ein DIN A4-Format formatiert werden (vgl. Anhang 9.1.4). Weiterhin musste die Einleitung des ersten Arbeitsblattes ebenfalls auf die analoge Methode umgeschrieben werden. Zusätzlich wurde bei den Arbeitsblättern eine dreidimensionale Umrandung eingeführt, um ein ansprechendes Layout zu schaffen, das mit dem digitalen Format vergleichbar ist. Es wurde außerdem auf eine gute Lesbarkeit der Inhalte geachtet. Die Arbeitsblätter wurden in Farbe doppelseitig beschriftet ausgedruckt. Bei mehr als einem Blatt wurden die Materialien zusammengeheftet.

Alternativ hätte die Vergleichsgruppe auch mit einfachen PDF-Arbeitsblättern unterrichtet werden können. Damit wäre eine digitale Bearbeitung des Arbeitsblattes möglich, allerdings stünden keine Hilfen zur Verfügung oder die Hilfen hätten in dem gleichen Dokument präsentiert werden müssen. Damit wären die Materialien sehr unübersichtlich oder umständlich in der Bedienung geworden, da beispielsweise die Hilfen nicht parallel zur Aufgabe betrachtet werden können. Beim Verzicht auf den Einsatz von Lernhilfen hätte die Studie zwei Variablen umfänglich manipuliert – die Art der Anwendung (HyperDocs / PDF-Reader) und den Grad der Differenzierung (keine Lernhilfen / Lernhilfen). Außerdem sollte, in Hinblick auf die Motivation und das Interesse, den Cognitive Load und die Nutzung der Lernhilfen, bewusst eine analoge mit einer digitalen Version verglichen werden, da überwiegend in den Schulen mit analogen Materialien unterrichtet wird.

### **Hilfen und Vertiefungsaufgabe**

Die größte Schwierigkeit beim Entwurf der Vergleichsgruppe lag in der Gestaltung des Differenzierungsangebots. Um eine vergleichbare Situation wie in der Untersuchungsgruppe zu schaffen, wurde jedem und jeder Schüler:in ein Satz Hilfen sowie die Vertiefungsaufgabe in jeder Stunde bereitgestellt. Damit konnten die Lernenden frei entscheiden, wann und wie sie eine Hilfe nutzen möchten, ohne dabei zum Beispiel aufstehen und an eine Sammelstelle für Hilfen laufen zu müssen. Die Hilfen wurden dabei in verbaler (Text) und visueller (Bild) Form bereitgestellt, wobei auch die Hilfen doppelseitig in der jeweilig eben beschriebenen Form bedruckt wurde (vgl. Abbildung 19).

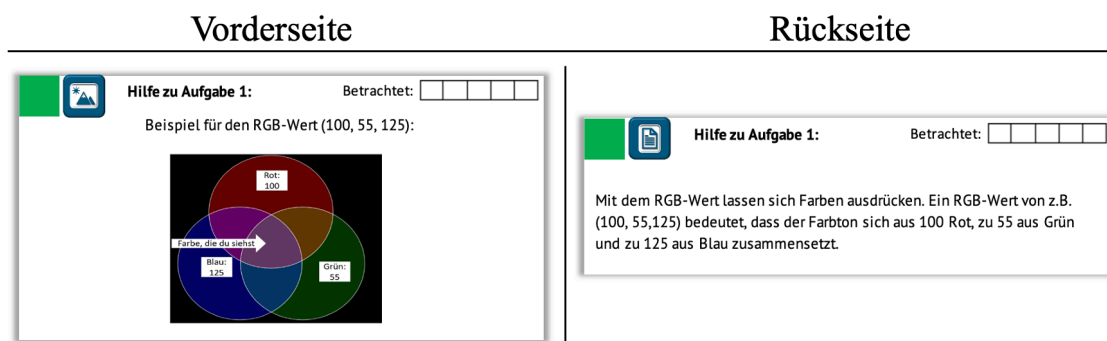


Abbildung 19: Vorder- und Rückseite einer Hilfe mit zwei verschiedenen Darstellungsformen (Links: Bild, rechts: Text).

Dazu wurden die Hilfen ebenfalls weitestgehend unverändert aus dem digitalen Format übernommen. Wenige Ausnahmen davon betrafen Hilfen, die aus mehreren Bildern zusammengesetzt waren, wie zum Beispiel die Versuchsdurchführung. Da für eine Hilfe maximal ein DIN A4-Blatt zur Verfügung stand, mussten die Bilder in ihrer Größe reduziert und neu angeordnet werden. Bei den HyperDocs erfolgte die Anordnung von oben nach unten. Bei der analogen Variante von links nach rechts und von oben nach unten. Audiovisuelle und auditive Hilfen standen aufgrund des analogen Formats nicht zur Verfügung. Zwei Hilfen aus der Interventionsgruppe wurden den Lernenden nicht zur Verfügung gestellt. Es handelt sich hierbei um kurze Erläuterungen zu den Begriffen „EFSA“ und „Azofarbstoffe“. Die Hilfen wurden in den HyperDocs direkt über die jeweiligen Wörter, ohne Symbole, verlinkt. Auf eine Implementation über analoge Lernhilfen wurde verzichtet, da die Kennzeichnung der Hilfen eine größere Veränderung des Layouts der Arbeitsblätter verursacht hätte. Die Informationen sind außerdem nicht relevant für die Aufgabenbearbeitung und wurden bei den Nutzungshäufigkeiten in der Interventionsgruppe nicht berücksichtigt (vgl. Kapitel 4.1.1).

Die Hilfen und Vertiefungsaufgaben wurden in einem DIN-LANG-Umschlag an die Arbeitsblätter angeheftet. Der Umschlag wurde außerdem mit einem freien Feld versehen, in das der oder die Schüler:in ihren jeweiligen anonymisierten Zuordnungscode des Arbeitsblattes schreiben konnten. Dadurch konnten lose Umschläge dem jeweiligen Arbeitsblatt zugeordnet werden. Hilfen, die nicht auf Anhieb in den Umschlag passten, wurden ein- oder zweimal gefaltet. Durch das Öffnen des Umschlags konnte festgestellt werden, ob der oder die Schüler:in die Hilfen betrachtete. Um eine genauere Analyse der Hilfenutzung auch im analogen Format zu gewährleisten, sollten die Lernenden ankreuzen, wie oft sie das jeweilige Differenzierungsangebot verwendet hatten (vgl. Abbildung 19). Leider zeigte sich während der Durchführung der Studie, dass nicht alle Schüler:innen

die jeweiligen Hilfen und Vertiefungsaufgaben markiert hatten, weshalb auf die Selbstauskunft auf dem Fragebogen zurückgegriffen wurde. Um die jeweilige Unterstützung in dem Umschlag schnell zu finden, wurden die Hilfen und die Vertiefung immer sortiert in den Umschlag gepackt. Dabei wurden die Hilfen lediglich randomisiert in Bezug auf ihre Darstellung (Text, Bild) einsortiert. Dieses Vorgehen wurde gewählt, um später eine bevorzugte Nutzung, hervorgerufen durch die vorgegebene Sortierung, zu vermeiden. Weiterhin wurde das Differenzierungsangebot farblich einzigartig markiert und jedes Kärtchen mit der Aufgabennummer versehen, um die Hilfen oder andere Informationen schnell zu finden (vgl. Anhang 9.1.6).

### **Externe Validität**

Externe Validität liegt vor, wenn sich die Ergebnisse auf andere Personen, Gruppen oder Situationen übertragen lassen (Schnell et al., 2018, S. 193). Die Ergebnisse sind damit verallgemeinerbar. An dieser Stelle werden nun verschiedene Störfaktoren der externen Validität erläutert (Schnell et al., 2018, S. 193 - 194; Rost, 2013, S. 129 - 130). Nehmen Versuchspersonen an einer Studie teil, sind sie darüber informiert, da unter anderem ihr Einverständnis erforderlich ist. Wissen sie oder vermuten sie zusätzlich den Grund oder Untersuchungsgegenstand der Studie, können die Teilnehmer:innen versuchen, die Erhebung durch gezielte Beantwortung der Fragen zu beeinflussen. Das oben genannte Phänomen wird dabei häufig durch künstliche Laborbedingungen verstärkt. Eine Verallgemeinerung auf Alltagssituationen wird erschwert. Die eigene Studie wurde explizit als Feldstudie angelegt, um den Effekt der Intervention beziehungsweise die Fragestellungen unter natürlichen Bedingungen zu testen. Dazu zählte zum Beispiel auch die Praktikabilität des Tableteinsatzes im Vergleich zu der analogen Variante. Schließlich werden die Geräte im Regelunterricht eingesetzt. Die Schüler:innen wurden weiterhin über die Erhebung aufgeklärt, jedoch wurde nicht speziell über den Zweck der Studie gesprochen. Die Interventions- und Vergleichsgruppen wussten nichts voneinander. Die Lehrkräfte erhielten keinen Einblick in die Befragungen der Schüler:innen, und die Stunden selbst wurden nicht von den Lehrkräften durchgeführt, sodass ein unnatürliches Verhalten, verursacht durch eine gewisse „Wunschhaltung“, der Schüler:innen eher auszuschließen war. Die Untersuchung fand außerdem im Rahmen des Mittelstufen- und Oberstufenunterrichts statt, womit auch ein breiterer Teil der Altersgruppen in den weiterführenden Schulen abgebildet wurde. Eine zu spezielle Auswahl der Stichprobe und deren spezifische Reaktion auf die experimentelle Situation kann zu einer Verzerrung der Ergebnisse führen. Es

handelt sich dann um eine Wechselwirkung zwischen den Auswahlverfahren und den experimentellen Variablen. Dadurch sind die Erkenntnisse nicht auf andere Stichproben übertragbar oder verallgemeinerbar. Durch eine Replikation des Experiments, kann der Einflussfaktor kontrolliert werden. Im Falle der Studie wurden verschiedene Schulen und Klassen interveniert. Dabei wurden die Stichproben, wie bereits weiter oben erläutert, quasi zufällig zugeordnet. Eine Vorauswahl durch den Versuchsleiter fand nicht statt. Die Klassen entsprachen regulären Schulklassen der jeweiligen Schulform.

### **4.5.5 Besonderheiten der Corona-Pandemie**

Die Corona-Pandemie hatte allgemein den Alltag erheblich verändert. Gerade das öffentliche Leben wurde aufgrund des Virus stark eingeschränkt. In den Schulen hat sich diese Einschränkungen durch die Schließungen, Wechselunterricht und bestimmte Verhaltensregeln im Schulalltag bemerkbar gemacht. Im nachfolgenden Abschnitt soll daher auf die verschiedenen Veränderungen im Alltag der Jugendlichen eingegangen und mögliche Konsequenzen für die vorliegende Studie genannt werden.

#### **Schulschließungen und Wechselunterricht**

Während die erste Pilotierung noch unter regulären Bedingungen stattfand, wurde die zweite Pilotierung im Wechselunterricht durchgeführt. Zwischen Mitte März und Mitte Mai 2020 waren in Deutschland die Schulen geschlossen und Schüler:innen wurden im Fernunterricht unterrichtet. Dadurch mussten viele Lernende aktiv ein digitales Medium zum Lernen nutzen. Gleichzeitig fehlten die sozialen Kontakte des Schulalltags. In der darauffolgenden Zeit erfolgte der Unterricht bis zu den Sommerferien im wöchentlichen Wechsel. Die Klassen und Kurse waren damit zum Teil auf zehn bis 15 Schüler:innen beschränkt. Die dritte Pilotierung wurde vor den Sommerferien 2020 durchgeführt. Es ist daher nicht auszuschließen, dass die pandemische Situation zu einem anderen Verhalten der Schüler:innen führte. Die eigentliche Interventionsstudie fand unmittelbar nach den Sommerferien zwischen September und Dezember 2020 statt. Dabei wurden die Klassen wieder regulär in der vollen Gruppengröße unterrichtet. Anschließend wurde erneut ein Lockdown verhängt, sodass die Schulen von Januar bis April 2021 erneut geschlossen wurden. Der anschließende Wechselunterricht dauerte bis Juni an. Ab Mitte Juni wurden die Klassen wieder regulär in voller Größe unterrichtet. Ab diesem Zeitpunkt fand die Untersuchung der Vergleichsgruppe statt, sodass unter den gleichen Bedingungen wie in der Interventionsgruppe unterrichtet wurde.

### **Unterrichtsgeschehen während der Pandemie**

Die eigentliche Durchführung des Unterrichts und die Zeitstruktur der Stunden wurden weder in der Interventions- noch in der Vergleichsgruppe beeinträchtigt. Die Veränderungen zu der Zeit vor der Pandemie waren regelmäßiges Lüften und das Tragen von Masken. Beide Maßnahmen hatten keine Auswirkungen auf die für die Studien vorgesehene Stundenstruktur. Aufgrund der bereits gemachten Erfahrungen mit der Pandemie während der letzten Pilotierung, dürften diese Maßnahmen zu keiner Veränderung im Verhalten der Schüler:innen in Bezug auf das Lernverhalten geführt haben. Weiterhin wurden die Tablets regelmäßig nach den Stunden desinfiziert. Auch hier bestand keinerlei Einfluss auf die eigentliche Stundenstruktur. Zum Zeitpunkt der Erhebung der Vergleichsgruppe bestand eine Testpflicht, sodass die Klassen zweimal in der Woche zu bestimmten Zeiten mit Schnelltests auf das Virus getestet wurden. Von den Testungen waren allerdings keine Erhebungspunkte betroffen.

## 5. Ergebnisse

In diesem Kapitel werden die Ergebnisse der Studie in Bezug auf die verschiedenen Forschungsfragen dargelegt. Eine Einordnung in den Forschungsstand und die Diskussion der Ergebnisse erfolgt in Kapitel 6. Zunächst wird die Stichprobe ausführlicher in Hinblick auf die Sozialdemografie und schulbezogene Merkmale beschrieben. Den Hauptteil des Kapitels bildet die Ergebnisdarstellung in Bezug auf die Forschungsfragen. Dazu erfolgt in jedem Unterkapitel die Darstellung der deskriptiven Statistik, gefolgt von den statistischen Ergebnissen der Inferenzanalysen. Als erstes werden die Ergebnisse in Bezug auf die Motivation und das Interesse dargelegt. Die Analysen erfolgen unter Beachtung der Vergleichs- und Interventionsgruppe sowie der Geschlechter und der Schulform. Die Zusammenhangsanalysen zu weiteren Variablen befinden sich in den anderen Teilkapiteln. Den zweiten Ergebnisteil bildet die Analyse der Usability. Hierbei wird die Benutzerfreundlichkeit zwischen dem ersten und vierten Einsatz verglichen. Zusätzlich erfolgt eine Analyse der Unterschiede zwischen den Geschlechtern und den Schulformen. Zuletzt erfolgen Zusammenhangsanalysen zu den tabletbezogenen Kompetenzen, dem Cognitive Load und der Motivation. Den letzten Analyseschwerpunkt bildet die Nutzung von Lernhilfen und Vertiefungsaufgaben. Dieses Kapitel weist im Gegensatz zu den vorherigen Kapiteln einen explorativen Charakter auf. Zunächst soll die Nutzung des Differenzierungsangebots zwischen der Interventions- und der Vergleichsgruppe betrachtet werden. Anschließend erfolgt eine Analyse der Nutzungshäufigkeiten über die Dauer der Erhebung nach Geschlecht und Schulform. Aufgrund des geringen Stichprobenumfangs der Gesamtschule erfolgen manche Analysen lediglich unter Einbezug der gymnasialen Schüler:innen, um eine bessere Vergleichbarkeit zwischen der Vergleichs- und Interventionsgruppe zu schaffen, da die Vergleichsgruppe lediglich in der Oberstufe, und auch dort nur zu einem geringen Teil, Datenpunkte aus der Gesamtschule beinhaltet. Aufgrund des geringen Stichprobenumfangs für diverse Lernende ( $n < 5$ ) erfolgen Analysen in Bezug auf das Geschlecht lediglich unter Beachtung der binären Geschlechter. In jedem Kapitel wird auf den genauen Umfang der Teilstichproben hingewiesen.

Viele der Analysen wurden aufgrund der Verbundenheit mehrerer Forschungsfragen als gemischte Varianzanalysen durchgeführt (vgl. Kapitel 4.1.3). Dazu wurde sich am Vorgehen nach Maxwell et al. (2017, S. 423) orientiert. Die Auswertung selbst erfolgte weitestgehend nach den Methoden von Wilcox (2021).



## 5.1 Beschreibung der Stichprobe

In diesem Kapitel erfolgt eine ausführlichere Beschreibung der Stichprobe unter Berücksichtigung der soziodemografischen und schulbezogenen Daten. Die Größe der Stichprobe richtet sich nach der größten, für die Analyse relevanten, Teilstichprobe, die in den meisten Auswertungen berücksichtigt wurde. Das umfasst bei den soziodemografischen Daten und den Fähigkeiten im jeweiligen Fachgebiet die Teilstichprobe für die Motivation beziehungsweise den Cognitive Load und bei den Nutzungshäufigkeiten der digitalen Endgeräte die Teilstichprobe der Usability. Die Stichprobengrößen sind angegeben.

### 5.1.1 Interventionsgruppe

#### Mittelstufe

In Tabelle 23 sind die Verteilung und das Alter der Teilnehmer:innen gezeigt. Die Gesamtschüler:innen sind im Schnitt fast ein halbes Jahr älter als die Gymnasiast:innen, und es überwiegt das männliche Geschlecht (vgl. Tabelle 23).

*Tabelle 23: Geschlecht und Alter (in Jahren) der Proband:innen in der Mittelstufe der Interventionsgruppe, aufgeteilt nach Schulform.*

	<i>männlich</i>	<i>weiblich</i>	<i>divers</i>	<i>Alter (Mittelwert)</i>
Gymnasium	99	109	0	13 – 16 (14.51)
Gesamtschule	68	53	0	13 – 17 (14.98)

Bei der Einschätzung der Fähigkeiten im Fach Chemie liegen die Schüler:innen der Gesamtschule ebenfalls über denen des Gymnasiums, während die anderen Werte nahezu gleich sind (vgl. Tabelle 24).

*Tabelle 24: Mittelwerte mit Standardabweichung (in Klammern) Einschätzungen der Fähigkeiten im Fach Chemie, den Naturwissenschaften (Nawi) und in der Schule allgemein in der Mittelstufe der Interventionsgruppe (7-stufige Skala vom Typ Likert mit den Polen „sehr schlecht“ (1) bis „sehr gut“ (7)).*

	<i>Fähigkeiten Chemie</i>		<i>Fähigkeiten Nawi</i>		<i>Fähigkeiten Schule</i>	
	w	m	w	m	w	m
Gymnasium	4.04 (1.32)	4.42 (1.25)	4.69 (1.16)	5.07 (0.96)	5.09 (1.05)	5.07 (0.96)
Gesamtschule	4.49 (1.23)	4.69 (1.37)	4.45 (1.35)	5.04 (1.10)	5.09 (1.04)	5.06 (0.94)

Es fällt auf, dass die Schüler:innen beider Schulformen ihre Fähigkeiten in Chemie und in den Naturwissenschaften als schlechter gegenüber den allgemeinen Fähigkeiten in der Schule einschätzten. Zwischen den Geschlechtern liegen in Bezug auf die Fähigkeiten in Chemie und in den Naturwissenschaften Unterschiede zugunsten der Schüler vor.

Im Gymnasium beträgt der Mittelwert des Nutzungsangebots von Lernhilfen im Unterricht auf einer 7-stufigen Skala vom Typ Likert mit den Polen „nie“ bis „immer“ bei 97 Schülern  $M = 1.62$  (1.21) und bei 107 Schülerinnen  $M = 1.61$  (1.08). In der Gesamtschule liegt die Angebotshäufigkeit bei 66 Schülern bei  $M = 1.61$  (0.91) und bei 52 Schülerinnen bei  $M = 1.85$  (1.48). Fehlende Werte wurden bei beiden Schulformen nicht imputiert. Insgesamt machten sieben Schüler:innen keine Angaben. In der Gesamtschule wurde demnach ähnlich häufig mit Lernhilfen differenziert wie am Gymnasium. Insgesamt ist das Angebot allerdings gering.

#### *Nutzung digitaler Endgeräte*

Die Ergebnisse umfassen lediglich Datenpunkte von Schüler:innen, die zuhause oder in der Schule die jeweiligen Geräte nutzen konnten (vgl. Anhang 9.3.1). In der Mittelstufe nutzten die Lernenden des Gymnasiums zuhause häufig ein Smartphone oder den Computer. Tablets wurden nur mittelhäufig verwendet. Die Schülerinnen nutzten dabei häufiger das Smartphone oder das Tablet als die Schüler (vgl. Tabelle 25). In der Gesamtschule wurden häufiger als am Gymnasium das Smartphone und das Tablet zum häuslichen Lernen herangezogen. Die Geschlechtsunterschiede zeigten sich auch dort, wonach vor allem Schülerinnen ein Tablet anstelle des PCs verwenden.

*Tabelle 25: Mittelwerte mit Standardabweichung (in Klammern) der häuslichen Nutzung digitaler Endgeräte in der Mittelstufe. 7-stufige Skala vom Typ Likert mit den Polen „nie“ (1) bis „immer“ (7). Es wurden lediglich Daten von Schüler:innen miteinbezogen, wenn sie das Gerät auch nutzen konnten.*

	<i>Nutzung Smartphone</i>		<i>Nutzung Tablet</i>		<i>Nutzung Computer</i>	
	w	m	w	m	w	m
Gymnasium	4.52 (1.47)	4.17 (1.73)	3.12 (1.90)	2.75 (1.71)	4.06 (1.66)	4.04 (1.72)
Gesamtschule	5.27 (1.25)	4.42 (1.42)	3.86 (2.01)	2.57 (1.78)	3.78 (1.93)	4.06 (1.64)

In der Schule nutzten die wenigsten Lernenden ein Smartphone oder ein Tablet zum Lernen. Der Computer wurde am häufigsten verwendet. Es sind bei der Nutzung des Tablets und des Computers leichte Unterschiede zwischen den Geschlechtern vorhanden (vgl.

Tabelle 26). Die Schüler:innen der Gesamtschule nutzten insgesamt deutlich häufiger das Smartphone oder Tablet beim Lernen in der Schule. Die Computernutzung war ähnlich hoch, wobei dort die Geschlechterunterschiede größer sind. Tablets wurden jedoch von Schülerinnen häufiger genutzt. Die Nutzung ist insgesamt jedoch niedrig.

*Tabelle 26: Mittelwerte mit Standardabweichung (in Klammern) der schulischen Nutzung digitaler Endgeräte in der Mittelstufe. 7-stufige Skala vom Typ Likert mit den Polen „nie“ (1) bis „immer“ (7). Es wurden lediglich Daten von Schüler:innen miteinbezogen, wenn sie das Gerät auch nutzen konnten.*

	<i>Nutzung Smartphone</i>		<i>Nutzung Tablet</i>		<i>Nutzung Computer</i>	
	w	m	w	m	w	m
Gymnasium	1.16 (0.82)	1.07 (0.25)	1.92 (0.76)	2.36 (1.43)	2.67 (1.02)	2.74 (1.43)
Gesamtschule	2.81 (1.02)	2.70 (1.37)	3.11 (1.37)	2.63 (1.22)	2.35 (1.23)	2.88 (1.48)

### **Oberstufe**

In der Oberstufe war die Geschlechterverteilung im Gymnasium ausgeglichen. Die Stichprobe der Gesamtschule umfasste lediglich eine Klasse, in der sich ausschließlich Schülerinnen befanden. Durch die geringe Stichprobe variierte auch das Alter stärker (vgl. Tabelle 27).

*Tabelle 27: Geschlecht und Alter (in Jahren) der Proband:innen in der Oberstufe der Interventionsgruppe, unterteilt nach Schulform.*

	<i>männlich</i>	<i>weiblich</i>	<i>divers</i>	<i>Alter (Durchschnitt)</i>
Gymnasium	56	46	0	16 – 19 (16.93)
Gesamtschule	0	8	0	16 – 18 (16.38)

Bei den Fähigkeiten lagen erneut Unterschiede zwischen den Geschlechtern bei den Gymnasiast:innen vor. Die Schülerinnen schätzten demnach ihre Fähigkeiten im Fach Chemie als deutlich niedriger ein. Die sonstigen Angaben sind ausgeglichen. Der Vergleich zur Gesamtschule zeigt deutlich niedrigere Fähigkeiten bei den Gesamtschülerinnen (vgl. Tabelle 28). Die Stichprobe ist jedoch klein und basiert, im Gegensatz zu der des Gymnasiums, lediglich auf einem Grundkurs.

*Tabelle 28: Mittelwerte mit Standardabweichung (in Klammern) der Einschätzungen der Fähigkeiten im Fach Chemie, den Naturwissenschaften (Nawi) und in der Schule allgemein in der Oberstufe der Interventionsgruppe. (7-stufige Skala vom Typ Likert mit den Polen „sehr schlecht“ (1) bis „sehr gut“ (7)).*

	<i>Fähigkeiten Chemie</i>		<i>Fähigkeiten Nawi</i>		<i>Fähigkeiten Schule</i>	
	w	m	w	m	w	m
Gymnasium	4.52 (1.38)	5.11 (1.30)	5.24 (0.87)	5.29 (1.06)	5.24 (0.82)	5.16 (1.13)
Gesamtschule	3.10 (1.00)	/	4.38 (0.74)	/	4.50 (1.07)	/

56 Schüler am Gymnasium gaben auf einer 7-stufigen Skala vom Typ Likert mit den Polen „nie“ bis „immer“ einen Mittelwert von  $M = 1.66$  (1.23) bezüglich des Nutzungsangebots von Hilfen an. Bei den 45 Schülerinnen liegt der Mittelwert bei  $M = 1.84$  (1.36). Eine Schülerin machte keine Angabe. In der Gesamtschule gaben die acht Schülerinnen einen Mittelwert von  $M = 3.13$  (2.23) an. Sie hatten damit häufiger im Unterricht die Möglichkeit, Hilfen zu nutzen. Die Standardabweichung ist allerdings hoch.

#### *Nutzung digitaler Endgeräte*

Die Ergebnisse umfassen erneut lediglich Datenpunkte von Schüler:innen, die zuhause oder in der Schule die jeweiligen Geräte nutzen konnten (vgl. Anhang 9.3.1). Auch in der Oberstufe nutzten Schülerinnen am Gymnasium gegenüber Schülern häufiger das Smartphone und das Tablet. Die männlichen Lernenden nutzten eher den Computer zum Lernen. Die Stichprobe der Gesamtschule besteht lediglich aus Schülerinnen. Sie nutzten insgesamt selten digitale Medien zum Lernen (vgl. Tabelle 29).

*Tabelle 29: Mittelwerte mit Standardabweichung (in Klammern) der häuslichen Nutzung digitaler Endgeräte in der Oberstufe. 7-stufige Skala vom Typ Likert mit den Polen „nie“ (1) bis „immer“ (7). Es wurden lediglich Daten von Schüler:innen miteinbezogen, wenn sie das Gerät auch nutzen konnten.*

	<i>Nutzung Smartphone</i>		<i>Nutzung Tablet</i>		<i>Nutzung Computer</i>	
	w	m	w	m	w	m
Gymnasium	4.74 (1.62)	3.96 (1.51)	3.64 (1.85)	2.37 (1.82)	3.95 (1.75)	4.49 (1.65)
Gesamtschule	1.25 (0.46)	/	2.00 (0.00)	/	1.13 (0.36)	/

In der Schule nutzten die Gymnasiast:innen am häufigsten den Computer. Schülerinnen nutzten ebenfalls häufig das Tablet. Das Smartphone wurde in der Schule kaum genutzt

(vgl. Tabelle 30). In der Gesamtschule sind die Daten aufgrund der kleinen Stichprobe kaum aussagekräftig. Die Daten für die Nutzung des Smartphones und des Computers basieren auf zwei beziehungsweise einem Datenpunkt.

*Tabelle 30: Mittelwerte mit Standardabweichung (in Klammern) der schulischen Nutzung digitaler Endgeräte in der Oberstufe. 7-stufige Skala vom Typ Likert mit den Polen „nie“ (1) bis „immer“ (7). Es wurden lediglich Daten von Schüler:innen miteinbezogen, wenn sie das Gerät auch nutzen konnten.*

	<i>Nutzung Smartphone</i>		<i>Nutzung Tablet</i>		<i>Nutzung Computer</i>	
	w	m	w	m	w	m
Gymnasium	1.10 (0.30)	1.50 (1.54)	3.40 (1.14)	1.75 (0.71)	3.14 (1.08)	3.38 (1.13)
Gesamtschule	6.50 (0.71)	/	3.13 (1.96)	/	7.00 (0.00)	/

### 5.1.2 Vergleichsgruppe

In der Vergleichsgruppe erfolgt die Darstellung nach der vorherigen Gliederung. Dabei entfallen allerdings die Angaben zur Nutzungshäufigkeit von digitalen Endgeräten, da diese Variablen aufgrund der fehlenden Relevanz bei dieser Gruppe nicht erhoben wurden.

#### Mittelstufe

In der Mittelstufe befanden sich insgesamt 151 Schüler:innen in der Erhebung (vgl. Tabelle 31). Die folgenden Beschreibungen finden aufgrund der geringeren Stichprobengröße der diversen Lernenden lediglich unter Beachtung des weiblichen und männlichen Geschlechts statt. Die Geschlechter sind in etwa gleich verteilt. Das Alter ist höher als in der Interventionsgruppe.

*Tabelle 31: Geschlecht und Alter (in Jahren) der Proband:innen in der Mittelstufe der Vergleichsgruppe.*

	<b>männlich</b>	<b>weiblich</b>	<b>divers</b>	<b>Alter (Durchschnitt)</b>
Gymnasium	78	71	2	13 – 17 (15.24)

Wie bereits in der Interventionsgruppe schätzten die Schülerinnen ihre Fähigkeiten in Chemie und in den Naturwissenschaften geringer ein als die Schüler. Insgesamt ist die Selbsteinschätzung für Chemie niedriger als in der Interventionsgruppe (vgl. Tabelle 32).

*Tabelle 32: Mittelwerte mit Standardabweichung (in Klammern) der Einschätzungen der Fähigkeiten in Chemie, den Naturwissenschaften (Nawi) und in der Schule allgemein in der Mittelstufe der Vergleichsgruppe (7-stufige Skala vom Typ Likert mit den Polen „sehr schlecht“ (1) bis „sehr gut“ (7)).*

	<i>Fähigkeiten Chemie</i>		<i>Fähigkeiten Nawi</i>		<i>Fähigkeiten Schule</i>	
	w	m	w	m	w	m
Gymnasium	3.82 (1.54)	4.31 (1.61)	4.61 (1.25)	5.08 (1.28)	5.17 (1.12)	5.08 (1.07)

Auf einer 7-stufigen Skala vom Typ Likert mit den Polen „nie“ bis „immer“ liegt der Mittelwert von 71 Schülerinnen bei  $M = 2.79$  (1.65) und von 78 Schülern bei  $M = 2.49$  (1.73). Die Vergleichsgruppe konnte damit etwas häufiger im Unterricht auf Lernhilfen zurückgreifen als die Interventionsgruppe.

### **Oberstufe**

In der Oberstufe ist ein kleiner Kurs der Gesamtschule vorhanden, der aufgrund von Dropouts weiter dezimiert wurde (vgl. Tabelle 33). Aufgrund der niedrigeren Teilstichprobengröße der diversen Lernenden und dem damit verbundenen Ausschluss aus den überwiegenden Auswertungsverfahren, erfolgt keine gesonderte Darstellung der Fähigkeiten im jeweiligen Fach.

*Tabelle 33: Geschlecht und Alter (in Jahren) der Proband:innen in der Oberstufe der Vergleichsgruppe, unterteilt nach Schulform.*

	<i>männlich</i>	<i>weiblich</i>	<i>divers</i>	<i>Alter (Durchschnitt)</i>
Gymnasium	24	24	2	15 – 18 (16.50)
Gesamtschule	1	2	0	17 (17.00)

In der Vergleichsgruppe schätzten die Schülerinnen ihre Fähigkeiten im Fach Chemie erneut geringer ein als die Schüler. In den Naturwissenschaften liegen sie jedoch dieses Mal über den männlichen Lernenden, ebenso bei den Fähigkeiten in der Schule (vgl. Tabelle 34). Insgesamt schätzten die Vergleichs- und Interventionsgruppe ihre Fähigkeiten in etwa gleich hoch ein.

Tabelle 34: Mittelwerte mit Standardabweichung (in Klammern) der Einschätzungen der Fähigkeiten in Chemie, den Naturwissenschaften (Nawi) und in der Schule allgemein in der Oberstufe der Vergleichsgruppe (7-stufige Skala vom Typ Likert mit den Polen „sehr schlecht“ (1) bis „sehr gut“ (7)).

	<i>Fähigkeiten Chemie</i>		<i>Fähigkeiten Nawi</i>		<i>Fähigkeiten Schule</i>	
	w	m	w	m	w	m
Gymnasium	4.63 (1.50)	5.00 (1.47)	5.00 (1.47)	4.79 (1.32)	5.33 (1.09)	5.17 (0.92)

Von den 23 Schülerinnen liegt der Mittelwert der Möglichkeit der Hilfenutzung bei  $M = 2.48$  (1.28) und bei den 24 Schülern bei  $M = 2.96$  (1.27). Eine Schülerin machte keine Angaben.

## 5.2 Intrinsische Motivation und Interesse

Die Auswertung der intrinsischen Motivation erfolgt unter Betrachtung der Variablen des Intrinsic Motivation Inventory in Bezug auf das Lernen mit Arbeitsblättern (Interesse / Vergnügen, wahrgenommene Wahlfreiheit (Autonomie), Kompetenzerleben und Druck). Die Ergebnisse des Fachinteresses Chemie sowie des Sachinteresses Naturwissenschaften werden ebenfalls an dieser Stelle dargestellt. Die Variablen wurden durch robuste gemischte Varianzanalysen (*engl.: mixed ANOVA*) analysiert. Die Voraussetzungen der mixed ANOVA wurden bei jeder Variable geprüft. Die Inner- und Zwischensubjektfaktoren sind den jeweiligen Kapiteln zu entnehmen. Bei Vorhandensein von Interaktionen wurden die Haupteffekte nicht weiter analysiert (Maxwell et al., 2017, S. 423).

### 5.2.1 Vergleich zwischen Interventions- und Vergleichsgruppe in Abhängigkeit vom Geschlecht und Messzeitpunkt in der Mittelstufe

#### Interesse: Geschlecht, Gruppe, Messzeitpunkt

##### *Interesse beim Lernen mit Arbeitsblättern*

Um den Einfluss der Intervention auf das Interesse nach der SDT beim Lernen mit HyperDocs in Abhängigkeit vom Geschlecht zu untersuchen, wurde eine gemischte Varianzanalyse (between-by-between-by-within) mit den Faktoren „Geschlecht“, „Gruppe“ und „Messzeitpunkt“ durchgeführt. Damit wurden die Haupteffekte der drei Faktoren sowie deren Interaktionen geprüft. Als erstes wurde die deskriptive Statistik gewonnen (vgl. Tabelle 35).

Tabelle 35: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf das Interesse (SDT) beim Lernen mit Arbeitsblättern in der Mittelstufe.

	<i>Interventionsgruppe</i>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	109				99			
Mittelwert	4.61	4.41	3.98	3.97	4.45	4.26	4.11	4.03
(Standardabweichung)	(0.52)	(0.64)	(0.86)	(0.91)	(0.78)	(0.81)	(0.75)	(0.86)
Median	4.75	4.50	4.25	4.00	4.75	4.5	4.25	4.25
	<i>Vergleichsgruppe</i>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	71				78			
Mittelwert	2.83	3.11	2.99	2.86	3.39	3.46	3.31	3.18
(Standardabweichung)	(1.04)	(1.00)	(1.03)	(1.06)	(0.91)	(0.98)	(0.92)	(0.92)
Median	2.75	3.25	3.00	3.00	3.50	3.62	3.25	3.00

Anschließend wurden die Voraussetzungen für eine gemischte Varianzanalyse geprüft. Mauchly's Test auf Sphärizität ist für den Innersubjektfaktor „Messzeitpunkt“ signifikant ( $p < 0.001$ ). Die Verteilung der Residuen zeigt auf Basis eines QQ-Plots und eines signifikanten Shapiro-Wilk-Tests ( $p < 0.001$ ) keine Normalverteilung. In der Stichprobe sind, gruppiert nach Messzeitpunkt, Geschlecht und Gruppenzugehörigkeit, 28 Ausreißer ( $1,5 \cdot \text{IQR}^9$ ), wovon zwei als extrem gelten ( $3 \cdot \text{IQR}$ ), vorhanden. Der Levenes-Test auf Varianzhomogenität zeigt für die ersten beiden Messzeitpunkte ein signifikantes Ergebnis (1:  $p < 0.001$ ; 2:  $p = 0.001$ ; 3:  $p = 0.088$ ; 4:  $p = 0.218$ ). Der Box-M-Test für die Homogenität der Kovarianzmatrizen ist sowohl für den Faktor „Geschlecht“ ( $p = 0.001$ ) als auch den Faktor „Gruppe“ ( $p < 0.001$ ) signifikant. Aufgrund der Verletzung der Voraussetzungen wurde eine robuste Variante der gemischten Varianzanalyse mit Bootstrapping durchgeführt. Tabelle 36 zeigt deren Ergebnisse.

<sup>9</sup> Interquartilsabstand (eng.: interquartile range)



Tabelle 36: Ergebnisse der robusten Varianzanalyse zum Interesse (SDT) beim Lernen mit Arbeitsblättern in der Mittelstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap-Verfahren.<sup>10</sup>

Effekt	$Q$	$p$
Geschlecht	6.24**	0.007
Gruppe	201.07**	< 0.001
Messzeitpunkt	13.66**	< 0.001
Geschlecht*Gruppe	6.39*	0.013
Geschlecht*Messzeitpunkt	0.36	0.702
Gruppe*Messzeitpunkt	9.44**	< 0.001
Geschlecht*Gruppe*Messzeitpunkt	1.39	0.093

Die Analyse zeigt eine nicht signifikante Dreiwege-Interaktion. In Abhängigkeit von der Zeit verlaufen die jeweiligen Graphen, bis auf die erste Stunde, bei Betrachtung des Geschlechts und der Gruppe nahezu parallel (vgl. Abbildung 20).

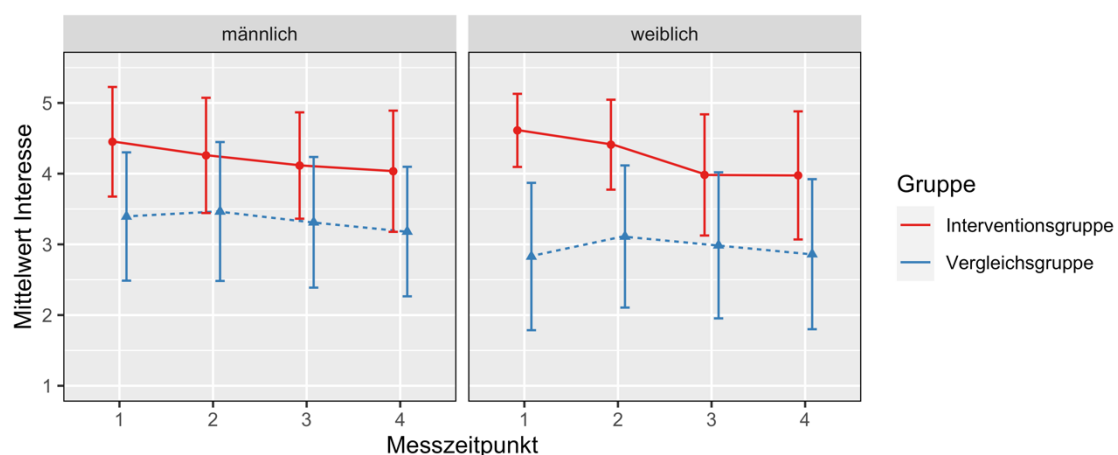


Abbildung 20: Mittelwerte mit Standardabweichung des Interesses (SDT) beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.

Zur weiteren Analyse wurden die beiden signifikanten Zweifache-Interaktionen betrachtet, in denen der dritte Faktor unberücksichtigt blieb. Aus Abbildung 20 wird ersichtlich, dass sich die Interventionsgruppe und die Vergleichsgruppe in Abhängigkeit von der Zeit unterscheiden. Während bei ersterer eine nahezu kontinuierliche Abnahme des Interesses zu beobachten ist, steigt bei letzterer das Interesse zunächst von der ersten zur zweiten Stunde, unabhängig von dem Geschlecht, an. Zusätzlich zeigt sich bei den Schülerinnen

<sup>10</sup> Im Folgenden werden signifikante Ergebnisse auf dem Niveau  $p < 0.05$  mit einem Stern (\*) und signifikante Ergebnisse auf dem Niveau  $p < 0.01$  mit zwei Sternen (\*\*) markiert.

der Interventionsgruppe eine starke Abnahme des Interesses von der zweiten zu der dritten Stunde. Die Schülerinnen der Vergleichsgruppe zeigen über die gesamte Erhebung ein geringeres Interesse als die Schüler.

Als nächstes wurde die Zweiwege-Interaktion „Geschlecht\*Gruppe“ näher betrachtet und die einfachen Haupteffekte geprüft. Dazu wurde das Interesse über alle vier Messzeitpunkte gemittelt. Die Ergebnisse sind in Tabelle 37 zusammengefasst.

*Tabelle 37: Ergebnisse der Untersuchung der einfachen Haupteffekte der Interaktion „Geschlecht\*Gruppe“ in der Mittelstufe.*

<i>Gruppe</i>	<i>t</i>	<i>df</i>	<i>p</i>	$\bar{\xi}$
Interventionsgruppe	0.06	125.90	0.961	0.02
Vergleichsgruppe	3.46**	75.46	< 0.001	0.40
<b><i>Geschlecht</i></b>				
weiblich	11.36**	70.08	< 0.001	0.86
männlich	10.03**	98.17	< 0.001	0.78

Es zeigt sich ein signifikanter Unterschied zwischen den Geschlechtern in der Vergleichsgruppe ( $M_w = 2.95 \pm 0.89$ ,  $M_m = 3.34 \pm 0.74$ )<sup>11</sup>, nicht jedoch in der Interventionsgruppe ( $M_w = 4.24 \pm 0.61$ ,  $M_m = 4.22 \pm 0.65$ ). Bei den Schülerinnen ( $M_I = 4.24 \pm 0.61$ ,  $M_V = 2.95 \pm 0.89$ )<sup>12</sup> und Schülern ( $M_I = 4.22 \pm 0.65$ ,  $M_V = 3.34 \pm 0.74$ ) gibt es signifikante Unterschiede zwischen den Gruppen mit einer hohen Effektstärke.

Zum Schluss wurde die Zweiwege-Interaktion „Gruppe\*Messzeitpunkt“ analysiert. Dazu erfolgte eine Auswertung der einfachen Haupteffekte. Für den Zwischensubjektfaktor ergeben sich die Ergebnisse nach Tabelle 38. Dafür wurden erneut unabhängige robuste t-Tests durchgeführt. Für jeden Messzeitpunkt unterscheiden sich die Gruppen signifikant voneinander, wobei das Interesse der Interventionsgruppe mit einer hohen Effektstärke über dem der Vergleichsgruppe liegt. Bei dem Vergleich der Mittelwerte wird innerhalb der Gruppen bereits eine tendenzielle Abnahme des Interesses beim Lernen mit HyperDocs sichtbar, die zunächst bei der Interventionsgruppe stärker erfolgt. Dennoch bleiben die Unterschiede mit einer hohen Effektstärke bestehen.

<sup>11</sup> Das tiefgestellte w steht für weiblich, m für männlich.

<sup>12</sup> Das tiefgestellte I steht für Interventionsgruppe, V für Vergleichsgruppe.

Tabelle 38: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Gruppe“ der Interaktion „Gruppe\*Messzeitpunkt“ (Interesse) in der Mittelstufe.

Messzeitpunkt	<i>t</i>	<i>df</i>	<i>p</i>	$\hat{\xi}$	<i>M<sub>I</sub></i> ( <i>SD</i> )	<i>M<sub>V</sub></i> ( <i>SD</i> )
1	15.53**	137.78	< 0.001	0.81	4.54 (0.66)	3.12 (1.01)
2	10.57**	163.27	< 0.001	0.72	4.34 (0.73)	3.30 (1.00)
3	8.55**	201.28	< 0.001	0.62	4.04 (0.81)	3.16 (0.99)
4	9.60**	180.77	< 0.001	0.68	4.00 (0.88)	3.03 (1.00)

Für die Interventionsgruppe ergibt sich in Bezug auf den Innersubjektfaktor „Messzeitpunkt“ ein signifikantes Ergebnis ( $robust F(2.71, 338.85) = 49.19, p < 0.001$ ), ebenso für die Vergleichsgruppe ( $robust F(2.94, 264.60) = 5.47, p < 0.001$ ). Die Post-hoc-Tests zeigen in der Interventionsgruppe lediglich zwischen der dritten und vierten Stunde keinen signifikanten Unterschied, während bei der Vergleichsgruppe ein signifikanter Unterschied lediglich zwischen der zweiten und vierten Erhebung zu finden ist. Die genauen Testergebnisse sind Anhang 9.3.2 zu entnehmen. Die Effektstärke in der Interventionsgruppe beträgt  $\delta_t = 0.39$  und ist damit als schwach einzustufen. Die Effektstärke zwischen der ersten und der vierten Stunde beträgt  $\delta_t = 0.64$  und ist damit mittelstark. In der Vergleichsgruppe beträgt die mittlere Effektstärke  $\delta_t = 0.08$  und diejenige, für den Vergleich zwischen der ersten und vierten Stunde ebenfalls  $\delta_t = 0.08$ . In der Interventionsgruppe nimmt damit, im Gegensatz zu der Vergleichsgruppe, das Interesse beim Lernen mit Arbeitsblättern deutlich ab.

#### *Fachinteresse Chemie und Sachinteresse Naturwissenschaften*

Neben dem Interesse beim Lernen mit HyperDocs wurde auch das Fachinteresse Chemie und das Sachinteresse Naturwissenschaften im Pre-/Post-Test bestimmt. Die Ergebnisse wurden im Folgenden nach den gleichen Verfahren wie oben bestimmt. Aufgrund der Verletzung der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0,001$ ), der Varianzhomogenität zu beiden Messzeitpunkten (*Pre*:  $p = 0.008$ , *Post*:  $p = 0.018$ ) und der Homogenität der Kovarianzmatrizen für beide Faktoren wurde eine robuste mixed ANOVA durchgeführt. In der Stichprobe ist ein Ausreißer vorhanden. In Tabelle 39 ist die deskriptive Statistik aufgeführt.

Tabelle 39: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf Fachinteresse Chemie in der Mittelstufe.

	<b>Interventionsgruppe</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt	1	2	1	2
Stichprobenumfang	101		94	
Mittelwert (Standardabweichung)	3.16 (1.01)	3.21 (1.15)	3.43 (0.84)	3.60 (0.86)
Median	3.00	3.00	3.50	3.50
	<b>Vergleichsgruppe</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt	1	2	1	2
Stichprobenumfang	69		76	
Mittelwert (Standardabweichung)	2.89 (1.15)	2.92 (1.26)	3.41 (1.17)	3.45 (1.17)
Median	3.00	3.00	3.50	3.50

Die nachfolgende Tabelle (vgl. Tabelle 40) zeigt die Ergebnisse der gemischten Varianzanalyse. Demnach existiert lediglich ein signifikanter Haupteffekt „Geschlecht“. Die Zweifache-Interaktion „Gruppe\*Messzeitpunkt“ liegt genau auf der Signifikanzgrenze.

Tabelle 40: Ergebnisse der robusten Varianzanalyse zum Fachinteresse Chemie in der Mittelstufe. Die *Q*-Werte stammen aus dem non-bootstrap-Verfahren.

<b>Effekt</b>	<b>Q</b>	<b>p</b>
Geschlecht	11.44**	< 0.001
Gruppe	0.66	0.381
Messzeitpunkt	1.31	0.229
Geschlecht*Gruppe	1.39	0.234
Geschlecht*Messzeitpunkt	0.68	0.381
Gruppe*Messzeitpunkt	3.84	0.050
Geschlecht*Gruppe*Messzeitpunkt	0.13	0.698

In Abbildung 21 zeigt sich der Unterschied zwischen den Geschlechtern.

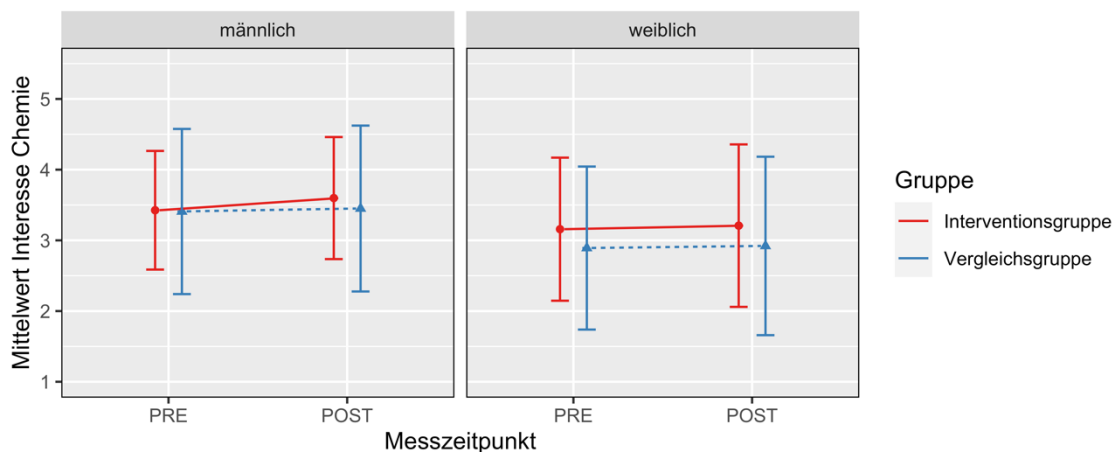


Abbildung 21: Mittelwerte mit Standardabweichung des Fachinteresses Chemie, aufgeteilt nach Geschlecht und Gruppe im Pre- und Posttest in der Mittelstufe.

Demnach besteht bei den Schülerinnen ein geringeres Interesse als bei den Schülern. Eine leichte Zunahme des Interesses ist bei allen Teilgruppen zu beobachten. Die Interaktion „Gruppe\*Messzeitpunkt“ liegt genau auf dem Signifikanzniveau von  $\alpha = 5\%$ . Auch wenn die Interaktion „Gruppe\*Messzeitpunkt“, unter der Annahme  $p < \alpha$ , nicht mehr signifikant ist, wurden die einfachen Haupteffekte analysiert. Der einfache Haupteffekt des Innersubjektfaktors „Messzeitpunkt“ zeigt für die Interventionsgruppe ( $M_{Pre} = 3.29 \pm 0.94$ ,  $M_{Post} = 3.40 \pm 1.04$ ) einen signifikanten Effekt mit schwacher Effektstärke (*robust t*(116) = 2.78,  $p = 0.006$ ,  $\hat{\xi} = 0.14$ ). Die Vergleichsgruppe ( $M_{Pre} = 3.16 \pm 1.19$ ,  $M_{Post} = 3.20 \pm 1.24$ ) zeigt keinen signifikanten Effekt (*robust t*(86) = 0.165,  $p = 0.869$ ,  $\hat{\xi} = 0.01$ )<sup>13</sup>. Der einfache Haupteffekt „Gruppe“ ist zum ersten Messzeitpunkt (*robust t*(150.66) = 0.06,  $p = 0.954$ ,  $\hat{\xi} = 0.03$ ;  $M_I = 3.29 \pm 0.94$ ,  $M_V = 3.16 \pm 1.19$ ) und zweiten Messzeitpunkt (*robust t*(159.17) = 1.27,  $p = 0.204$ ,  $\hat{\xi} = 0.11$ ;  $M_I = 3.40 \pm 1.04$ ,  $M_V = 3.20 \pm 1.24$ ) nicht signifikant, wobei die Interventionsgruppe einen größeren Interessenszuwachs verzeichnet als die Vergleichsgruppe, was durch den einfachen Haupteffekt „Messzeitpunkt“ bestätigt wird. Eine weitere Analyse des Haupteffektes „Geschlecht“ ( $M_w = 3.07 \pm 1.08$ ,  $M_m = 3.48 \pm 0.94$ ) zeigt einen signifikanten Unterschied zwischen den Geschlechtern mit einer schwachen bis mittleren Effektstärke (*robust t*(188.80) = 3.38,  $p < 0.001$ ,  $\hat{\xi} = 0.26$ ). Damit zeigen die Schülerinnen über die gesamte Erhebung ein niedrigeres Fachinteresse Chemie als die Schüler.

<sup>13</sup> Die t-Test-Statistik wird im Folgenden immer als positiver Wert dargelegt. Die Richtung des Effekts kann der Veränderung der Mittelwerte entnommen werden.

Neben dem Interesse am Fach Chemie wurde auch das Sachinteresse an den Naturwissenschaften bestimmt, da die Unterrichtsreihe verschiedene Aspekte der Fächer Mathematik und Biologie inkludiert, die wiederum jedoch nicht den Fokus des Fachinhaltes bilden. Die Werte der deskriptiven Statistik sind Tabelle 41 zu entnehmen.

*Tabelle 41: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf das Sachinteresse an Naturwissenschaften in der Mittelstufe.*

	<b>Interventionsgruppe</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt				
Stichprobenumfang	101		94	
Mittelwert	3.13 (1.03)	3.16 (1.15)	3.31 (0.94)	3.44 (0.96)
Median	3.50	3.50	3.00	3.00
	<b>Vergleichsgruppe</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt				
Stichprobenumfang	69		76	
Mittelwert	2.83 (1.01)	2.87 (1.18)	3.41 (1.02)	3.57 (1.07)
Median	3.00	3.00	3.50	3.50

Die robuste gemischte Varianzanalyse zeigt eine signifikante Zweifache-Interaktion „Geschlecht\*Gruppe“ und einen signifikanten Haupteffekt „Messzeitpunkt“ (vgl. Tabelle 42).

*Tabelle 42: Ergebnisse der robusten Varianzanalyse zum Sachinteresse Naturwissenschaften in der Mittelstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.*

<b>Effekt</b>	<b>Q</b>	<b>p</b>
Geschlecht	17.17**	< 0.001
Gruppe	0.50	0.494
Messzeitpunkt	6.79*	0.010
Geschlecht*Gruppe	5.57*	0.023
Geschlecht*Messzeitpunkt	1.25	0.265
Gruppe*Messzeitpunkt	0.04	0.835
Geschlecht*Gruppe*Messzeitpunkt	0.13	0.706

In Abbildung 22 sind die Mittelwerte des Sachinteresses an Naturwissenschaften aufgetragen. Demnach nimmt bei den Schülern bei beiden Gruppen das Interesse zu. Bei den Schülerinnen kommt es zu keiner wahrnehmbaren Verbesserung des Interesses. Hierbei liegt die Vergleichsgruppe auch unter der Interventionsgruppe. Bei den Jungen wiederum liegen die Gruppen nahe beieinander, wobei die Vergleichsgruppe hier leicht darüber liegt. Die Zuwächse sind bei den Schülern allerdings gleich groß.

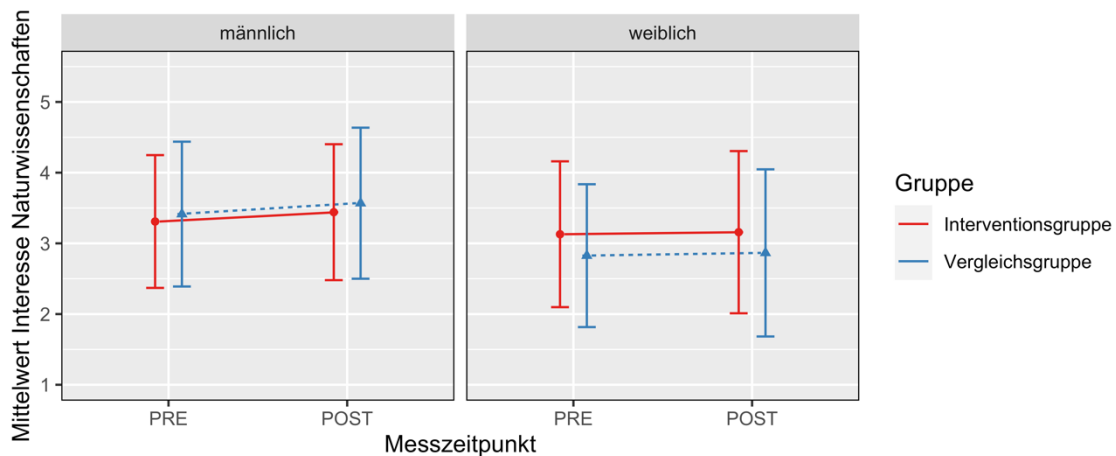


Abbildung 22: Mittelwert mit Standardabweichung des Sachinteresses Naturwissenschaften in der Mittelstufe, aufgeteilt nach Geschlecht und Gruppe im Pre- und Posttest in der Mittelstufe.

Der Haupteffekt Messzeitpunkt ist signifikant mit einer geringen Effektstärke (*robust*  $t(203) = 2.49$ ,  $p = 0.013$ ,  $\hat{\xi} = 0.09$ ;  $M_{PRE} = 3.18 \pm 1.02$ ,  $M_{POST} = 3.27 \pm 1.11$ ). Damit nimmt, unabhängig von der Gruppe und vom Geschlecht, das Sachinteresse zu. Die Auswertung der Zweibege-Interaktion ist in Tabelle 43 aufgeführt.

Tabelle 43: Ergebnisse der Untersuchung der einfachen Haupteffekte der Interaktion „Geschlecht\*Gruppe“ in der Mittelstufe für das Sachinteresse.

Gruppe	$t$	$df$	$p$	$\hat{\xi}$
Interventionsgruppe	1.39	107.51	0.166	0.15
Vergleichsgruppe	4.23**	84.96	< 0.001	0.44
Geschlecht				
weiblich	2.12*	91.56	0.037	0.22
männlich	1.31	83.18	0.195	0.14

Demnach unterscheiden sich in der Vergleichsgruppe die Geschlechter signifikant mit einer hohen Effektstärke voneinander, wonach die Schüler ( $M = 3.49 \pm 0.99$ ) insgesamt ein höheres Sachinteresse erfahren als die Schülerinnen ( $M = 2.85 \pm 1.03$ ). Weiterhin unterschieden sich die beiden Gruppen bei den Schülerinnen. Die weiblichen Lernenden

der Interventionsgruppe zeigen bereits im Pre-Test ein signifikant höheres Sachinteresse, das sich jedoch über die Zeit nicht verändert ( $M_I = 3.14 \pm 0.99$ ,  $M_V = 2.85 \pm 1.03$ ).

### **Wahrgenommene Wahlfreiheit: Geschlecht, Gruppe, Messzeitpunkt**

Der Einfluss auf die wahrgenommene Wahlfreiheit (Autonomie) beim Lernen mit Arbeitsblättern unter Beachtung des Geschlechts, der Einsatzdauer und der Gruppenzugehörigkeit wurde ebenfalls mit einer gemischten Varianzanalyse untersucht. Die deskriptiven Beschreibungen der Stichproben sind Tabelle 44 zu entnehmen.

*Tabelle 44: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf die wahrgenommene Wahlfreiheit beim Lernen mit Arbeitsblättern in der Mittelstufe.*

		<i>Interventionsgruppe</i>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		109				99			
Mittelwert		4.17	4.08	3.69	3.69	4.10	4.07	3.84	3.75
(Standardabweichung)		(0.67)	(0.76)	(0.95)	(0.92)	(0.78)	(0.68)	(0.81)	(0.96)
Median		4.25	4.25	3.75	3.75	4.25	4.00	4.00	3.75
		<i>Vergleichsgruppe</i>							
		Weiblich				Männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		71				78			
Mittelwert		3.32	3.22	3.02	2.93	3.36	3.27	3.21	3.20
(Standardabweichung)		(0.89)	(0.96)	(1.05)	(0.97)	(0.95)	(1.00)	(1.00)	(0.97)
Median		3.25	3.25	3.00	3.00	3.50	3.25	3.18	3.17

Die Prüfung der Voraussetzungen zeigt eine Verletzung der Sphärizität ( $p < 0.001$ ), Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ), der Varianzhomogenität für die erste ( $p = 0.005$ ) und zweite Stunde ( $p = 0.002$ ) sowie der Homogenität der Kovarianzmatrizen für den Faktor „Geschlecht“ ( $p < 0.001$ ). Insgesamt sind 26 Ausreißer vorhanden, wovon keiner als extrem eingestuft wird. Aus diesem Grund wurde erneut eine robuste mixed ANOVA mit Bootstrapping gewählt. Die Ergebnisse sind Tabelle 45 zu entnehmen.



Tabelle 45: Ergebnisse der robusten Varianzanalyse zur wahrgenommenen Wahlfreiheit beim Lernen mit Arbeitsblättern in der Mittelstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap-Verfahren.

Effekt	$Q$	$p$
Geschlecht	0.66	0.424
Gruppe	84.40**	< 0.001
Messzeitpunkt	14.89**	< 0.001
Geschlecht*Gruppe	0.48	0.479
Geschlecht*Messzeitpunkt	0.76	0.356
Gruppe*Messzeitpunkt	1.20	0.177
Geschlecht*Gruppe*Messzeitpunkt	0.07	0.965

Die Analyse zeigt keine signifikanten Dreiwege- und Zweiwege-Interaktionen. Abbildung 23 verdeutlicht diesen Befund.

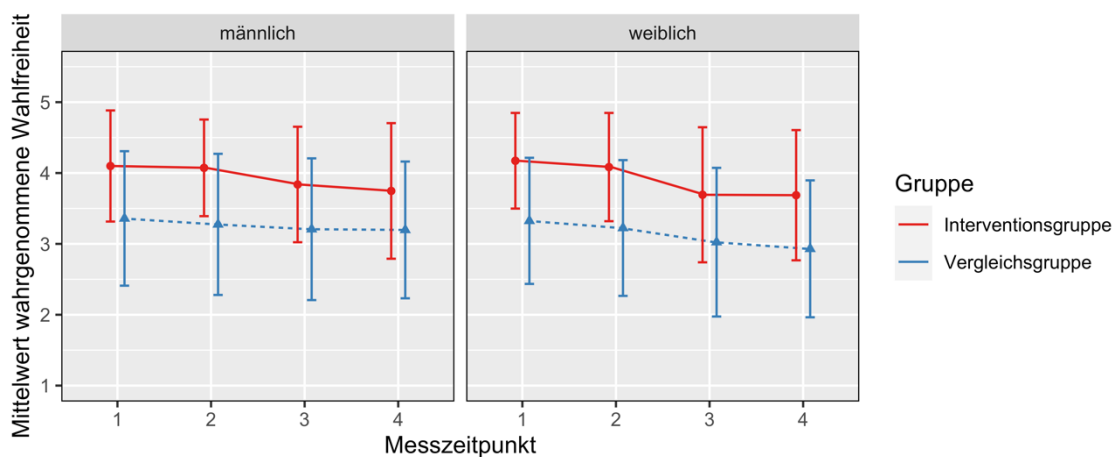


Abbildung 23: Mittelwerte mit Standardabweichung der wahrgenommenen Wahlfreiheit beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.

Die beiden Graphen verlaufen für die Geschlechter im Verlauf der Untersuchung nahezu parallel. Es ist über alle Gruppen hinweg eine leichte Abnahme der Wahlfreiheit beim Lernen mit Arbeitsblättern sowie ein Unterschied zwischen der Interventions- und Vergleichsgruppe zu sehen, was sich in den signifikanten Haupteffekten widerspiegelt. Um einen besseren Eindruck über die Stärke des Haupteffekts „Gruppe“ zu erhalten, wurde ein robuster t-Test durchgeführt ( $robust\ t(186.39) = 9.92, p < 0.001, \hat{\xi} = 0.65$ ). Die Effektstärke ist als hoch einzustufen, wodurch die Interventionsgruppe ( $M_I = 3.92 \pm 0.65$ ) über den gesamten Zeitraum eine höhere Wahlfreiheit erlebt als die Vergleichsgruppe

( $M_V = 3.20 \pm 0.82$ ). Die Effektstärke des Haupteffektes „Messzeitpunkt“ beträgt  $\delta_t = 0.25$ . Zwischen der ersten und vierten Stunde ist die Effektstärke  $\delta_t = 0.40$ . Sie ist damit auf einem schwachen bis mittleren Niveau, sodass über die Unterrichtsreihe, unabhängig von der Methodik, die wahrgenommene Wahlfreiheit abnimmt.

### **Kompetenzerleben: Geschlecht, Gruppe, Messzeitpunkt**

Auch das Kompetenzerleben beim Bearbeiten der Arbeitsblätter wurde nach einer gemischten Varianzanalyse untersucht. Aufgrund der Verletzung der Sphärizität ( $p < 0.001$ ), Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ), Varianzhomogenität für die erste Stunde ( $p < 0.001$ ) sowie der Homogenität der Kovarianzmatrizen für den Faktor „Geschlecht“ ( $p = 0.012$ ) und „Gruppe“ ( $p < 0.001$ ) wurde erneut eine robuste Variante der Varianzanalyse mit getrimmten Mittelwerten und Bootstrapping durchgeführt. Die deskriptive Statistik ist Tabelle 46 zu entnehmen. Insgesamt sind 16 Ausreißer vorhanden, wovon einer als extrem eingestuft wird.

*Tabelle 46: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf das Kompetenzerleben beim Lernen mit Arbeitsblättern in der Mittelstufe.*

	<i>Interventionsgruppe</i>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	109				99			
Mittelwert	4.35	4.23	3.82	3.77	4.31	4.28	4.02	3.93
(Standardabweichung)	(0.56)	(0.65)	(0.76)	(0.79)	(0.59)	(0.69)	(0.70)	(0.76)
Median	4.50	4.25	4.00	3.75	4.50	4.50	4.00	4.00
	<i>Vergleichsgruppe</i>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	71				78			
Mittelwert	3.77	3.46	3.42	2.94	4.04	3.83	3.75	3.50
(Standardabweichung)	(0.88)	(0.76)	(0.61)	(0.85)	(0.58)	(0.78)	(0.86)	(0.83)
Median	4.00	3.50	3.25	3.07	4.00	3.80	4.00	3.50

Die Testergebnisse der ANOVA sind in nachfolgender Tabelle (47) fixiert.

Tabelle 47: Ergebnisse der robusten Varianzanalyse zum Kompetenzerleben beim Lernen mit Arbeitsblättern in der Mittelstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.

Effekt	Q	p
Geschlecht	12.86**	< 0.001
Gruppe	60.40**	< 0.001
Messzeitpunkt	32.37**	< 0.001
Geschlecht*Gruppe	3.92	0.058
Geschlecht*Messzeitpunkt	2.36*	0.020
Gruppe*Messzeitpunkt	4.31**	0.003
Geschlecht*Gruppe*Messzeitpunkt	0.19	0.853

Die Ergebnisse sind für die beiden Zweiwege-Interaktionen „Geschlecht\*Messzeitpunkt“ und „Gruppe\*Messzeitpunkt“ signifikant, weshalb nun die einfachen Haupteffekte analysiert werden. Die Interaktion ist in Abbildung 24 zu sehen. Demnach nimmt bei beiden Geschlechtern in der Vergleichsgruppe das Kompetenzerleben mit zunehmender Unterrichtszeit stärker ab. Gleichzeitig bewerten Schüler in der Vergleichsgruppe das Kompetenzerleben höher als Schülerinnen. In der Interventionsgruppe fällt insbesondere bei den weiblichen Lernenden in der dritten Stunde das Kompetenzerleben ab. Es verbleibt jedoch, wie auch bei den Jungen, in der vierten Stunde auf einem ähnlichen Niveau, während es bei den Schüler:innen der Vergleichsgruppe weiter abfällt.

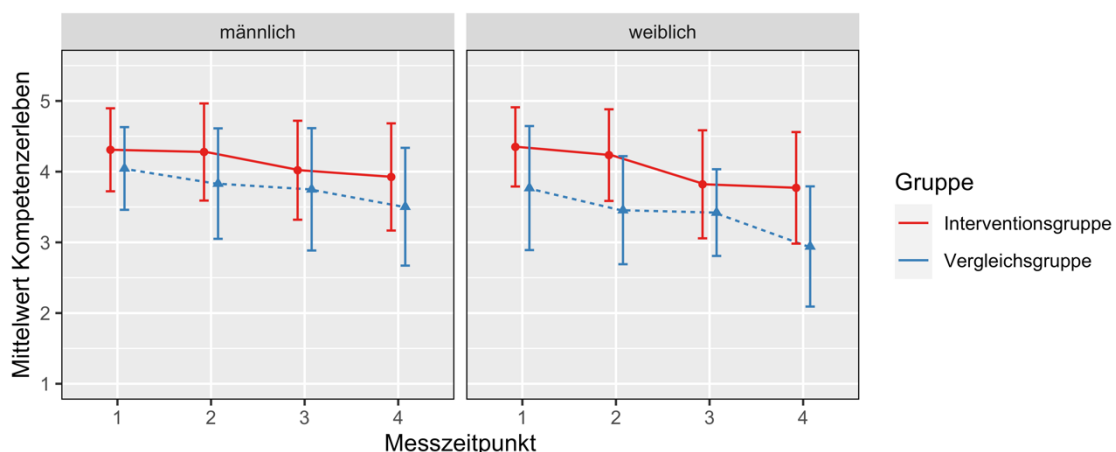


Abbildung 24: Mittelwerte mit Standardabweichung des Kompetenzerlebens beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.

Für die Interaktion, die das Geschlecht betrifft, wurde zunächst der Zwischensubjektfaktor zu den unterschiedlichen Messzeitpunkten untersucht (vgl. Tabelle 48).

Tabelle 48: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Geschlecht“ der Interaktion „Geschlecht\*Messzeitpunkt“ (Kompetenzerleben) in der Mittelstufe.

Messzeitpunkt	<i>t</i>	<i>df</i>	<i>p</i>	$\hat{\xi}$	$M_w(SD)$	$M_m(SD)$
1	0.34	206.26	0.735	0.03	4.12 (0.76)	4.19 (0.60)
2	1.56	210.06	0.120	0.12	3.93 (0.79)	4.08 (0.76)
3	3.67**	207.20	< 0.001	0.31	3.66 (0.73)	3.90 (0.79)
4	2.74**	208.54	0.007	0.24	3.44 (0.91)	3.74 (0.82)

Die Geschlechter unterscheiden sich, unter Vernachlässigung der Gruppenzugehörigkeit, signifikant ab der dritten Stunde, wobei Schüler mehr Kompetenzerleben beim Lernen mit Arbeitsblättern empfinden als Schülerinnen. Als nächstes wurde der Innersubjektfaktor für die Schülerinnen mit einer robusten ANOVA mit Messwertwiederholung analysiert. Dabei zeigte sich ein signifikantes Ergebnis (*robust*  $F(3,321) = 56.88, p < 0.001$ ). Die Post-hoc-Tests zeigten zwischen allen Messzeitpunkten signifikante Unterschiede (vgl. Anhang 9.3.2). Die Effektstärke beträgt dabei  $\delta_t = 0.44$ . Zwischen der ersten und vierten Stunde ist die Effektstärke  $\delta_t = 0.70$ . Die Effektstärken liegen auf einem mittleren bis starken Niveau. Auch bei den Schülern zeigte sich die ANOVA signifikant (*robust*  $F(2.84, 300.93) = 20.72, p < 0.001$ ). Im Post-hoc-Test zeigt sich allerdings nur zwischen der zweiten und dritten Stunde ein signifikanter Unterschied, der von Interesse ist (vgl. Anhang 9.3.2). Dieser Unterschied ist bereits in Abbildung 24 erkennbar. Die Effektstärke beträgt  $\delta_t = 0.31$ . Zwischen der ersten und vierten Stunde liegt sie bei  $\delta_t = 0.51$ . Bei den Schülerinnen ist die Abnahme folglich stärker.

Als nächstes wurde die signifikante Interaktion „Gruppe\*Messzeitpunkt“ in Bezug auf einfache Haupteffekte untersucht. Dazu wurde wie zuvor der Zwischensubjektfaktor (vgl. Tabelle 49) untersucht.

Tabelle 49: Untersuchung des Zwischensubjektfaktors „Gruppe“ der Interaktion „Gruppe\*Messzeitpunkt“ (Kompetenzerleben) in der Mittelstufe.

Messzeitpunkt	<i>t</i>	<i>df</i>	<i>p</i>	$\hat{\xi}$	$M_I(SD)$	$M_V(SD)$
1	4.92**	150.82	< 0.001	0.39	4.33 (0.57)	3.91 (0.75)
2	8.43**	170.32	< 0.001	0.55	4.26 (0.67)	3.65 (0.79)
3	4.03**	205.31	< 0.001	0.30	3.92 (0.74)	3.59 (0.77)
4	5.97**	187.60	< 0.001	0.54	3.84 (0.78)	3.24 (0.88)

Bei der Untersuchung zeigt sich zu jedem Messzeitpunkt ein signifikanter Unterschied zwischen den Gruppen, der auf einer hohen Effektstärke beruht. Demnach bewerten Lernende, die mit HyperDocs arbeiteten, das Kompetenzerleben als deutlich höher als die Lernenden der analogen Variante. Zuletzt erfolgt eine Auswertung des Innersubjektfaktors. Die robuste messwertwiederholte ANOVA zeigt ein signifikantes Ergebnis für die Interventionsgruppe ( $robust F(2.81, 350.68) = 41.09, p < 0.001$ ), wobei sich lediglich, wie bereits beim Faktor „Geschlecht“, interessante signifikante Unterschiede zwischen der zweiten und dritten Stunde zeigen (vgl. Anhang 9.3.2). In der Vergleichsgruppe zeigt sich ebenso ein Unterschied über die Zeit ( $robust F(2.93, 263.74) = 30.71, p < 0.001$ ). Hierbei unterscheiden sich insbesondere die erste und zweite sowie die dritte und vierte Stunde voneinander, während zwischen der zweiten und dritten Stunde kein Unterschied im Kompetenzerleben festzustellen ist (vgl. Anhang 9.3.2). Die Effektstärke beträgt in der Interventionsgruppe  $\delta_t = 0.37$ . Zwischen der ersten und vierten Stunde ist die Effektstärke  $\delta_t = 0.56$ . Für die Vergleichsgruppe liegen die Effektstärken bei  $\delta_t = 0.36$  beziehungsweise  $\delta_t = 0.65$ . Damit nimmt das Kompetenzerleben von der ersten zur vierten Stunde in der Vergleichsgruppe stärker ab.

### **Druck: Geschlecht, Gruppe, Messzeitpunkt**

Aufgrund der Vielzahl an bereits durchgeführten Signifikanztests und der geringen Unterschiede zwischen der Interventions- und Vergleichsgruppe, wird die Variable „Druck“ (SDT) lediglich deskriptiv und graphisch analysiert.

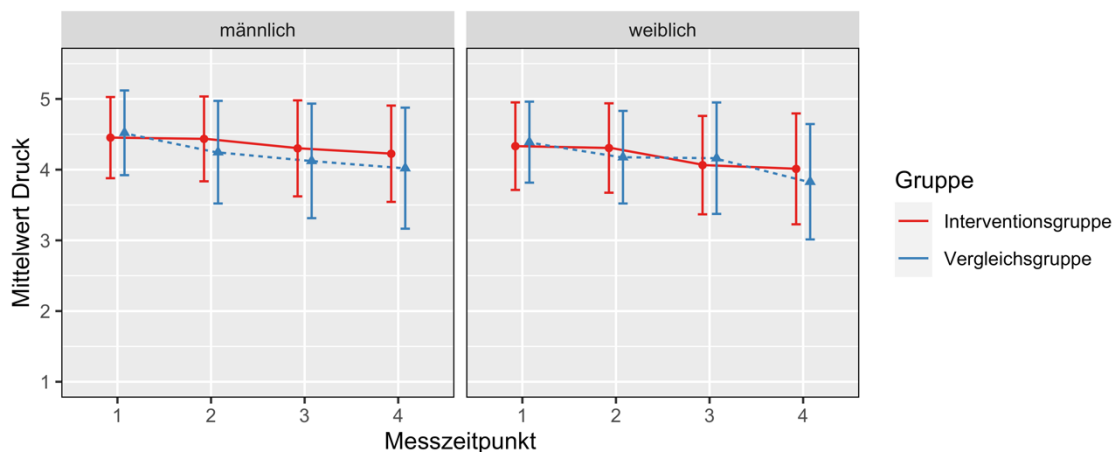


Abbildung 25: Mittelwerte mit Standardabweichung des Drucks beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe. Ein hoher Mittelwert entspricht einem geringer wahrgenommenen Druck.

Sowohl die Geschlechter als auch die Gruppen unterscheiden sich über die Einsatzdauer nicht wahrnehmbar voneinander (vgl. Abbildung 25). Es ist jedoch eine tendenzielle Zunahme des Drucks beim Lernen mit Arbeitsblättern in beiden Gruppen zu erkennen, da hohe Werte einem geringeren Druck entsprechen. Verdeutlicht wird dies durch die deskriptive Statistik (vgl. Tabelle 50). Tendenziell nimmt bei den Schülerinnen der Druck jedoch von der ersten zur letzten Stunde gegenüber den Schülern stärker zu.

*Tabelle 50: Deskriptive Statistik der Interventions- und Vergleichsgruppe zum Druck beim Lernen mit Arbeitsblättern in den einzelnen Unterrichtsstunden, aufgeteilt nach Gruppe und Geschlecht in der Mittelstufe.*

	<i>Interventionsgruppe</i>								
	weiblich				männlich				
	Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang		109				99			
Mittelwert	4.33	4.31	4.06	4.01	4.45	4.43	4.30	4.22	
(Standardabweichung)	(0.62)	(0.63)	(0.70)	(0.78)	(0.57)	(0.60)	(0.68)	(0.68)	
Median	4.25	4.50	4.00	4.00	4.50	4.50	4.50	4.25	
	<i>Vergleichsgruppe</i>								
	weiblich				männlich				
	Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang		71				78			
Mittelwert	4.39	4.18	4.16	3.83	4.52	4.25	4.12	4.02	
(Standardabweichung)	(0.57)	(0.66)	(0.79)	(0.82)	(0.62)	(0.73)	(0.81)	(0.86)	
Median	4.50	4.25	4.18	3.75	4.75	4.38	4.25	4.00	

### 5.2.2 Vergleich zwischen Interventions- und Vergleichsgruppe in Abhängigkeit vom Geschlecht und Messzeitpunkt in der Oberstufe

Das Vorgehen der Auswertung der Datensätze für die Oberstufe ist äquivalent zu dem der Mittelstufe. Es werden daher zunächst die Voraussetzungen für eine gemischte Varianzanalyse geprüft und diese im Anschluss durchgeführt. Danach folgt die Auswertung der Interaktionen oder Haupteffekte.

**Interesse: Geschlecht, Gruppe, Messzeitpunkt***Interesse beim Lernen mit Arbeitsblättern*

Bei Prüfung der Voraussetzungen ergibt sich eine Verletzung der Sphärizität ( $p = 0.002$ ) und der Normalverteilung der Residuen (*QQ-Plot, Shapiro-Wilk-Test:  $p < 0.001$* ). Zusätzlich befinden sich in der Stichprobe 14 Ausreißer, wovon keiner als extrem einzustufen ist. Aufgrund der Verletzung der Sphärizität wurde sich für eine robuste Variante der ANOVA entschieden. In Tabelle 51 ist die deskriptive Statistik dargestellt.

*Tabelle 51: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf das Interesse (SDT) beim Lernen mit Arbeitsblättern in der Oberstufe.*

		<i>Interventionsgruppe</i>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		45				56			
Mittelwert		4.38	3.96	3.58	3.77	4.09	3.91	3.66	3.70
(Standardabweichung)		(0.62)	(0.85)	(0.97)	(0.88)	(0.79)	(0.90)	(0.95)	(0.98)
Median		4.50	4.05	3.75	3.75	4.12	4.00	3.75	3.88
		<i>Vergleichsgruppe</i>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		22				23			
Mittelwert		3.55	3.08	2.88	3.5	3.02	3.19	3.07	3.45
(Standardabweichung)		(0.93)	(0.96)	(0.93)	(1.11)	(0.83)	(0.96)	(1.06)	(0.98)
Median		3.50	3.12	2.75	3.62	3.00	3.25	3.25	3.50

Aufgrund der niedrigen Teilstichproben in Bezug auf das Geschlecht wurde die Varianzanalyse mit Bootstrapping durchgeführt. Die Ergebnisse sind Tabelle 52 zu entnehmen.

Tabelle 52: Ergebnisse der robusten Varianzanalyse zum Interesse (SDT) beim Lernen mit Arbeitsblättern in der Oberstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.

Effekt	Q	p
Geschlecht	0.26	0.633
Gruppe	23.91**	< 0.001
Messzeitpunkt	4.32**	< 0.001
Geschlecht*Gruppe	< 0.01	0.980
Geschlecht*Messzeitpunkt	1.42*	0.013
Gruppe*Messzeitpunkt	4.01**	< 0.001
Geschlecht*Gruppe*Messzeitpunkt	0.42	0.366

Die Analyse zeigt keine signifikante Dreiwege-Interaktion und zwei signifikante Zweiwege-Interaktionen. Die Haupteffekte werden daher nicht analysiert.

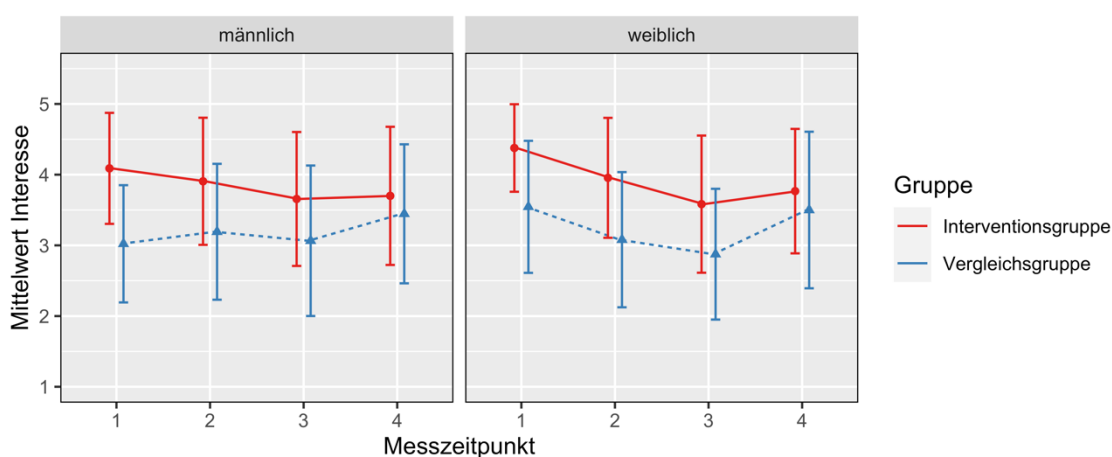


Abbildung 26: Mittelwerte mit Standardabweichung des Interesses (SDT) beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe.

In Abbildung 26 sind die Zweiwege-Interaktionen identifizierbar. Demnach gibt es eine stärkere Abnahme des Interesses beim Lernen mit Arbeitsblättern bei Schülerinnen in beiden Gruppen, wobei zur vierten Stunde eine Zunahme des Interesses zu sehen ist. Gleichzeitig unterscheiden sich die Schüler der Interventions- und Vergleichsgruppe im zeitlichen Verlauf von der ersten zur zweiten und von der dritten zur vierten Stunde. Als nächstes wurden die einfachen Haupteffekte analysiert. Dazu wurde mit der Interaktion „Geschlecht\*Messzeitpunkt“ begonnen. Für den Zwischensubjekteffekt ergibt sich lediglich für die erste Stunde ein signifikanter Unterschied (vgl. Tabelle 53).



Tabelle 53: Untersuchung des Zwischensubjektfaktors „Geschlecht“ der Interaktion „Geschlecht\*Messzeitpunkt“ (Interesse) in der Oberstufe.

Messzeitpunkt	$t$	$df$	$p$	$\hat{\xi}$	$M_w(SD)$	$M_m(SD)$
1	2.51*	86.60	0.014	0.29	4.10 (0.83)	3.78 (0.93)
2	0.15	84.33	0.879	0.03	3.67 (0.97)	3.70 (0.97)
3	0.79	80.65	0.430	0.10	3.35 (1.01)	3.48 (1.01)
4	0.08	85.01	0.937	0.03	3.68 (0.96)	3.63 (0.98)

Der Innersubjekteffekt zeigt sich bei den Schülerinnen signifikant (*robust*  $F(3, 120) = 15.11, p < 0.001$ ). Die Unterschiede sind zwischen allen Stunden, bis auf den Vergleich zwischen der zweiten und vierten Stunde signifikant (vgl. Anhang 9.3.3). Die Effektstärke beträgt  $\delta_t = 0.25$  und ist damit als schwach einzustufen. Die Effektstärke zwischen der ersten und der vierten Stunde beträgt  $\delta_t = 0.42$  und ist damit tendenziell mittelstark. Auch bei den männlichen Lernenden zeigt sich ein signifikanter Einfluss der Zeit auf das Interesse, wobei der Abfall nicht so stark ausfällt wie bei den Schülerinnen (*robust*  $F(2.78, 133.28) = 3.67, p < 0.016$ ). Es sind keine interessanten signifikanten Unterschiede zwischen den Stunden ersichtlich (vgl. Anhang 9.3.3). Die Effektstärke zwischen der ersten und der vierten Stunde beträgt  $\delta_t = 0.25$  und ist damit schwach. Die mittlere Effektstärke liegt bei  $\delta_t = 0.16$ . Unabhängig von der Unterrichtsmethode nimmt bei den Schülerinnen das Interesse beim Lernen mit Arbeitsblättern stärker ab als bei den Schülern.

Danach wurde die Interaktion „Gruppe\*Messzeitpunkt“ auf die einfachen Haupteffekte untersucht. Die Gruppen unterscheiden sich zu den ersten drei Messzeitpunkten signifikant voneinander (vgl. Tabelle 54).

Tabelle 54: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Gruppe“ der Interaktion „Gruppe\*Messzeitpunkt“ (Interesse) in der Oberstufe.

Messzeitpunkt	$t$	$df$	$p$	$\hat{\xi}$	$M_I(SD)$	$M_V(SD)$
1	6.48	41.95**	< 0.001	0.72	4.22 (0.73)	3.28 (0.91)
2	3.72	49.17**	< 0.001	0.58	3.93 (0.87)	3.14 (0.95)
3	3.95	67.86**	< 0.001	0.45	3.62 (0.95)	2.97 (0.99)
4	1.00	63.65	0.321	0.14	3.73 (0.93)	3.47 (1.03)

Die Effektstärken sind dabei als hoch anzusehen. Unter Betrachtung der Mittelwerte fällt auf, dass beide Gruppen in der vierten Stunde eine Zunahme des Interesses beim Lernen

mit Arbeitsblättern verzeichnen, wobei der Ausschlag bei der Vergleichsgruppe deutlich größer ausfällt. Anschließend wurde der Innersubjektfaktor untersucht. Sowohl in der Interventionsgruppe (*robust*  $F(2.7, 161.98) = 19.47, p < 0.001$ ) als auch in der Vergleichsgruppe (*robust*  $F(2.71, 70.57) = 6.00, p = 0.002$ ) zeigt sich ein signifikanter Effekt der Einsatzdauer. Während sich in der Interventionsgruppe insbesondere die ersten Stunden signifikant voneinander unterscheiden, sind es in der Vergleichsgruppe lediglich die dritte und vierte Stunde (vgl. Anhang 9.3.3). Die mittlere Effektstärke beträgt dabei in der Interventionsgruppe  $\delta_t = 0.33$  und die Effektstärke zwischen der ersten und vierten Stunde  $\delta_t = 0.58$ . In der Vergleichsgruppe liegt die mittlere Effektstärke bei  $\delta_t = 0.14$  und die Effektstärke zwischen der ersten und vierten Stunde bei  $\delta_t = 0.22$ . In der Vergleichsgruppe nimmt das Interesse mit einer sehr schwachen Effektstärke, verursacht durch die Schüler, zu, während in der Interventionsgruppe das Interesse mit einer schwachen bis mittleren Effektstärke abnimmt.

*Fachinteresse Chemie und Sachinteresse Naturwissenschaften*

Aufgrund der Verletzung der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ) und dem Vorhandensein von drei Ausreißern wurde eine robuste Variante der Varianzanalyse gewählt. Die deskriptive Statistik ist in Tabelle 55 aufgeführt.

*Tabelle 55: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf das Fachinteresse Chemie in der Oberstufe.*

	<i>Interventionsgruppe</i>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt				
Stichprobenumfang	43		50	
Mittelwert (Standardabweichung)	3.62 (1.01)	3.53 (0.98)	3.96 (0.88)	4.04 (0.93)
Median	4.00	3.50	4.00	4.00
	<i>Vergleichsgruppe</i>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt				
Stichprobenumfang	23		24	
Mittelwert (Standardabweichung)	3.76 (1.03)	3.89 (1.00)	4.00 (0.88)	3.85 (0.98)
Median	4.00	3.50	4.00	4.00

Die Ergebnisse der robusten gemischten Varianzanalyse zeigen eine signifikante Dreiwege-Interaktion, womit sich die Geschlechter, abhängig von der Gruppe zu den Messzeitpunkten unterscheiden (vgl. Tabelle 56).

Tabelle 56: Ergebnisse der robusten Varianzanalyse zum Fachinteresse Chemie in der Oberstufe. Die *Q*-Werte stammen aus dem non-bootstrap-Verfahren.

<i>Effekt</i>	<i>Q</i>	<i>p</i>
Geschlecht	2.51	0.115
Gruppe	0.26	0.583
Messzeitpunkt	0.10	0.735
Geschlecht*Gruppe	0.74	0.406
Geschlecht*Messzeitpunkt	0.07	0.790
Gruppe*Messzeitpunkt	< 0.01	0.998
Geschlecht*Gruppe*Messzeitpunkt	4.05*	0.030

Es folgt nun die Analyse der einfachen, einfachen Haupteffekte<sup>14</sup> (Dreiwege-Interaktion). Dazu wird die Interaktion „Geschlecht\*Gruppe“ zu den beiden Messzeitpunkten ausgewertet. Die Analyse erfolgt mittels des R-Pakets „welchADF“ (Villacorta, 2017), das eine mixed ANOVA mit Bootstrapping beinhaltet. Das Paket nutzt die gleiche Teststatistik, den Welch-James-Test (Johansen, 1980), wie bereits Mair & Wilcox (2020). Die Ergebnisse sind in Tabelle 57 zusammengefasst.

Tabelle 57: Teststatistik der jeweiligen Untersuchung der Interaktion „Geschlecht\*Gruppe“ (Fachinteresse Chemie) im Pre- und Posttest in der Oberstufe.

<i>Messzeitpunkt</i>	<i>WJ</i>	<i>df<sub>N</sub></i>	<i>df<sub>D</sub></i>	<i>p</i>
PRE	0.04	1	45.14	0.83
POST	2.34	1	52.29	0.122

Demnach ist zu keinem der beiden Messzeitpunkte die Zweiwege-Interaktion signifikant. Es sind auch keine signifikanten Haupteffekte vorhanden. In Abbildung 27 wird der Grund für die signifikante Dreiwege-Interaktion deutlich: Bei der Interventionsgruppe nimmt das Fachinteresse bei den Schülern zu, während es bei den Schülerinnen abnimmt. Der umgekehrte Fall liegt bei der Vergleichsgruppe vor. Hier nimmt das Interesse der weiblichen Lernenden zu, während das der männlichen abnimmt.

<sup>14</sup> Nach Maxwell et al. (2017, S. 426) „simple, simple main effects“

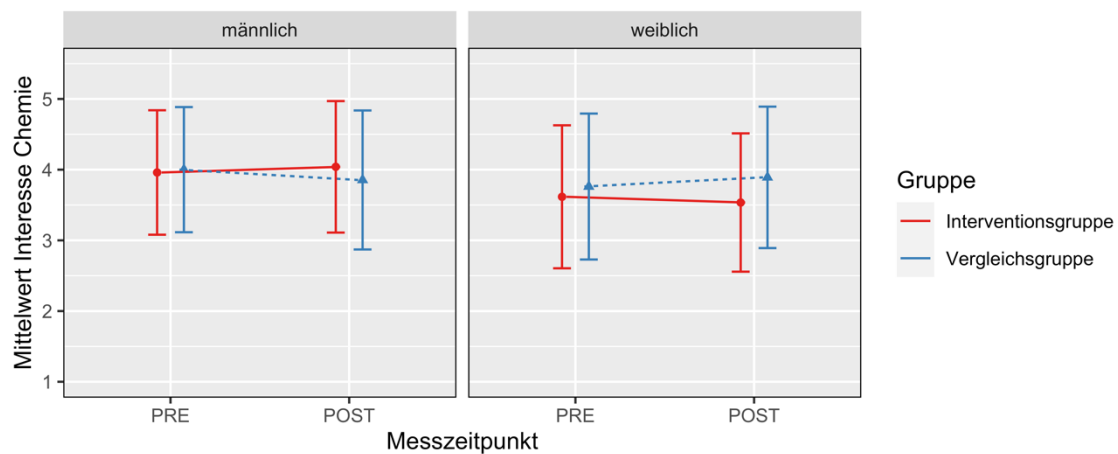


Abbildung 27: Mittelwerte mit Standardabweichung des Fachinteresses Chemie, aufgeteilt nach Geschlecht und Gruppe im Pre- und Posttest in der Oberstufe.

Wie auch in der Mittelstufe erfolgt die Analyse des Sachinteresses mit einer robusten gemischten Varianzanalyse. Die Statistik ist in Tabelle 58 aufgeführt.

Tabelle 58: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf das Sachinteresse an Naturwissenschaften in der Oberstufe.

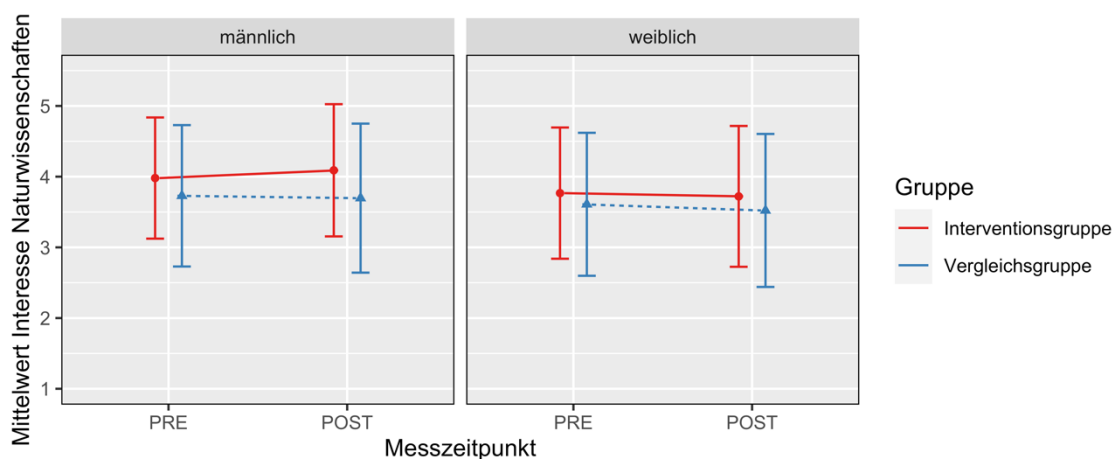
Messzeitpunkt	<b>Interventionsgruppe</b>			
	weiblich		männlich	
	1	2	1	2
Stichprobenumfang	43		50	
Mittelwert (Standardabweichung)	3.77 (0.93)	3.72 (1.00)	3.98 (0.86)	4.09 (0.94)
Median	4.00	4.00	4.25	4.50
Messzeitpunkt	<b>Vergleichsgruppe</b>			
	weiblich		männlich	
	1	2	1	2
Stichprobenumfang	23		24	
Mittelwert (Standardabweichung)	3.61 (1.01)	3.52 (1.08)	3.73 (1.00)	3.70 (1.05)
Median	3.50	4.00	3.50	3.75

Die Varianzanalyse zeigt keine signifikanten Ergebnisse (vgl. Tabelle 59).

*Tabelle 59: Ergebnisse der robusten Varianzanalyse zum Sachinteresse Naturwissenschaften in der Oberstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.*

<i>Effekt</i>	<i>Q</i>	<i>p</i>
Geschlecht	1.94	0.194
Gruppe	1.93	0.160
Messzeitpunkt	0.85	0.374
Geschlecht*Gruppe	0.27	0.591
Geschlecht*Messzeitpunkt	1.20	0.277
Gruppe*Messzeitpunkt	0.64	0.399
Geschlecht*Gruppe*Messzeitpunkt	0.95	0.327

Das anfängliche Interesse der Interventionsgruppe ist höher als das der Vergleichsgruppe. Während es bei den Schülern der Interventionsgruppe zu einem Zuwachs des Interesses kommt, nimmt das Sachinteresse an Naturwissenschaften bei allen anderen Teilgruppen leicht ab. Dabei ist bei den Schülerinnen die Abnahme bei beiden Gruppen gleich groß (vgl. Abbildung 28). Zwischen den Geschlechtern selbst bestehen nur geringe Unterschiede.



*Abbildung 28: Mittelwerte mit Standardabweichung des Sachinteresses Naturwissenschaften, aufgeteilt nach Geschlecht und Gruppe im Pre- und Posttest in der Oberstufe.*

### **Wahrgenommene Wahlfreiheit: Geschlecht, Gruppe, Messzeitpunkt**

Bei der Auswertung der wahrgenommenen Wahlfreiheit (Autonomie) beim Lernen mit Arbeitsblättern zeigen sich die Sphärizität ( $p = 0.002$ ), Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ), die Varianzhomogenität in der zweiten ( $p = 0.031$ ) und in der vierten Stunde ( $p = 0.017$ ) und die Homogenität der Kovarianzmatrizen für den Faktor Gruppe ( $p < 0.001$ ) verletzt. Es sind sieben normale Ausreißer vorhanden. Es wurde daher erneut eine robuste gemischte Varianzanalyse mit Bootstrapping angewandt. Die deskriptive Statistik ist Tabelle 60 zu entnehmen.

*Tabelle 60: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf die wahrgenommene Wahlfreiheit beim Lernen mit Arbeitsblättern in der Oberstufe.*

	<i>Interventionsgruppe</i>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	45				56			
Mittelwert (Standardabweichung)	3.92 (0.72)	3.67 (0.77)	3.61 (0.76)	3.67 (0.88)	3.85 (0.81)	3.65 (0.90)	3.57 (0.92)	3.54 (0.80)
Median	4.00	3.75	3.50	3.75	3.88	3.50	3.50	3.50
	<i>Vergleichsgruppe</i>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	22				23			
Mittelwert (Standardabweichung)	3.89 (1.09)	3.34 (1.19)	3.06 (1.15)	3.3 (1.35)	3.18 (0.99)	2.82 (0.99)	2.85 (1.01)	3.29 (0.96)
Median	4.00	3.00	3.12	3.75	3.00	3.00	3.00	3.00

Die Auswertung zeigt keine Dreiwege- und Zweiwege-Interaktion auf signifikantem Niveau. Der Haupteffekt der beiden Zwischensubjektfaktoren „Gruppe“ und „Messzeitpunkt“ sind beide signifikant (vgl. Tabelle 61).

Tabelle 61: Ergebnisse der robusten Varianzanalyse zur wahrgenommenen Wahlfreiheit beim Lernen mit Arbeitsblättern in der Oberstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap-Verfahren.

Effekt	$Q$	$p$
Geschlecht	2.27	0.182
Gruppe	7.00*	0.015
Messzeitpunkt	5.02**	< 0.001
Geschlecht*Gruppe	1.55	0.265
Geschlecht*Messzeitpunkt	0.55	0.282
Gruppe*Messzeitpunkt	0.73	0.147
Geschlecht*Gruppe*Messzeitpunkt	1.00	0.090

In Abbildung 29 werden die Haupteffekte sichtbar. Demnach unterscheidet sich die Wahlfreiheit der Gruppen in den einzelnen Stunden ( $M_I = 3.68 \pm 0.70$ ;  $M_V = 3.21 \pm 0.96$ ). Der robuste t-Test zeigt eine mittelstarke Effektstärke ( $\text{robust } t(35.23) = 2.87, p = 0.007, \xi = 0.40$ ).

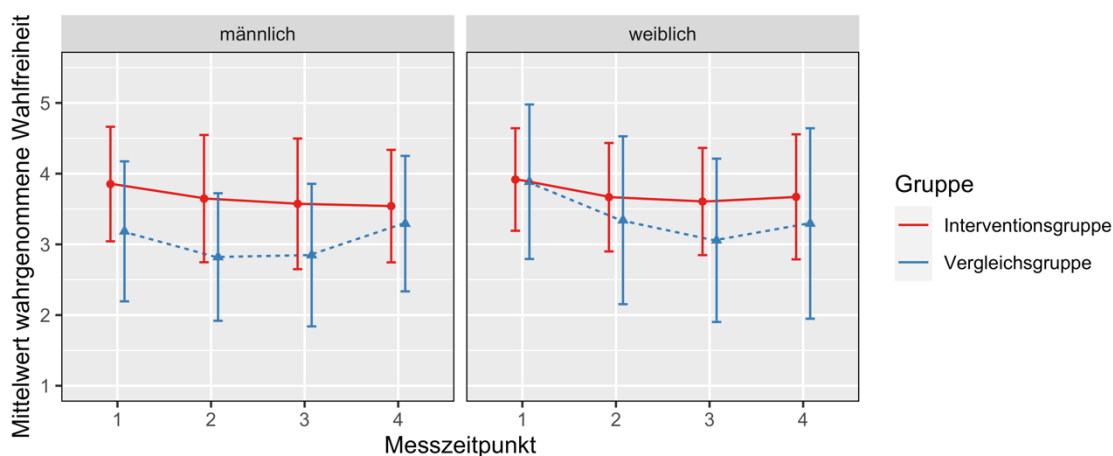


Abbildung 29: Mittelwerte mit Standardabweichung der wahrgenommenen Wahlfreiheit beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe.

Zunächst ist dabei eine Abnahme und dann zur letzten Stunde eine Zunahme der wahrgenommenen Wahlfreiheit beim Lernen mit den Arbeitsblättern zu erkennen. Die Zunahme ist bei der Vergleichsgruppe höher. Die genauen signifikanten Unterschiede zwischen den einzelnen Stunden sind Anhang 9.3.3 zu entnehmen. Die Effektstärke beträgt  $\delta_t = 0.22$  und ist damit als schwach einzustufen. Die Effektstärke zwischen der ersten und der vierten Stunde beträgt  $\delta_t = 0.34$  und ist damit ebenfalls schwach. Die wahrgenommene



Wahlfreiheit nimmt daher, unabhängig von der Methodik, leicht ab. Bei den Schülerinnen der Vergleichsgruppe liegt sie über der der Schüler.

### **Kompetenzerleben: Geschlecht, Gruppe, Messzeitpunkt**

Aufgrund der Verletzung der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ) und der Homogenität der Kovarianzmatrizen für die Faktoren „Gruppe“ ( $p = 0.022$ ) und „Geschlecht“ ( $p = 0.005$ ) sowie dem Vorhandensein von drei normalen Ausreißern wurde erneut eine robuste mixed ANOVA mit Bootstrapping eingesetzt. In Tabelle 62 ist die deskriptive Statistik aufgeführt.

*Tabelle 62: Deskriptive Statistik zum Kompetenzerleben beim Lernen mit Arbeitsblättern in der Oberstufe.*

		<i>Interventionsgruppe</i>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		45				56			
Mittelwert		4.26	3.9	3.77	3.91	4.42	4.06	3.73	3.9
(Standardabweichung)		(0.62)	(0.70)	(0.78)	(0.71)	(0.50)	(0.61)	(0.68)	(0.64)
Median		4.25	3.75	4.00	3.75	4.50	4.00	3.75	4.00
		<i>Vergleichsgruppe</i>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		22				23			
Mittelwert		4.22	3.28	2.82	3.44	4.05	3.62	3.41	3.71
(Standardabweichung)		(0.72)	(0.58)	(0.74)	(0.82)	(0.74)	(0.57)	(0.66)	(0.67)
Median		4.25	3.00	2.75	3.25	4.25	3.75	3.50	3.75

Die ANOVA zeigt eine signifikante Dreiwege-Interaktion, womit alle weiteren signifikanten Ergebnisse nicht analysiert werden (vgl. Tabelle 63).

*Tabelle 63: Ergebnisse der robusten Varianzanalyse zum Kompetenzerleben beim Lernen mit Arbeitsblättern in der Oberstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.*

<i>Effekt</i>	<i>Q</i>	<i>p</i>
Geschlecht	3.09	0.080
Gruppe	16.07**	< 0.001
Messzeitpunkt	15.43**	< 0.001
Geschlecht*Gruppe	1.40	0.260
Geschlecht*Messzeitpunkt	0.91	0.102
Gruppe*Messzeitpunkt	1.91**	0.005
Geschlecht*Gruppe*Messzeitpunkt	1.40*	0.038

Um die Dreiwege-Interaktion näher zu analysieren, werden nun die einfachen, einfachen Haupteffekte analysiert. Dazu wird zu jedem Messzeitpunkt die Interaktion „Gruppe\*Geschlecht“ geprüft. Die Ergebnisse der Zweiwege-Interaktionen zu den jeweiligen Stunden sind in Tabelle 64 festgehalten.

*Tabelle 64: Teststatistik der jeweiligen Untersuchung der Interaktion „Gruppe\*Geschlecht“ (Kompetenzerleben) zu den einzelnen Messzeitpunkten.*

<i>Messzeitpunkt</i>	<i>WJ</i>	<i>df<sub>N</sub></i>	<i>df<sub>D</sub></i>	<i>p</i>
1	1.52	1	43.15	0.222
2	1.98	1	61.05	0.173
3	5.32*	1	48.50	0.024
4	1.22	1	47.69	0.280

Die signifikante Zweiwege-Interaktion in der dritten Stunde zeigt, dass sich die Gruppen in Abhängigkeit vom Geschlecht unterscheiden. Hierbei zeigen vor allem die Schülerinnen in der Vergleichsgruppe ein niedrigeres Kompetenzerleben als die Schüler dieser Gruppe und als die Schülerinnen der Interventionsgruppe (vgl. Abbildung 30 & Tabelle 62). Die Analyse der einfachen Haupteffekte der Zweiwege-Interaktion ist in Tabelle 65 gezeigt. Die Analyse zeigt, dass sich die Geschlechter in der Vergleichsgruppe signifikant voneinander unterscheiden. Gleichzeitig liegt lediglich ein signifikanter Unterschied zwischen den Schülerinnen der Interventions- und Vergleichsgruppe vor.

Tabelle 65: Ergebnisse der Untersuchung der einfachen Haupteffekte der Interaktion „Geschlecht\*Gruppe“ (Kompetenzerleben) in der Oberstufe.

Gruppe	t	df	p	$\hat{\xi}$
Interventionsgruppe	0.37	43.62	0.713	0.06
Vergleichsgruppe	2.52*	26.11	0.018	0.59
Geschlecht				
weiblich	4.77**	32.39	< 0.001	0.70
männlich	1.48	19.54	0.155	0.31

Die Mittelwerte können Tabelle 62 entnommen werden. Zusätzlich zu der signifikanten Zweiwege-Interaktion in der dritten Stunde liegen noch weitere signifikante Haupteffekte in den anderen Stunden vor. Zur zweiten Stunde zeigen sich die beiden Haupteffekte „Gruppe“ ( $WJ(1, 61.05) = 25.38, p < 0.001$ ) und „Geschlecht“ ( $WJ(1, 61.05) = 8.24, p = 0.001$ ) der Zweiwege-Interaktion signifikant. Ebenso liegt ein signifikanter Haupteffekt der „Gruppe“ ( $WJ(1, 48.5) = 19.42, p < 0.001$ ) in der dritten Stunde vor, der jedoch aufgrund der signifikanten Zweiwege-Interaktion nicht interpretiert werden sollte. Zuletzt zeigt sich in der letzten Stunde auch ein signifikanter Haupteffekt der Gruppe ( $WJ(1, 47.69) = 4.76, p = 0.04$ ). Die jeweiligen Haupteffekte können anhand von Abbildung 30 nachvollzogen werden.

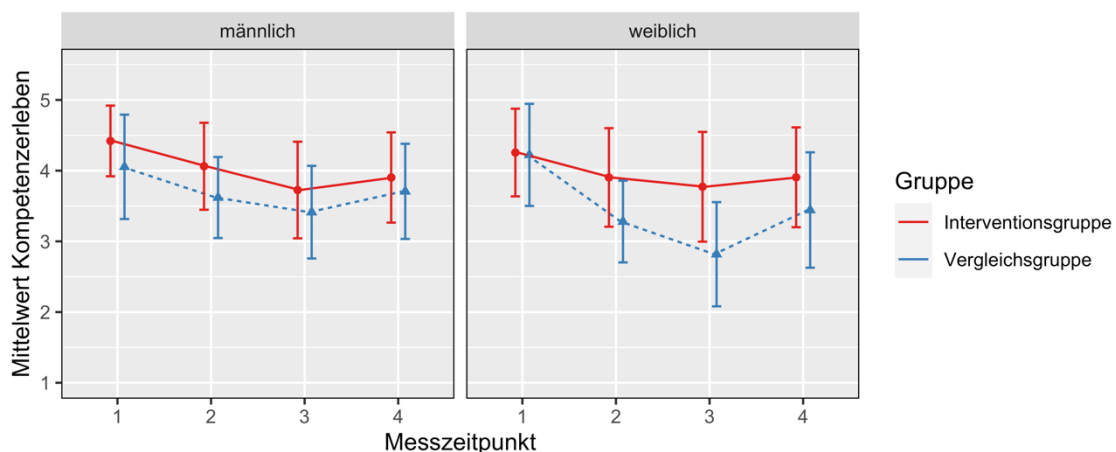


Abbildung 30: Mittelwerte mit Standardabweichung des Kompetenzerlebens beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe.

Demnach unterscheiden sich die beiden Gruppen insbesondere ab der zweiten Stunde. Um einen Eindruck über die Effektstärke zu erhalten, wurde der Haupteffekt „Gruppe“ über alle Stunden mittels eines robusten t-Tests analysiert ( $robust\ t(45.12) = 4.09$ ,

$p < 0.001$ ,  $\hat{\xi} = 0.49$ ). Dabei zeigt sich ein starker Effekt. Bei Betrachtung des Messzeitpunktes, in Abhängigkeit von der Gruppenzugehörigkeit, zeigt sich in der Interventionsgruppe über alle Stunden ein schwacher Effekt  $\delta_t = 0.37$  und zwischen der ersten und vierten Stunde ein mittelstarker Effekt  $\delta_t = 0.59$ . In der Vergleichsgruppe liegen die Effektstärken mit  $\delta_t = 0.30$  beziehungsweise  $\delta_t = 0.60$  (*erste – vierte Stunde*) auf einem ähnlichen Niveau, sodass in beiden Gruppen das Kompetenzerleben in etwa gleich stark abnimmt. Bei den Schülerinnen der Vergleichsgruppe ist die Abnahme jedoch bis zur dritten Stunde stärker, wie in Abbildung 30 zu erkennen ist.

### **Druck: Geschlecht, Gruppe, Messzeitpunkt**

Wie bereits bei der Mittelstufe wird der Druck beim Lernen mit Arbeitsblättern lediglich auf Basis der deskriptiven Beschreibung (vgl. Tabelle 66) und des Plots ausgewertet.

*Tabelle 66: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Druck beim Lernen mit Arbeitsblättern in der Oberstufe.*

		<i>Interventionsgruppe</i>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		45				56			
Mittelwert		4.41	4.23	4.19	4.27	4.53	4.39	4.3	4.4
(Standardabweichung)		(0.62)	(0.72)	(0.67)	(0.70)	(0.53)	(0.56)	(0.62)	(0.67)
Median		4.75	4.25	4.25	4.50	4.75	4.50	4.50	4.72
		<i>Vergleichsgruppe</i>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		22				23			
Mittelwert		4.32	3.90	3.99	4.31	4.62	3.96	3.82	4.22
(Standardabweichung)		(0.86)	(0.87)	(0.93)	(0.84)	(0.63)	(0.67)	(0.81)	(0.64)
Median		4.75	4.00	4.12	4.75	5.00	3.75	3.75	4.25

Anhand von Abbildung 31 zeigen sich keine größeren Unterschiede zwischen den Geschlechtern. Die Vergleichsgruppe liegt bei beiden Geschlechtern in der zweiten und dritten Stunde unter der Interventionsgruppe, womit Schüler:innen, die analog arbeiteten,

einen höheren Druck empfanden. Es ist eine leichte Zunahme des Drucks für alle Teilgruppen über die Zeit bis zur dritten Stunde erkennbar.

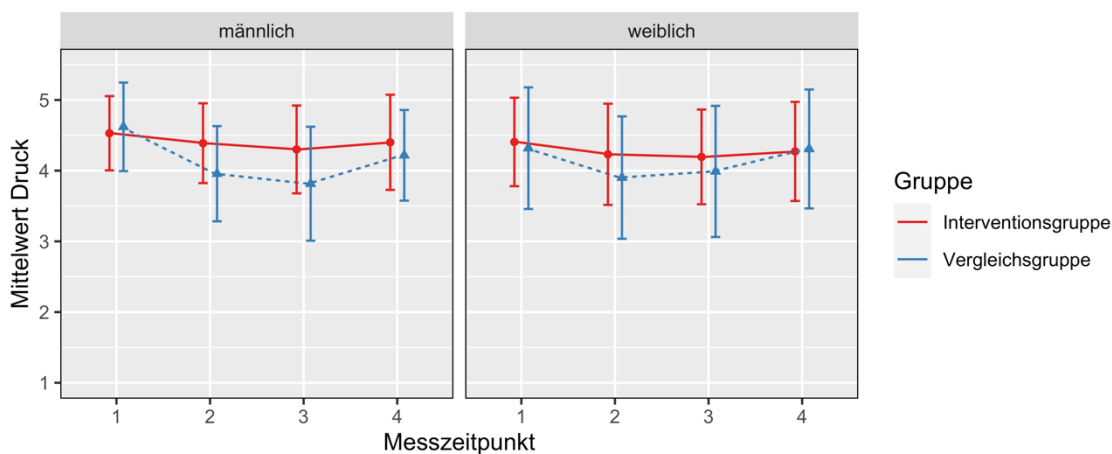


Abbildung 31: Mittelwerte mit Standardabweichung des Drucks beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe. Hohe Werte entsprechen einem geringeren Druck.

### 5.2.3 Vergleich zwischen den Schulformen in der Interventionsgruppe in der Mittelstufe

Um gegebenenfalls Unterschiede bezüglich der motivationalen Variablen bei Lernenden verschiedener Schulformen bei der Nutzung von HyperDocSystems festzustellen, die auch das Geschlecht betreffen, wurde erneut eine gemischte Varianzanalyse durchgeführt, wobei der Faktor „Gruppe“ des vorherigen Kapitels durch „Schulform“ ersetzt wurde. Es werden zunächst wieder die Voraussetzungen für eine mixed ANOVA im between-between-within Design geprüft. Anschließend erfolgt die deskriptive Beschreibung der Stichprobe mit anschließender Auswertung der Varianzanalyse.

#### **Interesse: Geschlecht, Schulform, Messzeitpunkt**

##### *Interesse beim Lernen mit HyperDocs*

Die Prüfung der Voraussetzungen ergab eine Verletzung der Sphärizität ( $p < 0.001$ ), der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ) und der Varianzhomogenität in der dritten Stunde ( $p < 0.014$ ). Es sind 24 starke Ausreißer vorhanden. Wegen dieser Verletzungen der Testvoraussetzungen der Varianzanalyse wurde eine robuste Lösung angewandt. Die deskriptive Statistik ist in Tabelle 67 zu finden.

Tabelle 67: Deskriptive Statistik des Interesses (SDT) beim Lernen mit HyperDocs im Schulformvergleich in der Mittelstufe.

		<b>Gymnasium</b>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		109				99			
Mittelwert		4.61	4.41	3.98	3.97	4.45	4.26	4.11	4.03
(Standardabweichung)		(0.52)	(0.64)	(0.86)	(0.91)	(0.78)	(0.81)	(0.75)	(0.86)
Median		4.75	4.50	4.25	4.00	4.75	4.50	4.25	4.25
		<b>Gesamtschule</b>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		53				64			
Mittelwert		4.44	4.27	3.85	3.76	4.41	4.26	4.05	3.97
(Standardabweichung)		(0.75)	(0.84)	(1.04)	(0.98)	(0.68)	(0.75)	(0.82)	(0.87)
Median		4.75	4.75	4.00	3.75	4.75	4.50	4.00	4.12

Die Analyse zeigt keine signifikanten Dreiwege- und Zweiwege-Interaktionen. Es liegt ein signifikanter Haupteffekt der Zeit vor (vgl. Tabelle 68).

Tabelle 68: Ergebnisse der robusten Varianzanalyse des Interesses (SDT) beim Lernen mit HyperDocs im Schulformvergleich in der Mittelstufe. Die *Q*-Werte stammen aus dem non-bootstrap-Verfahren.

<b>Effekt</b>	<b><i>Q</i></b>	<b><i>p</i></b>
Geschlecht	0.12	0.736
Schulform	1.38	0.232
Messzeitpunkt	29.38**	< 0.001
Geschlecht*Schulform	0.14	0.700
Geschlecht*Messzeitpunkt	1.51	0.073
Schulform*Messzeitpunkt	0.60	0.445
Geschlecht*Schulform*Messzeitpunkt	0.20	0.836

In Abbildung 32 ist ein Einfluss der Einsatzdauer deutlich erkennbar.

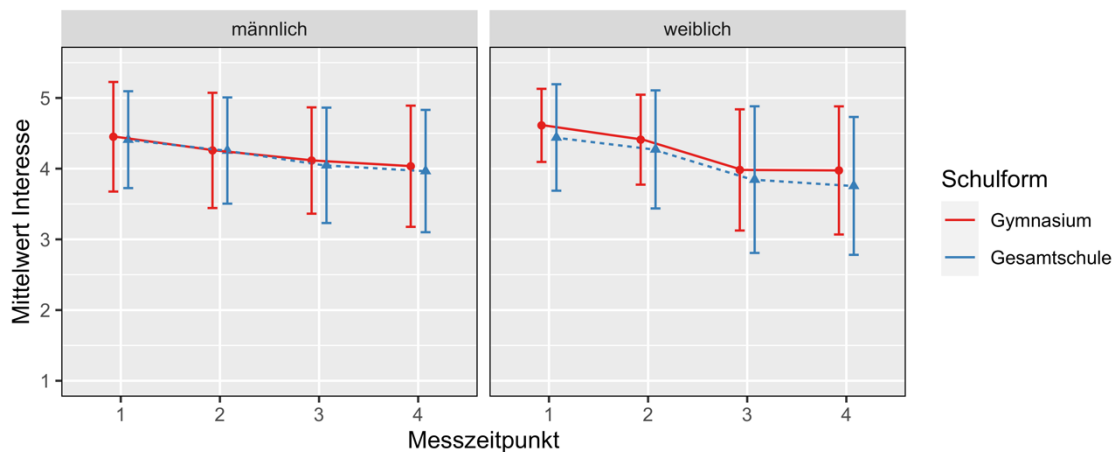


Abbildung 32: Mittelwerte mit Standardabweichung des Interesses (SDT) beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.

Es handelt sich um eine kontinuierliche Abnahme des Interesses, die bei den Schülerinnen zunächst stärker ausfällt. Die Abnahme verläuft in beiden Schulformen nahezu parallel, wobei bei den Schülerinnen der Gesamtschule die Abnahme von der dritten zur vierten Stunde etwas stärker ist. Die genauen signifikanten Unterschiede zwischen den einzelnen Stunden sind Anhang 9.3.4 zu entnehmen. Die Effektstärke beträgt  $\delta_t = 0.39$  und ist damit als schwach bis mittelstark einzustufen. Die Effektstärke zwischen der ersten und der vierten Stunde beträgt  $\delta_t = 0.55$  und ist damit mittelstark.

*Fachinteresse Chemie und Sachinteresse Naturwissenschaften*

Wie auch zuvor erfolgt die Auswertung des Fachinteresses Chemie mit einer robusten gemischten Varianzanalyse. In der Stichprobe sind vier Ausreißer vorhanden. In Tabelle 69 ist die deskriptive Statistik aufgeführt.

*Tabelle 69: Deskriptive Statistik des Fachinteresses Chemie im Schulformvergleich in der Mittelstufe.*

Messzeitpunkt	<b>Gymnasium</b>			
	weiblich		männlich	
	1	2	1	2
Stichprobenumfang	101		94	
Mittelwert (Standardabweichung)	3.16 (1.01)	3.21 (1.15)	3.43 (0.84)	3.60 (0.86)
Median	3.00	3.00	3.50	3.50
Messzeitpunkt	<b>Gesamtschule</b>			
	weiblich		männlich	
	1	2	1	2
Stichprobenumfang	48		56	
Mittelwert (Standardabweichung)	3.75 (0.91)	4.04 (0.79)	3.47 (1.01)	3.38 (1.26)
Median	4.00	3.00	3.50	3.50



Die gemischte Varianzanalyse zeigt keine signifikanten Interaktionen und drei signifikante Haupteffekte (vgl. Tabelle 70).

Tabelle 70: Ergebnisse der robusten Varianzanalyse des Fachinteresses Chemie im Schulformvergleich in der Mittelstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap-Verfahren.

Effekt	$Q$	$p$
Geschlecht	7.83**	0.005
Schulform	6.29*	0.017
Messzeitpunkt	5.80*	0.018
Geschlecht*Sschulform	0.39	0.501
Geschlecht*Messzeitpunkt	2.54	0.115
Sschulform*Messzeitpunkt	0.01	0.942
Geschlecht*Sschulform*Messzeitpunkt	0.42	0.509

In Abbildung 33 sind die Mittelwerte der jeweiligen Teilgruppen gezeigt. Es bestehen deutlich sichtbare Unterschiede zwischen den Schulformen. Demnach ist das Fachinteresse der Schüler:innen der Gesamtschule höher als derer des Gymnasiums. Bei beiden Schulformen liegt das Interesse an Chemie der weiblichen Lernenden unter dem der männlichen. Bezüglich der Entwicklung nimmt bei allen Teilgruppen mit Ausnahme der Schülerinnen der Gesamtschule das Fachinteresse zu. Bei den Jungen fällt dieser Effekt stärker aus.

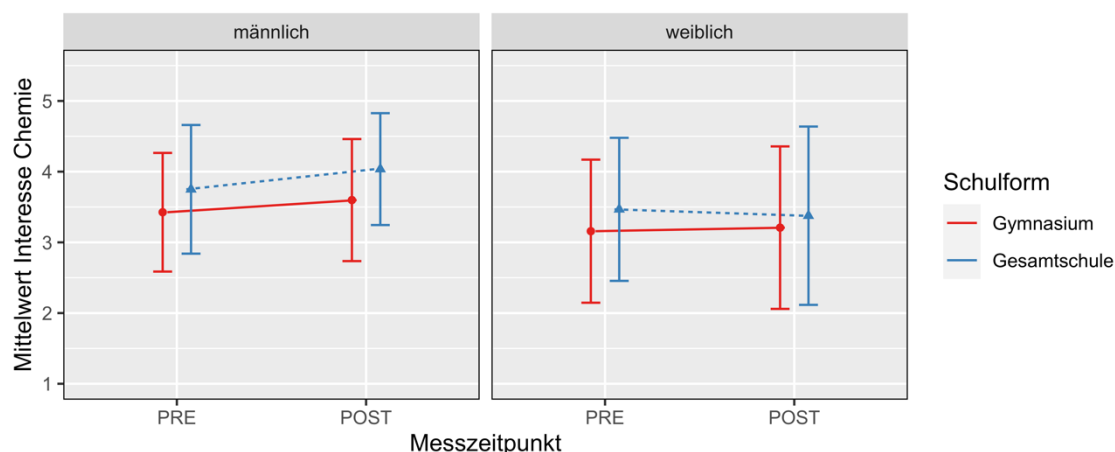


Abbildung 33: Mittelwerte mit Standardabweichung des Fachinteresses Chemie, aufgeteilt nach Geschlecht und Schulform im Pre- und Posttest in der Mittelstufe.

Die Analyse des Haupteffektes „Geschlecht“ ( $M_w = 3.26 \pm 1.03$ ,  $M_m = 3.65 \pm 0.80$ ) zeigt einen signifikanten Unterschied zwischen den Geschlechtern mit einer mittleren

Effektstärke (*robust t*(170.70) = 2.97,  $p = 0.003$ ,  $\xi = 0.29$ ). Der Haupteffekt „Schulform“ ( $M_{Gym} = 3.34 \pm 0.92$ ,  $M_{IGS} = 3.68 \pm 0.93$ ) ist ebenfalls signifikant mit einem schwachen bis mittleren Effekt (*robust t*(114.59) = 2.65,  $p = 0.009$ ,  $\xi = 0.24$ ). Das gleiche gilt für den Effekt „Messzeitpunkt“ ( $M_{Pre} = 3.40 \pm 0.96$ ,  $M_{Post} = 3.51 \pm 1.06$ ), wobei die Effektstärke gering ist (*robust t*(180) = 3.52,  $p < 0.001$ ,  $\xi = 0.11$ ).

Die Statistik zum Sachinteresse an Naturwissenschaften ist Tabelle 71 zu entnehmen.

Tabelle 71: Deskriptive Statistik des Sachinteresses Naturwissenschaften im Schulformvergleich in der Mittelstufe

	<b>Gymnasium</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt				
Stichprobenumfang	101		94	
Mittelwert (Standardabweichung)	3.13 (1.03)	3.16 (1.15)	3.31 (0.94)	3.44 (0.96)
Median	3.00	3.00	3.50	3.50
	<b>Gesamtschule</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt				
Stichprobenumfang	48		56	
Mittelwert (Standardabweichung)	2.91 (1.18)	2.75 (1.39)	3.44 (0.87)	3.76 (0.92)
Median	3.00	2.50	3.50	3.50

Die robuste mixed ANOVA zeigt eine signifikante Zweifache-Interaktion „Geschlecht\*Messzeitpunkt“ (vgl. Tabelle 72).

Tabelle 72: Ergebnisse der robusten Varianzanalyse des Sachinteresses an Naturwissenschaften im Schulformvergleich in der Mittelstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.

Effekt	Q	p
Geschlecht	9.77	< 0.001
Schulform	0.52	0.441
Messzeitpunkt	1.38	0.217
Geschlecht*Sschulform	3.45	0.068
Geschlecht*Messzeitpunkt	7.48	0.010
Sschulform*Messzeitpunkt	0.44	0.482
Geschlecht*Sschulform*Messzeitpunkt	2.57	0.110

In Abbildung 34 sind die Mittelwerte der Teilgruppen aufgetragen. Hierbei nimmt das Sachinteresse erneut bei den Schülern zu, wobei die Lernenden der Gesamtschule einen größeren Interessenszuwachs verzeichnen. Bei den Schülerinnen des Gymnasiums bleibt der Wert nahezu konstant, während er bei den weiblichen Lernenden der Gesamtschule deutlich abnimmt. Insgesamt liegen die Werte der Schülerinnen unter denen der Schüler.

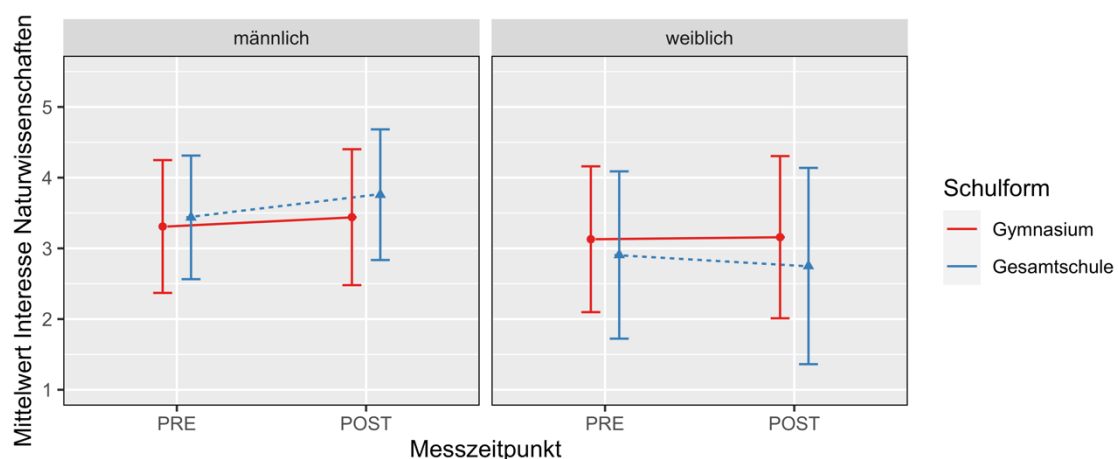


Abbildung 34: Mittelwerte mit Standardabweichung des Sachinteresses Naturwissenschaften, aufgeteilt nach Geschlecht und Schulform im Pre- und Posttest in der Mittelstufe.

Die Ergebnisse der Zweifache-Interaktion werden nachfolgend dargestellt. Demnach unterscheiden sich die Schüler, unabhängig von der Gruppe, signifikant im Pre- und Post-Test mit einer schwachen Effektstärke ( $robust\ t(89) = 3.21, p = 0.002, \xi = 0.16$ ,

$M_{Pre} = 3.36 \pm 0.92$ ,  $M_{Post} = 3.56 \pm 0.96$ ). Bei den weiblichen Lernenden liegt kein signifikanter Unterschied vor (*robust t*(90) = 0.67,  $p = 0.506$ ,  $\hat{\xi} = 0.03$ ,  $M_{Pre} = 3.06 \pm 1.08$ ,  $M_{Post} = 3.03 \pm 1.24$ ). Während sich die Geschlechter zum ersten Messzeitpunkt (*robust t*(168.31) = 1.74,  $p = 0.083$ ,  $\hat{\xi} = 0.17$ ) nicht signifikant voneinander unterscheiden, liegt zum zweiten Zeitpunkt ein signifikanter Unterschied zwischen den Geschlechtern vor (*robust t*(169.28) = 3.92,  $p < 0.001$ ,  $\hat{\xi} = 0.32$ ).

### **Wahrgenommene Wahlfreiheit: Geschlecht, Schulform, Messzeitpunkt**

Aufgrund der Verletzung der Sphärizität ( $p < 0.001$ ) und der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ) und dem Vorhandensein von zwölf Ausreißern wurde eine robuste gemischte Varianzanalyse durchgeführt. Die deskriptive Statistik kann Tabelle 73 entnommen werden.

*Tabelle 73: Deskriptive Statistik der wahrgenommenen Wahlfreiheit beim Lernen mit HyperDocs im Schulformvergleich in der Mittelstufe.*

		<b>Gymnasium</b>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		109				99			
Mittelwert		4.17	4.08	3.69	3.69	4.1	4.07	3.84	3.75
(Standardabweichung)		(0.67)	(0.76)	(0.95)	(0.92)	(0.78)	(0.68)	(0.81)	(0.96)
Median		4.25	4.25	3.75	3.75	4.25	4.00	4.00	3.75
		<b>Gesamtschule</b>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		53				64			
Mittelwert		3.98	3.88	3.53	3.69	3.9	3.74	3.63	3.65
(Standardabweichung)		(0.71)	(0.86)	(0.88)	(0.85)	(0.80)	(0.80)	(0.79)	(0.79)
Median		4.00	3.75	3.50	3.50	4.00	3.50	3.50	3.50

Wie bereits beim Interesse (SDT), zeigt sich bei der wahrgenommenen Wahlfreiheit (Autonomie) beim Lernen mit HyperDocs keine signifikante Zweifache- oder Dreifache-Interaktion. Es sind zwei signifikante Haupteffekte vorhanden, die die Schulform und den Messzeitpunkt betreffen (vgl. Tabelle 74).

Tabelle 74: Ergebnisse der robusten Varianzanalyse zur wahrgenommenen Wahlfreiheit beim Lernen mit Arbeitsblättern in der Mittelstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap-Verfahren.

Effekt	$Q$	$p$
Geschlecht	0.12	0.721
Schulform	8.01**	0.005
Messzeitpunkt	19.19**	< 0.001
Geschlecht*Schulform	0.21	0.662
Geschlecht*Messzeitpunkt	1.41	0.081
Schulform*Messzeitpunkt	0.82	0.259
Geschlecht*Schulform*Messzeitpunkt	0.42	0.568

Nach dem Haupteffekt „Schulform“ unterscheiden sich die Schüler:innen des Gymnasiums und der Gesamtschule in Bezug auf ihre wahrgenommene Wahlfreiheit ( $M_{Gym} = 3.92 \pm 0.65$ ,  $M_{IGS} = 3.75 \pm 0.67$ ). Der robuste T-Test ( $robust\ t(135.46) = 2.62$ ,  $p = 0.010$ ,  $\hat{\xi} = 0.22$ ) zeigt einen schwachen bis mittelstarken Effekt. Demnach fühlen sich Gymnasiast:innen beim Lernen mit HyperDocs autonomer. Insgesamt nimmt bei beiden Gruppen, unabhängig vom Geschlecht, die wahrgenommene Wahlfreiheit mit der Einsatzdauer ab, was durch den signifikanten Haupteffekt „Messzeitpunkt“ verdeutlicht wird (vgl. Abbildung 35).

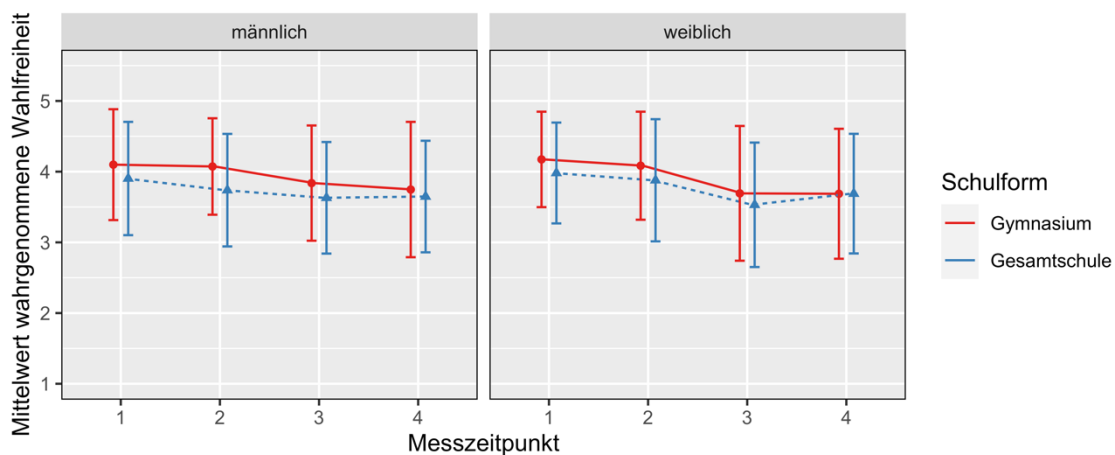


Abbildung 35: Mittelwerte mit Standardabweichung der wahrgenommenen Wahlfreiheit beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.

Die Effektstärke ist mit  $\delta_t = 0.28$  als schwach einzustufen. Die Effektstärke des Unterschieds zwischen der ersten und letzten Stunde ist mit  $\delta_t = 0.42$  tendenziell mittelstark.

**Kompetenzerleben: Geschlecht, Schulform, Messzeitpunkt**

Die Prüfung der Voraussetzungen der gemischten Varianzanalyse zeigte eine Verletzung der Sphärizität ( $p < 0.001$ ), Normalverteilung (*QQ-Plot, Shapiro-Wilk-Test*:  $p < 0.001$ ), der Varianzhomogenität in der ersten Stunde ( $p = 0.022$ ) und der Homogenität der Kovarianzmatrizen des Faktors „Geschlecht“ ( $p < 0.046$ ). Es sind zudem 13 Ausreißer vorhanden, wovon einer als extrem eingestuft wird. In Tabelle 75 ist die deskriptive Statistik der Stichprobe abgebildet.

*Tabelle 75: Deskriptive Statistik des Kompetenzerlebens beim Lernen mit HyperDocs im Schulformvergleich in der Mittelstufe.*

		<b>Gymnasium</b>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		109				99			
Mittelwert		4.35	4.23	3.82	3.77	4.31	4.28	4.02	3.93
(Standardabweichung)		(0.56)	(0.65)	(0.76)	(0.79)	(0.59)	(0.69)	(0.70)	(0.76)
Median		4.50	4.25	4.00	3.75	4.50	4.50	4.00	4.00
		<b>Gesamtschule</b>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		53				64			
Mittelwert		4.20	4.08	3.76	3.64	4.31	4.26	3.98	3.94
(Standardabweichung)		(0.76)	(0.75)	(0.72)	(0.74)	(0.57)	(0.66)	(0.71)	(0.71)
Median		4,50	4.25	3.75	3.50	4.50	4.25	4.00	4.00

Die robuste ANOVA zeigte keine signifikanten Dreiwege- und Zweiwege-Interaktionen. Die Haupteffekte „Geschlecht“ und „Messzeitpunkt“ sind signifikant (vgl. Tabelle 76).

Tabelle 76: Ergebnisse der robusten Varianzanalyse des Kompetenzerlebens beim Lernen mit HyperDocs im Schulformvergleich in der Mittelstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap-Verfahren.

Effekt	$Q$	$p$
Geschlecht	4.55*	0.032
Schulform	0.79	0.379
Messzeitpunkt	29.60**	< 0.001
Geschlecht*Schulform	0.42	0.492
Geschlecht*Messzeitpunkt	1.49	0.077
Schulform*Messzeitpunkt	0.20	0.825
Geschlecht*Schulform*Messzeitpunkt	0.20	0.833

Abbildung 36 verdeutlicht die Haupteffekte. Demnach nimmt das Kompetenzerleben beim Lernen mit HyperDocs in beiden Gruppen und Geschlechtern über die Einsatzdauer von HyperDocs ab. Bei den weiblichen Lernenden fällt das Kompetenzerleben ab der dritten Stunde stärker ab als bei den Schülern.

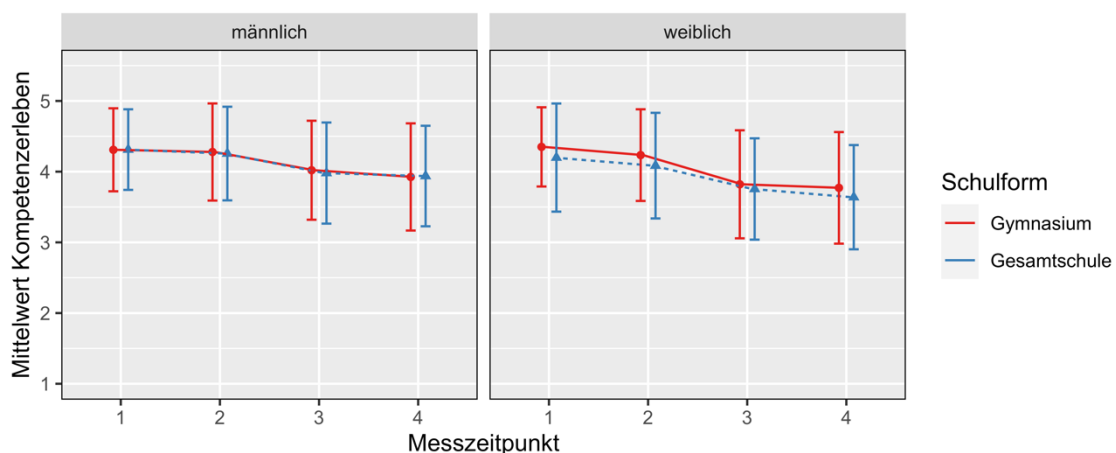


Abbildung 36: Mittelwerte mit Standardabweichung des Kompetenzerlebens beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.

Dadurch lässt sich der signifikante Haupteffekt „Geschlecht“ erklären. Der robuste t-Test zeigt kein signifikantes Ergebnis und lediglich eine schwache Effektstärke (*robust*  $t(192.24) = 1.78, p = 0.076, \hat{\xi} = 0.14$ ). Die fehlende Signifikanz des t-Tests gegenüber der robusten gemischten Varianzanalyse lässt sich womöglich dadurch begründen, dass bei der ANOVA Bootstrapping angewandt wird. Die Effektstärke des Haupteffektes „Messzeitpunkt“ beträgt  $\delta_t = 0.35$  und für den Vergleich zwischen der ersten und vierten Stunde  $\delta_t = 0.57$ . Die Effekte sind damit schwach bis mittelstark.

**Druck: Geschlecht, Schulform, Messzeitpunkt**

Wie in den vorherigen Analysen erfolgt die Auswertung des Drucks deskriptiv (vgl. Tabelle 77).

*Tabelle 77: Deskriptive Statistik des Drucks beim Lernen mit HyperDocs im Schulformvergleich in der Mittelstufe.*

		<i>Gymnasium</i>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		109				99			
Mittelwert		4.33	4.31	4.06	4.01	4.45	4.43	4.3	4.22
(Standardabweichung)		(0.62)	(0.63)	(0.70)	(0.78)	(0.57)	(0.60)	(0.68)	(0.68)
Median		4.25	4.50	4.00	4.00	4.50	4.50	4.50	4.25
		<i>Gesamtschule</i>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		53				64			
Mittelwert		4.39	4.29	4.01	3.8	4.37	4.18	4.2	3.99
(Standardabweichung)		(0.64)	(0.64)	(0.83)	(0.84)	(0.69)	(0.73)	(0.78)	(0.85)
Median		4.50	4.50	4.00	3.75	4.50	4.25	4.25	4.00

In Abbildung 37 zeigt sich im Verlauf der Unterrichtsreihe eine Zunahme des Drucks beim Lernen mit HyperDocs, der in der Gesamtschule insbesondere in der vierten Stunde stärker auftritt. Zudem empfinden die Schülerinnen mehr Druck als die Schüler. Charakteristisch ist bei den Schülerinnen der Einbruch in der dritten Stunde, der zuvor bereits bei anderen Variablen beobachtet werden konnte.



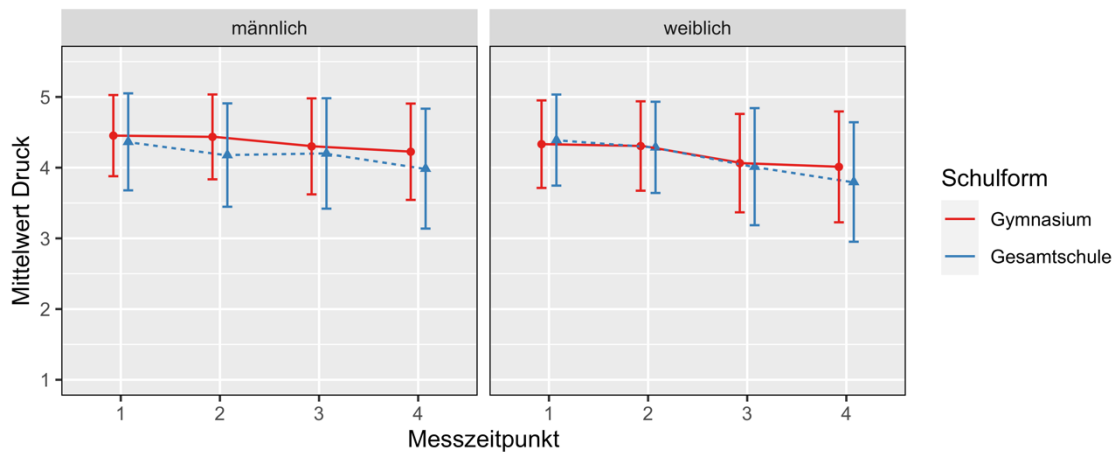


Abbildung 37: Mittelwerte mit Standardabweichung des Drucks beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe. Hohe Werte entsprechen einem geringeren Druck.

### 5.3 Usability

Im folgenden Kapitel wird die Usability von HyperDocSystems näher betrachtet. Dazu erfolgt die Auswertung zunächst für die Mittelstufe und für die Oberstufe. Danach werden Zusammenhänge zu weiteren Variablen untersucht, um mögliche Ergebnisse und Veränderungen in der Usability begründen zu können. Für die Auswertung werden lediglich Schüler:innen betrachtet, die an allen vier Stunden teilgenommen haben.

#### 5.3.1 Analyse im zeitlichen Verlauf Mittelstufe nach Geschlecht

Da bei einer explorativen Analyse größere Unterschiede zwischen den Schulformen im Nachtest der Usability festgestellt wurden, erfolgt die Auswertung der Einsatzdauer und des Geschlechts unter Berücksichtigung des dritten Faktors „Schulform“. Die Auswertung findet daher unter Zuhilfenahme einer gemischten Varianzanalyse mit den Faktoren „Geschlecht“, „Messzeitpunkt“ und „Schulform“ statt. Bei jeder Variablen werden zunächst die Voraussetzungen dieser Methode geprüft. Die Auswertung der Daten erfolgt analog zu Kapitel 5.2.

#### **Nützlichkeit: Geschlecht\*Schulform\*Messzeitpunkt**

Die Prüfung der Voraussetzungen ergibt eine Verletzung der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ) und der Varianzhomogenität im Nachtest. Zudem sind zehn Ausreißer vorhanden, weshalb eine robuste mixed ANOVA eingesetzt wurde.

Aufgrund der geringen Teilstichproben wurde zusätzlich Bootstrapping angewandt. Die deskriptive Statistik ist in Tabelle 78 aufgeführt.

*Tabelle 78: Deskriptive Statistik der Nützlichkeit (Usability) im Schulformvergleich in der Mittelstufe.*

	<b>Gymnasium</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt	1	2	1	2
Stichprobenumfang	101		96	
Mittelwert (Standardabweichung)	4.21 (0.81)	3.81 (1.09)	4.22 (1.04)	4.02 (1.00)
Median	4.33	4.00	4.67	4.00
	<b>Gesamtschule</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt	1	2	1	2
Stichprobenumfang	48		60	
Mittelwert (Standardabweichung)	4.18 (0.90)	3.56 (1.28)	4.13 (0.87)	3.89 (0.95)
Median	4.33	3.56	4.30	4.00

Die Analyse zeigt lediglich einen signifikanten Haupteffekt „Messzeitpunkt“ (vgl. Tabelle 79).

*Tabelle 79: Ergebnisse der robusten Varianzanalyse der Nützlichkeit im Schulformvergleich in der Mittelstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.*

<b>Effekt</b>	<b>Q</b>	<b>p</b>
Geschlecht	1.06	0.359
Schulform	1.52	0.250
Messzeitpunkt	31.30**	< 0.001
Geschlecht*Schulform	0.20	0.689
Geschlecht*Messzeitpunkt	2.84	0.075
Schulform*Messzeitpunkt	0.76	0.374
Geschlecht*Schulform*Messzeitpunkt	0.59	0.477

In Abbildung 38 zeigt sich zum einen deutlich der Haupteffekt „Messzeitpunkt“, da alle Geschlechter in allen Gruppen die Nützlichkeit im Nachtest schlechter bewerten. Dieser Effekt ist bei den Schülerinnen stärker ausgeprägt.

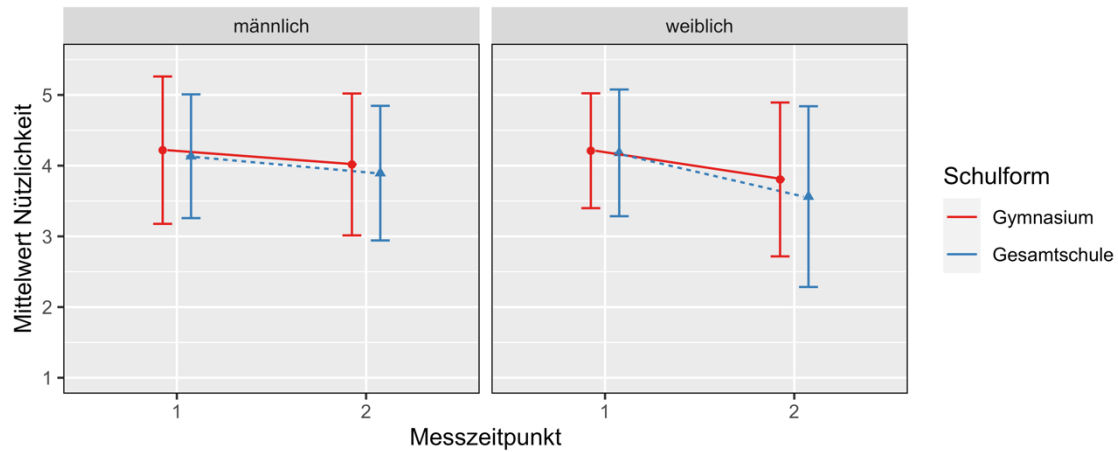


Abbildung 38: Mittelwerte mit Standardabweichung der Nützlichkeit (Usability) der HyperDocs, aufgeteilt nach Geschlecht und Schulform nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Mittelstufe.

Die Durchführung eines robusten abhängigen t-Tests zeigt einen signifikanten Unterschied zwischen beiden Messzeitpunkten ( $M_{MZP1} = 4.19 \pm 0.91$ ;  $M_{MZP2} = 3.85 \pm 1.08$ ) mit einer schwachen bis mittleren Effektstärke ( $\text{robust } t(182) = 6.34, p < 0.001, \hat{\xi} = 0.24$ ).

**Zufriedenheit: Geschlecht\*Schulform\*Messzeitpunkt**

Aufgrund der Verletzung der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ) und dem Vorhandensein von 20 Ausreißern, einer davon extrem, wurde eine robuste gemischte Varianzanalyse durchgeführt. Die deskriptive Statistik ist Tabelle 80 zu entnehmen.

*Tabelle 80: Deskriptive Statistik der Zufriedenheit im Schulformvergleich in der Mittelstufe.*

	<b>Gymnasium</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt				
Stichprobenumfang	101		96	
Mittelwert (Standardabweichung)	4.63 (0.46)	4.29 (0.71)	4.47 (0.79)	4.33 (0.79)
Median	4.50	4.50	5.00	4.50
	<b>Gesamtschule</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt				
Stichprobenumfang	48		60	
Mittelwert (Standardabweichung)	4.57 (0.61)	4.07 (1.03)	4.51 (0.59)	4.32 (0.81)
Median	5.00	4.50	4.50	4.50

In Abbildung 39 zeigt sich, wie auch bei der Nützlichkeit, eine leichte Abnahme über den Messzeitpunkt für alle Gruppen und für jede Schulform, wobei erneut bei den Schülerinnen in der Gesamtschule der Abfall deutlicher ausfällt. Die Werte liegen jedoch selbst zum zweiten Messzeitpunkt noch auf einem hohen Niveau von über  $M = 4.00$ . Auffallend ist die hohe Standardabweichung bei den Schülerinnen der Gesamtschule zu diesem Messzeitpunkt, was auf große Unterschiede in der Bewertung der Usability schließen lässt.

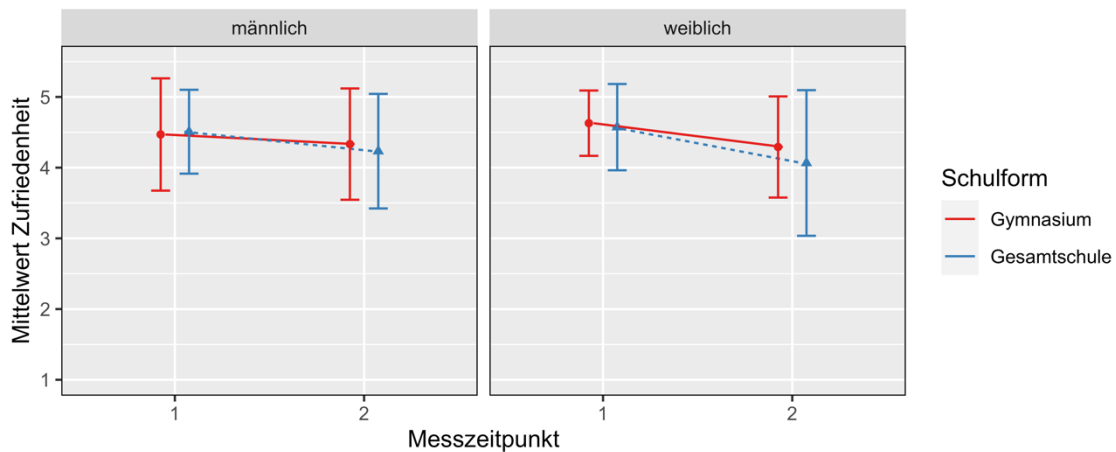


Abbildung 39: Mittelwerte mit Standardabweichung der Zufriedenheit (Usability) mit Hyper-Docs, aufgeteilt nach Geschlecht und Schulform nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Mittelstufe.

Eine genaue Analyse zeigt ebenfalls einen signifikanten Haupteffekt „Messzeitpunkt“. Die vollständigen Ergebnisse sind Tabelle 81 zu entnehmen.

Tabelle 81: Ergebnisse der robusten Varianzanalyse der Zufriedenheit im Schulformvergleich in der Mittelstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap-Verfahren.

Effekt	$Q$	$p$
Geschlecht	0.01	0.940
Schulform	0.65	0.451
Messzeitpunkt	36.54**	< 0.001
Geschlecht*Sschulform	0.10	0.755
Geschlecht*Messzeitpunkt	2.45	0.124
Sschulform*Messzeitpunkt	1.31	0.267
Geschlecht*Sschulform*Messzeitpunkt	0.01	0.922

Daraufhin wurde der Haupteffekt „Messzeitpunkt“ analysiert. Dabei zeigt sich zwischen den Messzeitpunkten ( $M_{MZP1} = 4.55 \pm 0.63$ ,  $M_{MZP2} = 4.26 \pm 0.81$ ) ein signifikanter Unterschied anhand eines robusten gepaarten t-Tests ( $robust\ t(182) = 5.98$ ,  $p < 0.001$ ,  $\xi = 0.30$ ). Der Effekt ist als mittelstark einzuordnen.

**Einfachheit der Nutzung: Geschlecht\*Schulform\*Messzeitpunkt**

Die Prüfung der Voraussetzungen zeigt eine Verletzung der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ), der Varianzhomogenität für den zweiten Messzeitpunkt ( $p = 0.005$ ) und eine Verletzung der Homogenität der Kovarianzmatrizen für den Faktor „Schulform“ ( $p = 0.002$ ). Zusätzlich sind zehn Ausreißer vorhanden, einer davon extrem. Aus diesem Grund wurde eine robuste gemischte Varianzanalyse mit Bootstrapping durchgeführt. Die deskriptive Statistik ist in Tabelle 82 aufgeführt.

*Tabelle 82: Deskriptive Statistik der Einfachheit der Nutzung im Schulformvergleich in der Mittelstufe.*

	<i>Gymnasium</i>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt				
Stichprobenumfang	101		96	
Mittelwert (Standardabweichung)	4.64 (0.49)	4.50 (0.58)	4.68 (0.56)	4.48 (0.63)
Median	4.67	4.67	5.00	4.67
	<i>Gesamtschule</i>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt				
Stichprobenumfang	48		60	
Mittelwert (Standardabweichung)	4.51 (0.74)	4.20 (0.88)	4.57 (0.47)	4.42 (0.62)
Median	5.00	4.50	4.67	4.67

Die mixed ANOVA zeigt keine signifikante Interaktion und einen signifikanten Haupteffekt „Messzeitpunkt“ (vgl. Tabelle 83).

Tabelle 83: Ergebnisse der robusten Varianzanalyse der Einfachheit der Nutzung im Schulformvergleich in der Mittelstufe. Die Q-Werte stammen aus dem non-bootstrap-Verfahren.

Effekt	Q	p
Geschlecht	0.22	0.664
Schulform	2.35	0.115
Messzeitpunkt	22.05**	< 0.001
Geschlecht*Schulform	0.02	0.914
Geschlecht*Messzeitpunkt	0.73	0.377
Schulform*Messzeitpunkt	0.31	0.529
Geschlecht*Schulform*Messzeitpunkt	3.52	0.067

In Abbildung 40 ist die Abnahme der Werte der Variable deutlich in beiden Schulformen zu erkennen. Bei den Schülerinnen der Gesamtschule sinken sie über den Erhebungszeitraum am deutlichsten ab, während sie beim Gymnasium lediglich leicht abnehmen. Beide Geschlechter stufen die Einfachheit der Nutzung hoch ein.

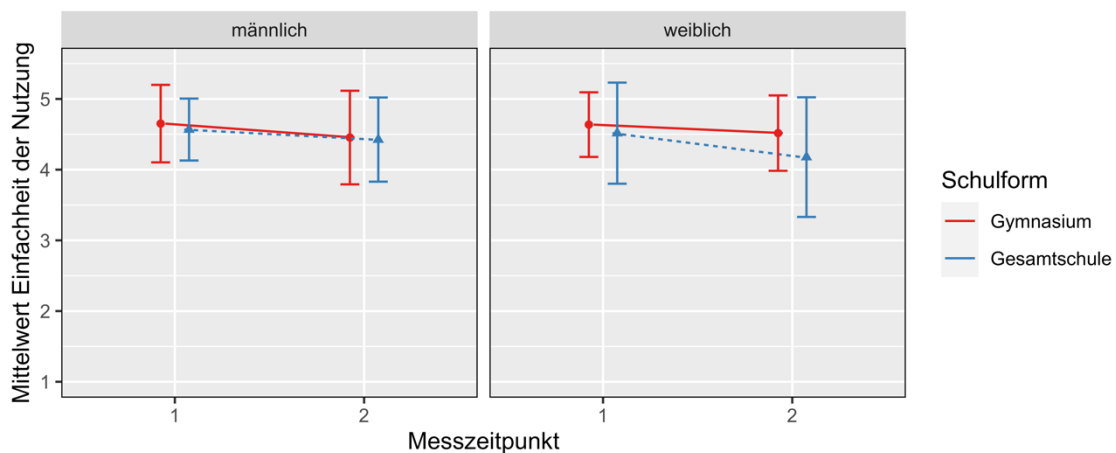


Abbildung 40: Mittelwerte mit Standardabweichung der Einfachheit der Nutzung (Usability) der HyperDocs, aufgeteilt nach Geschlecht und Schulform nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Mittelstufe.

Um die Effektstärke zu ermitteln, wurde ein robuster t-Test durchgeführt (*robust*  $t(182) = 5.62, p < 0.001, \hat{\xi} = 0.22, M_{MZP1} = 4.61 \pm 0.53, M_{MZP2} = 4.43 \pm 0.65$ ). Der Effekt ist als schwach bis mittelstark einzustufen.

### **Einfachheit des Lernens der Nutzung: Geschlecht\*Schulform\*Messzeitpunkt**

Zunächst wurden wieder die Voraussetzungen für eine gemischte Varianzanalyse geprüft. Dabei zeigte sich eine Verletzung der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ) und der Homogenität der Kovarianzmatrizen für den Faktor „Schulform“. Zudem sind 31 Ausreißer, sieben davon extrem, in der Stichprobe vorhanden. Aus diesem Grund wurde eine robuste mixed ANOVA mit Bootstrapping eingesetzt. Die deskriptive Statistik ist in Tabelle 84 aufgeführt.

*Tabelle 84: Deskriptive Statistik der Einfachheit des Lernens der Nutzung im Schulformvergleich in der Mittelstufe.*

	<i>Gymnasium</i>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt	1	2	1	2
Stichprobenumfang	101		96	
Mittelwert (Standardabweichung)	4.72 (0.39)	4.64 (0.50)	4.72 (0.47)	4.65 (0.50)
Median	5.00	5.00	5.00	5.00
	<i>Gesamtschule</i>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt	1	2	1	2
Stichprobenumfang	48		60	
Mittelwert (Standardabweichung)	4.61 (0.54)	4.44 (0.71)	4.66 (0.50)	4.48 (0.67)
Median	5.00	4.67	5.00	5.00

Die gemischte Varianzanalyse zeigt einen signifikanten Haupteffekt „Messzeitpunkt“. Die anderen Effekte sind nicht signifikant (vgl. Tabelle 85), wobei sich der Haupteffekt „Schulform“ und die Zweifache-Interaktion „Schulform\*Messzeitpunkt“ dem Signifikanzniveau annähern.



Tabelle 85: Ergebnisse der robusten Varianzanalyse Einfachheit des Lernens der Nutzung im Schulformvergleich in der Mittelstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap-Verfahren.

Effekt	$Q$	$p$
Geschlecht	0.49	0.496
Schulform	2.66	0.112
Messzeitpunkt	10.19**	0.005
Geschlecht*Schulform	0.03	0.831
Geschlecht*Messzeitpunkt	0.28	0.593
Schulform*Messzeitpunkt	1.74	0.165
Geschlecht*Schulform*Messzeitpunkt	0.24	0.603

Der Haupteffekt zeigt sich in Abbildung 41. Demnach nimmt bei beiden Geschlechtern und Schulformen die Einfachheit des Lernens der Nutzung über die Messzeitpunkte ab. Tendenziell nimmt sie bei Schüler:innen der Gesamtschule stärker ab. Der robuste gepaarte t-Test zeigt zwischen den Messzeitpunkten ( $M_{MZP1} = 4.69 \pm 0.46$ ,  $M_{MZP2} = 4.58 \pm 0.58$ ) einen signifikanten Unterschied mit einer schwachen Effektstärke ( $\text{robust } t(182) = 2.61$ ,  $p < 0.001$ ,  $\xi = 0.13$ ).

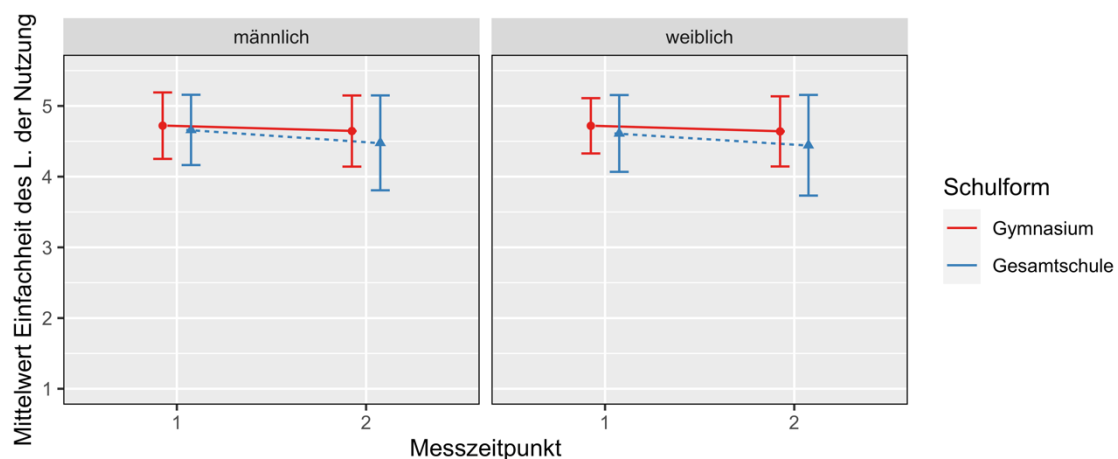


Abbildung 41: Mittelwerte mit Standardabweichung der Einfachheit des Lernens der Nutzung (Usability) der HyperDocs, aufgeteilt nach Geschlecht und Schulform nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Mittelstufe.

### 5.3.2 Analyse im zeitlichen Verlauf Oberstufe

Da für die Oberstufe nicht ausreichend Datenpunkte für die Gesamtschule bestehen, findet die Analyse lediglich in Form einer mixed ANOVA unter Beachtung des Geschlechts und des Messzeitpunktes Anwendung. Es werden erneut nur Lernende betrachtet, die in allen vier Stunden anwesend waren.

#### **Nützlichkeit: Geschlecht\*Messzeitpunkt**

Aufgrund der Verletzung der Normalverteilung (*QQ-Plot, Shapiro-Wilk-Test:  $p < 0.001$* ) und 14 Ausreißern, wovon zwei extrem sind, wurde eine robuste gemischte Varianzanalyse mit Bootstrapping durchgeführt. Tabelle 86 zeigt die deskriptive Statistik.

*Tabelle 86: Deskriptive Statistik der Nützlichkeit in der Oberstufe.*

Messzeitpunkt	Gymnasium			
	weiblich		männlich	
	1	2	1	2
Stichprobenumfang	43		50	
Mittelwert (Standardabweichung)	3.85 (0.94)	3.11 (0.90)	3.79 (1.06)	3.32 (0.91)
Median	4.00	3.00	4.00	3.33

Die Analyse zeigt keine signifikante Interaktion, jedoch einen signifikanten Haupteffekt „Messzeitpunkt“ (vgl. Tabelle 87).

*Tabelle 87: Teststatistik der robusten gemischten Varianzanalyse bezüglich der Nützlichkeit in der Oberstufe.*

Effekt	WJ	df <sub>N</sub>	df <sub>D</sub>	p
Geschlecht	0.40	1	47.79	0.551
Messzeitpunkt	42.96**	1	54.15	< 0.001
Geschlecht*Messzeitpunkt	0.61	1	54.15	0.424

Der signifikante Haupteffekt ist in Abbildung 42 erkennbar. Bei beiden Geschlechtern nimmt die Nützlichkeit über die Einsatzdauer ab. Dabei zeigt sich, wie auch schon in der Mittelstufe, eine größere Abnahme bei den Schülerinnen. Die Oberstufe stuft die Nützlichkeit geringer ein als die Mittelstufe. Die Abnahme ist zudem stärker.

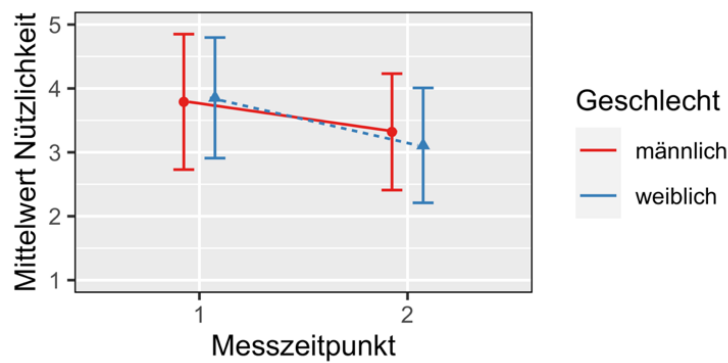


Abbildung 42: Mittelwerte mit Standardabweichung der Nützlichkeit (Usability) der HyperDocs, aufgeteilt nach Geschlecht nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Oberstufe.

Der robuste gepaarte t-Test zeigt zwischen den Messzeitpunkten ( $M_{MZP1} = 3.82 \pm 1.00$ ,  $M_{MZP2} = 3.22 \pm 0.91$ ) eine signifikante Abnahme mit einer hohen Effektstärke (*robust*  $t(56) = 6.48$ ,  $p < 0.001$ ,  $\xi = 0.47$ ).

### **Zufriedenheit: Geschlecht\*Messzeitpunkt**

Die Prüfung der Voraussetzungen zeigt eine Verletzung der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ) und der Varianzhomogenität in der ersten Stunde. In der Stichprobe ist ein Ausreißer vorhanden. Es wurde erneut eine robuste gemischte Varianzanalyse mit Bootstrapping durchgeführt. Die deskriptive Statistik ist in Tabelle 88 gezeigt.

Tabelle 88: Deskriptive Statistik der Zufriedenheit in der Oberstufe.

Messzeitpunkt	Gymnasium			
	weiblich		männlich	
	1	2	1	2
Stichprobenumfang	43		50	
Mittelwert (Standardabweichung)	4.45 (0.47)	3.99 (0.81)	4.29 (0.63)	3.88 (0.78)
Median	4.50	4.00	4.50	4.00

Die Varianzanalyse zeigt erneut einen signifikanten Haupteffekt „Messzeitpunkt“ (vgl. Tabelle 89).

Tabelle 89: Teststatistik der robusten gemischten Varianzanalyse bezüglich der Zufriedenheit in der Oberstufe.

Effekt	WJ	df <sub>N</sub>	df <sub>D</sub>	p
Geschlecht	0.86	1	54.00	0.360
Messzeitpunkt	41.37**	1	54.07	< 0.001
Geschlecht*Messzeitpunkt	0.03	1	54.07	0.868

In Abbildung 43 ist eine Abnahme der Zufriedenheit zu sehen. Dabei verläuft die Abnahme über beide Geschlechter gleich stark. Tendenziell sind die Schülerinnen zufriedener als die Schüler. Die Werte liegen bei beiden Geschlechtern auf einem hohen Niveau.

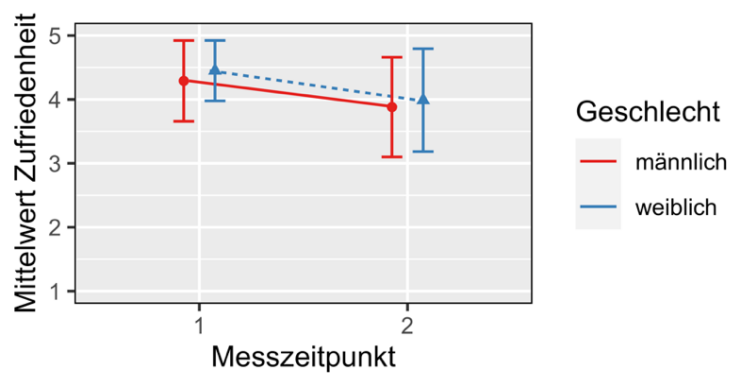


Abbildung 43: Mittelwerte mit Standardabweichung der Zufriedenheit (Usability) mit HyperDocs, aufgeteilt nach Geschlecht nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Oberstufe.

Der robuste gepaarte t-Test zeigt zwischen den Messzeitpunkten ( $M_{MZP1} = 4.36 \pm 0.57$ ,  $M_{MZP2} = 3.93 \pm 0.79$ ) eine signifikante Abnahme mit einer mittleren Effektstärke (*robust*  $t(56) = 7.17$ ,  $p < 0.001$ ,  $\xi = 0.36$ ).

### Einfachheit der Nutzung: Geschlecht\*Messzeitpunkt

Aufgrund der Verletzung der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ), eines Ausreißers und für eine bessere Vergleichbarkeit gegenüber den anderen Analysen wurde erneut eine robuste Varianzanalyse mit Bootstrapping durchgeführt. Die deskriptive Statistik ist Tabelle 90 zu entnehmen.

*Tabelle 90: Deskriptive Statistik der Einfachheit der Nutzung in der Oberstufe.*

Messzeitpunkt	<i>Gymnasium</i>			
	weiblich		männlich	
	1	2	1	2
Stichprobenumfang	43		50	
Mittelwert (Standardabweichung)	4.57 (0.50)	4.26 (0.63)	4.60 (0.48)	4.31 (0.62)
Median	4.67	4.33	4.67	4.33

Erneut liegt keine signifikante Zweifache-Interaktion, jedoch ein signifikanter Haupteffekt „Messzeitpunkt“ vor. Die Teststatistik ist in Tabelle 91 aufgeführt.

*Tabelle 91: Teststatistik der robusten gemischten Varianzanalyse bezüglich der Einfachheit der Nutzung in der Oberstufe.*

<i>Effekt</i>	<i>WJ</i>	<i>df<sub>N</sub></i>	<i>df<sub>D</sub></i>	<i>p</i>
Geschlecht	0.20	1	52.79	0.673
Messzeitpunkt	16.21**	1	51.77	< 0.001
Geschlecht*Messzeitpunkt	0.04	1	51.77	0.841

Die Abnahme über den Messzeitpunkt ist in Abbildung 44 deutlich zu erkennen. Die Abnahme verläuft bei beiden Geschlechtern, ausgehend von dem nahezu gleichen Mittelwert, parallel.

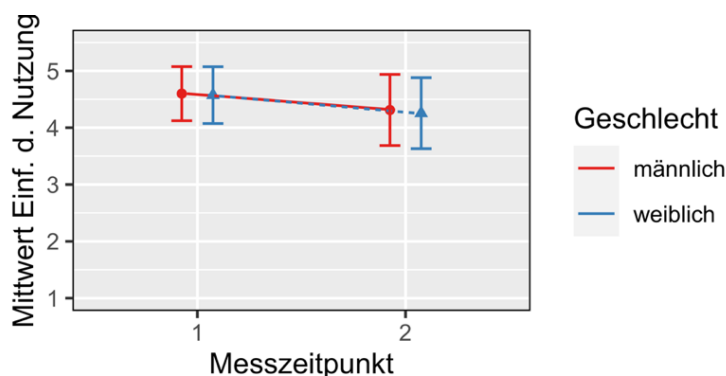


Abbildung 44: Mittelwerte mit Standardabweichung der Einfachheit der Nutzung (Usability) der HyperDocs, aufgeteilt nach Geschlecht nach der ersten (MZIP 1) und vierten (MZIP 2) Stunde in der Oberstufe.

Auch hier zeigt der robuste gepaarte t-Test zwischen den Messzeitpunkten ( $M_{MZIP1} = 4.61 \pm 0.44$ ,  $M_{MZIP2} = 4.33 \pm 0.59$ ) eine signifikante Abnahme mit einer mittleren Effektstärke ( $robust\ t(56) = 3.98$ ,  $p < 0.001$ ,  $\xi = 0.37$ ).

### **Einfachheit des Lernens der Nutzung: Geschlecht\*Messzeitpunkt**

Eine Prüfung der Voraussetzungen zeigt eine Verletzung der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ) und sieben Ausreißer, einer davon extrem. Daher wurde eine robuste gemischte Varianzanalyse mit Bootstrapping durchgeführt. Die deskriptive Statistik ist in Tabelle 92 aufgeführt.

Tabelle 92: Deskriptive Statistik der Einfachheit des Lernens der Nutzung in der Oberstufe.

Messzeitpunkt	Gymnasium			
	weiblich		männlich	
	1	2	1	2
Stichprobenumfang	43		50	
Mittelwert (Standardabweichung)	4.73 (0.42)	4.63 (0.46)	4.79 (0.34)	4.65 (0.47)
Median	5.00	5.00	5.00	5.00

Es zeigt sich erneut ein signifikanter Haupteffekt „Messzeitpunkt“. Die anderen Effekte sind nicht signifikant (vgl. Tabelle 93).

Tabelle 93: Teststatistik der robusten gemischten Varianzanalyse bezüglich der Einfachheit des Lernens der Nutzung in der Oberstufe.

Effekt	WJ	df <sub>N</sub>	df <sub>D</sub>	p
Geschlecht	0.35	1	46.37	0.562
Messzeitpunkt	4.93*	1	54.81	0.023
Geschlecht*Messzeitpunkt	< 0.01	1	54.81	0.973

Abbildung 45 verdeutlicht die Abnahme der Einfachheit des Lernens der Nutzung über die beiden Messzeitpunkte. Erneut bewerten beide Geschlechter diesen Aspekt der Usability in etwa gleich hoch.

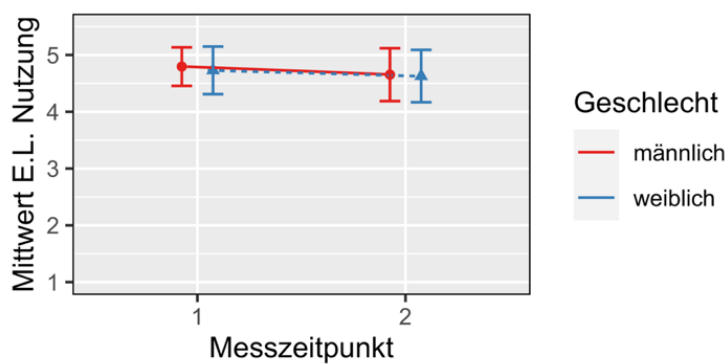


Abbildung 45: Mittelwerte mit Standardabweichung der Einfachheit des Lernens der Nutzung (Usability) der HyperDocs, aufgeteilt nach Geschlecht nach der ersten (MZP 1) und vierten (MZP 2) Stunde in der Oberstufe.

Der robuste gepaarte t-Test zeigt zwischen den Messzeitpunkten ( $M_{MZP1} = 4.76 \pm 0.38$ ,  $M_{MZP2} = 4.64 \pm 0.46$ ) eine signifikante Abnahme mit einer schwachen Effektstärke ( $robust\ t(56) = 2.76$ ,  $p = 0.007$ ,  $\xi = 0.20$ ).

### 5.3.3 Zusammenhangsanalysen zu weiteren Variablen

In diesem Teilkapitel werden Zusammenhangsanalysen zu den Variablen und Konstrukten „Nutzungshäufigkeit von Tablets“, „Tabletbezogene Kompetenzen“, „Cognitive Load“ und „Motivation“ betrachtet. Die Datenauswertung erfolgt getrennt nach Mittel- und Oberstufe, nicht jedoch nach Schulform. Aufgrund der Anwesenheit von Ausreißern werden die Korrelationen mittels *Spearman's rho* ( $\rho$ ) analysiert. Der Stichprobenumfang variiert aufgrund von fehlenden Werten, die nicht vollständig ersetzt werden konnten.

### **Nutzungshäufigkeit von Tablets**

Zunächst erfolgt die Zusammenhangsanalyse zwischen den Faktoren der Usability und der Nutzungshäufigkeit von Tablets im privaten und schulischen Bereich. In die Auswertung werden nur Fälle miteinbezogen, die angaben, ein Tablet zuhause oder in der Schule nutzen zu können.

#### *Mittelstufe*

In der Mittelstufe zeigen sich keine signifikanten Zusammenhänge zwischen der Nutzungshäufigkeit von Tablets im häuslichen oder schulischen Kontext und den Variablen der Usability (vgl. Tabelle 94).

*Tabelle 94: Korrelationen (Spearman's rho) zwischen der Nutzungshäufigkeit von Tablets im häuslichen und schulischen Umfeld sowie den Faktoren der Usability in der ersten Stunde der Mittelstufe.*

	<b>Häusliche Nutzung</b> n = 158	<b>Schulische Nutzung</b> n = 103
Nützlichkeit	$p = 0.762, \rho = 0.02$	$p = 0.139, \rho = -0.15$
Zufriedenheit	$p = 0.860, \rho = -0.01$	$p = 0.751, \rho = 0.03$
Einfachheit der Nutzung	$p = 0.410, \rho = 0.07$	$p = 0.388, \rho = -0.09$
Einfachheit des Lernens	$p = 0.994, \rho < 0.01$	$p = 0.566, \rho = -0.06$

#### *Oberstufe*

Auch in der Oberstufe zeigen sich keine signifikanten Korrelationen (vgl. Tabelle 95).

*Tabelle 95: Korrelationen (Spearman's rho) zwischen der Nutzungshäufigkeit von Tablets im häuslichen und schulischen Umfeld sowie den Faktoren der Usability in der ersten Stunde der Oberstufe.*

	<b>Häusliche Nutzung</b> n = 60	<b>Schulische Nutzung</b> n = 21
Nützlichkeit	$p = 0.347, \rho = -0.12$	$p = 0.256, \rho = -0.26$
Zufriedenheit	$p = 0.211, \rho = -0.16$	$p = 0.750, \rho = -0.07$
Einfachheit der Nutzung	$p = 0.775, \rho = 0.04$	$p = 0.530, \rho = -0.15$
Einfachheit des Lernens	$p = 0.737, \rho = -0.04$	$p = 0.070, \rho = -0.40$



Der Zusammenhang zwischen der Einfachheit des Lernens und der Nutzungshäufigkeit von Tablets im schulischen Bereich zeigt eine mittelstarke negative Ausprägung, die knapp nicht-signifikant ist. Der Stichprobenumfang ist jedoch klein, um eine gesicherte Aussage zu treffen.

### **Cognitive Load und Mental Effort**

Für die Berechnung der Korrelationen wurden die beiden Datensätze zur Usability und dem Cognitive Load sowie dem Mental Effort abgeglichen und die gemeinsamen Fälle identifiziert. Die Auswertung erfolgt differenziert für die erste und die vierte Stunde, da unmittelbar nach diesen Stunden auch die Erhebung der Usability erfolgte.

#### *Mittelstufe*

In Tabelle 96 sind die Korrelationskoeffizienten sowie das Signifikanzniveau gezeigt. Es herrschen zwischen allen Variablen der Usability und den verschiedenen Loadtypen in der ersten Stunde negative Zusammenhänge auf einem schwachen bis mittleren Niveau. In Bezug auf den Mental Effort ist lediglich die negative Korrelation zur Einfachheit des Lernens signifikant. In der vierten Stunde bleiben die Korrelationen auf einem hohen Signifikanzniveau und sie nehmen in Teilen an Stärke zu. Eine Ausnahme bildet der Intrinsic Load. Hierbei nimmt die Stärke des Zusammenhangs in Bezug auf die Zufriedenheit und die Einfachheit des Lernens ab.

*Tabelle 96: Korrelationen zwischen den Faktoren der Usability und den verschiedenen Cognitive Loadtypen sowie dem Mental Effort in der Mittelstufe in der ersten und vierten Stunde (n = 300).*

		<i>Stunde 1</i>			
		Intrinsic L.	Extraneous L.	Germane L.	Mental E.
Nützlichkeit		$p < 0.001,$	$p < 0.001,$	$p < 0.001,$	$p = 0.356,$
		$\rho = -0.24^{**}$	$\rho = -0.27^{**}$	$\rho = 0.29^{**}$	$\rho = -0.05$
Zufriedenheit		$p < 0.001,$	$p < 0.001,$	$p < 0.001,$	$p = 0.876,$
		$\rho = -0.21^{**}$	$\rho = -0.22^{**}$	$\rho = 0.32^{**}$	$\rho < 0.01$
Einfachheit der Nutzung		$p < 0.001,$	$p < 0.001,$	$p < 0.001,$	$p = 0.062,$
		$\rho = -0.34^{**}$	$\rho = -0.39^{**}$	$\rho = 0.28^{**}$	$\rho = -0.11$
Einfachheit des Lernens		$p < 0.001,$	$p < 0.001,$	$p < 0.001,$	$p = 0.02,$
		$\rho = -0.31^{**}$	$\rho = -0.41^{**}$	$\rho = 0.33^{**}$	$\rho = -0.13^*$

		<i>Stunde 4</i>			
		Intrinsic L.	Extraneous L.	Germane L.	Mental E.
Nützlichkeit		$p = 0.018,$	$p < 0.001,$	$p < 0.001,$	$p = 0.729,$
		$\rho = -0.14^*$	$\rho = -0.24^{**}$	$\rho = 0.31^{**}$	$\rho = 0.02$
Zufriedenheit		$p = 0.15,$	$p < 0.001,$	$p < 0.001,$	$p = 0.350,$
		$\rho = -0.09$	$\rho = -0.31^{**}$	$\rho = 0.42^{**}$	$\rho = 0.05$
Einfachheit der Nutzung		$p = 0.003,$	$p < 0.001,$	$p < 0.001,$	$p = 0.900,$
		$\rho = -0.17^{**}$	$\rho = -0.41^{**}$	$\rho = 0.32^{**}$	$\rho = -0.01$
Einfachheit des Lernens		$p = 0.192,$	$p < 0.001,$	$p < 0.001,$	$p = 0.960,$
		$\rho = -0.08$	$\rho = -0.34^{**}$	$\rho = 0.33^{**}$	$\rho < -0.01$

Bei letzterem ist die Korrelation zudem nicht mehr signifikant. Das gleiche Ergebnis lässt sich bei dem Mental Effort beobachten, wonach ebenfalls kein signifikanter Zusammenhang mehr vorliegt.

#### *Oberstufe*

Auch in der Oberstufe bestehen signifikante Korrelationen zwischen den Faktoren der Usability und den einzelnen Loadtypen sowie dem Mental Effort. Dabei ist die Richtung des Zusammenhangs gleich. Teilweise werden jedoch in der ersten Stunde nicht die niedrigen  $p$ -Werte der Mittelstufe erreicht. Als Beispiele sind hier der Extraneous und Intrinsic Load zu nennen, bei denen bei der Nützlichkeit beziehungsweise bei der Zufriedenheit kein signifikanter Zusammenhang besteht (vgl. Tabelle 97). In der vierten Stunde sind nahezu alle Zusammenhänge, bis auf die des Mental Effort, mit einem mittelstarken Effekt signifikant. Die Korrelation zwischen der Nützlichkeit und dem Extraneous Load verbleibt auf einem nicht signifikanten Niveau.

Tabelle 97: Korrelationen zwischen den Faktoren der Usability und den verschiedenen Cognitive Loadtypen sowie dem Mental Effort in der Oberstufe in der ersten und vierten Stunde ( $n = 100$ ).

		<i>Stunde 1</i>			
		Intrinsic L.	Extraneous L.	Germane L.	Mental E.
Nützlichkeit		$p = 0.003,$	$p = 0.086,$	$p = 0.390,$	$p = 0.016,$
		$\rho = -0.30^{**}$	$\rho = -0.17$	$\rho = 0.09$	$\rho = -0.24^{**}$
Zufriedenheit		$p = 0.482,$	$p = 0.002,$	$p < 0.001,$	$p = 0.486,$
		$\rho = -0.07$	$\rho = -0.31^{**}$	$\rho = 0.42^{**}$	$\rho = -0.07$
Einfachheit der Nutzung		$p < 0.001,$	$p < 0.001,$	$p = 0.062,$	$p < 0.001,$
		$\rho = -0.42^{**}$	$\rho = -0.37^{**}$	$\rho = 0.06$	$\rho = -0.40^{**}$
Einfachheit des Lernens		$p < 0.001,$	$p < 0.001,$	$p = 0.587,$	$p = 0.002,$
		$\rho = -0.42^{**}$	$\rho = -0.38^{**}$	$\rho = 0.12$	$\rho = -0.30^{**}$
		<i>Stunde 4</i>			
		Intrinsic L.	Extraneous L.	Germane L.	Mental E.
Nützlichkeit		$p = 0.005$	$p = 0.125$	$p = 0.009$	$p = 0.304$
		$\rho = -0.28^{**}$	$\rho = -0.15$	$\rho = 0.26^{**}$	$\rho = -0.10$
Zufriedenheit		$p = 0.024$	$p = 0.018$	$p < 0.001$	$p = 0.278$
		$\rho = -0.32^*$	$\rho = -0.24^*$	$\rho = 0.47^{**}$	$\rho = -0.11$
Einfachheit der Nutzung		$p < 0.001$	$p = 0.005$	$p < 0.001$	$p = 0.012$
		$\rho = -0.35^{**}$	$\rho = -0.28^{**}$	$\rho = 0.41^{**}$	$\rho = -0.19$
Einfachheit des Lernens		$p = 0.004$	$p = 0.003$	$p < 0.001$	$p = 0.484$
		$\rho = -0.29^{**}$	$\rho = -0.30^{**}$	$\rho = 0.36^{**}$	$\rho = -0.07$

### **Intrinsische Motivation**

Für die Betrachtung der Zusammenhänge zwischen den Variablen der intrinsischen Motivation (SDT) und der Usability wurden die Datensätze erneut abgeglichen. Für die Korrelationen wurden die Datensätze aus der ersten und der letzten Stunde verwendet.

#### *Mittelstufe*

In der Mittelstufe sind alle Korrelationen zwischen den Variablen signifikant mit einem mittelstarken bis starken Zusammenhang (vgl. Tabelle 98). Die Stärke der Korrelationen

nimmt in der vierten Stunde bei fast allen Zusammenhängen noch einmal zu. Der Druck, den die Schüler:innen beim Lernen mit HyperDocSystems empfinden, weist im Vergleich zu den anderen Variablen der intrinsischen Motivation den geringsten Zusammenhang zur Usability auf.

Tabelle 98: Korrelationen zwischen den Faktoren der Usability und den Faktoren der intrinsischen Motivation (SDT) in der Mittelstufe in der ersten und vierten Stunde ( $n = 306$ ).

		<i>Stunde 1</i>			
		Interesse	Wahrgenommene Wahlfreiheit	Kompetenzerleben	Druck
Nützlichkeit		$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,
		$\rho = 0.38^{**}$	$\rho = 0.28^{**}$	$\rho = 0.39^{**}$	$\rho = 0.15^{**}$
Zufriedenheit		$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,	$p = 0.027$ ,
		$\rho = 0.57^{**}$	$\rho = 0.40^{**}$	$\rho = 0.29^{**}$	$\rho = 0.13^*$
Einfachheit der Nutzung		$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,
		$\rho = 0.36^{**}$	$\rho = 0.30^{**}$	$\rho = 0.50^{**}$	$\rho = 0.28^{**}$
Einfachheit des Lernens		$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,
		$\rho = 0.32^{**}$	$\rho = 0.29^{**}$	$\rho = 0.39^{**}$	$\rho = 0.30^{**}$
		<i>Stunde 4</i>			
		Interesse	Wahrgenommene Wahlfreiheit	Kompetenzerleben	Druck
Nützlichkeit		$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,
		$\rho = 0.47^{**}$	$\rho = 0.33^{**}$	$\rho = 0.40^{**}$	$\rho = 0.27^{**}$
Zufriedenheit		$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,
		$\rho = 0.68^{**}$	$\rho = 0.41^{**}$	$\rho = 0.49^{**}$	$\rho = 0.34^{**}$
Einfachheit der Nutzung		$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,
		$\rho = 0.50^{**}$	$\rho = 0.33^{**}$	$\rho = 0.48^{**}$	$\rho = 0.43^{**}$
Einfachheit des Lernens		$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,	$p < 0.001$ ,
		$\rho = 0.34^{**}$	$\rho = 0.27^{**}$	$\rho = 0.44^{**}$	$\rho = 0.38^{**}$

### *Oberstufe*

Im Vergleich zur Mittelstufe sind nicht alle Korrelationen auf einem signifikanten Niveau. Zudem ist die Stärke des Zusammenhangs oftmals geringer. Gerade in Bezug auf

die wahrgenommene Wahlfreiheit (Autonomie) sind die Ergebnisse überwiegend nicht signifikant. Außerdem sind die Korrelationen der Usability in der vierten Stunde nicht einheitlich stärker, wie es bei der Mittelstufe der Fall war (vgl. Tabelle 99).

*Tabelle 99: Korrelationen zwischen den Faktoren der Usability und Faktoren der intrinsischen Motivation (SDT) in der Oberstufe in der ersten und vierten Stunde (n = 101).*

		<i>Stunde 1</i>			
		Interesse	Wahrgenommene Wahlfreiheit	Kompetenzer- leben	Druck
Nützlichkeit		$p < 0.001$ ,	$p = 0.545$ ,	$p < 0.001$ ,	$p = 0.003$ ,
		$\rho = 0.36^{**}$	$\rho = 0.06$	$\rho = 0.37^{**}$	$\rho = 0.29^*$
Zufriedenheit		$p < 0.001$ ,	$p = 0.029$ ,	$p < 0.001$ ,	$p = 0.013$ ,
		$\rho = 0.72^{**}$	$\rho = 0.22^*$	$\rho = 0.41^{**}$	$\rho = 0.25^*$
Einfachheit der Nutzung		$p = 0.004$ ,	$p = 0.148$ ,	$p < 0.001$ ,	$p < 0.001$ ,
		$\rho = 0.29^*$	$\rho = 0.15$	$\rho = 0.55^{**}$	$\rho = 0.42^{**}$
Einfachheit des Lernens		$p = 0.001$ ,	$p = 0.150$ ,	$p < 0.001$ ,	$p < 0.001$ ,
		$\rho = 0.32^{**}$	$\rho = 0.14$	$\rho = 0.47^{**}$	$\rho = 0.45^{**}$
		<i>Stunde 4</i>			
		Interesse	Wahrgenommene Wahlfreiheit	Kompetenzer- leben	Druck
Nützlichkeit		$p < 0.001$ ,	$p = 0.038$ ,	$p = 0.045$ ,	$p = 0.562$ ,
		$\rho = 0.41^{**}$	$\rho = 0.21^*$	$\rho = 0.20^*$	$\rho = 0.06$
Zufriedenheit		$p < 0.001$ ,	$p < 0.001$ ,	$p = 0.014$ ,	$p = 0.017$ ,
		$\rho = 0.77^{**}$	$\rho = 0.35^{**}$	$\rho = 0.36^*$	$\rho = 0.34^*$
Einfachheit der Nutzung		$p < 0.001$ ,	$p = 0.012$ ,	$p < 0.001$ ,	$p < 0.001$ ,
		$\rho = 0.44^{**}$	$\rho = 0.25$	$\rho = 0.48^{**}$	$\rho = 0.45^{**}$
Einfachheit des Lernens		$p < 0.001$ ,	$p = 0.008$ ,	$p < 0.001$ ,	$p < 0.001$ ,
		$\rho = 0.38^{**}$	$\rho = 0.26^{**}$	$\rho = 0.47^{**}$	$\rho = 0.59^{**}$

### **Tabletbezogene Kompetenzen**

Für die Berechnung der Korrelation zwischen der tabletbezogenen Selbstwirksamkeit sowie der Nützlichkeit von Tablets beim Lernen und der Usability wurden zunächst wieder

beide Datensätze angeglichen. Die Korrelationen werden für die erste und die letzte Stunde berechnet.

### Mittelstufe

In der Mittelstufe zeigt sich zwischen allen Variablen der Usability und der tabletbezogenen Selbstwirksamkeit sowie der Nützlichkeit von Tablets beim Lernen ein signifikanter Zusammenhang auf einem schwachen bis mittelstarken Niveau (vgl. Tabelle 100).

*Tabelle 100: Korrelationen zwischen den Faktoren der Usability und der tabletbezogenen Selbstwirksamkeit sowie der Variable „Tablets beim Lernen“ in der Mittelstufe in der ersten und vierten Stunde (n = 287).*

	<i>Stunde 1</i>	
	Tab. Selbstwirksamkeit	Tab. Lernen
Nützlichkeit	$p = 0.002, \rho = 0.18^{**}$	$p < 0.001, \rho = 0.39^{**}$
Zufriedenheit	$p = 0.014, \rho = 0.15^*$	$p < 0.001, \rho = 0.38^{**}$
Einfachheit der Nutzung	$p < 0.001, \rho = 0.25^{**}$	$p < 0.001, \rho = 0.28^{**}$
Einfachheit des Lernens	$p < 0.001, \rho = 0.34^{**}$	$p < 0.001, \rho = 0.30^{**}$
	<i>Stunde 4</i>	
	Tab. Selbstwirksamkeit	Tab. Lernen
Nützlichkeit	$p < 0.001, \rho = 0.29^{**}$	$p < 0.001, \rho = 0.69^{**}$
Zufriedenheit	$p < 0.001, \rho = 0.42^{**}$	$p < 0.001, \rho = 0.69^{**}$
Einfachheit der Nutzung	$p < 0.001, \rho = 0.44^{**}$	$p < 0.001, \rho = 0.64^{**}$
Einfachheit des Lernens	$p < 0.001, \rho = 0.53^{**}$	$p < 0.001, \rho = 0.53^{**}$

Insgesamt korreliert die Variable „Tablets beim Lernen“ stärker mit der Usability als die Selbstwirksamkeit. Die Stärke aller Zusammenhänge nimmt erneut in der vierten Stunde zu. Die jeweiligen Variablen korrelieren auf einem starken Niveau. Auch dort besteht ein stärkerer Zusammenhang der Usability zu der Nützlichkeit von Tablets zum Lernen als zu der tabletbezogenen Selbstwirksamkeit.

### Oberstufe

In der Oberstufe zeigen sich ebenfalls viele signifikante Korrelationen zwischen den einzelnen Variablen. Insgesamt ist die Stärke der Zusammenhänge geringer und die p-Werte sind höher. Hierbei sind die Zusammenhänge zwischen der Einfachheit der Nutzung und Tablets beim Lernen sowie der Einfachheit des Lernens und der Nützlichkeit nicht mehr

signifikant. Im Gegensatz zu der Mittelstufe bestehen zwischen den Variablen der Usability und der Nützlichkeit von Tablets beim Lernen schwächere Korrelationen als zur tabletbezogenen Selbstwirksamkeit. Im Post-Test dreht sich dann dieses Ergebnis um, sodass hohe Zusammenhänge zwischen der Usability und der Nützlichkeit von Tablets beim Lernen bestehen. Die Korrelationen zur Selbstwirksamkeit nehmen in ihrer Stärke dabei ab. Der Zusammenhang zur Nützlichkeit ist nicht mehr signifikant (vgl. Tabelle 101).

*Tabelle 101: Korrelationen zwischen den Faktoren der Usability und der tabletbezogenen Selbstwirksamkeit sowie Tablets beim Lernen in der Oberstufe in der ersten und vierten Stunde (n = 90).*

<b>Stunde 1</b>		
	Tab. Selbstwirksamkeit	Tab. Lernen
Nützlichkeit	$p = 0.002, \rho = 0.32^{**}$	$p = 0.031, \rho = 0.23^*$
Zufriedenheit	$p = 0.039, \rho = 0.22^*$	$p = 0.026, \rho = 0.24^*$
Einfachheit der Nutzung	$p < 0.001, \rho = 0.43^{**}$	$p = 0.271, \rho = 0.12$
Einfachheit des Lernens	$p < 0.001, \rho = 0.50^{**}$	$p = 0.063, \rho = 0.20$
<b>Stunde 4</b>		
	Tab. Selbstwirksamkeit	Tab. Lernen
Nützlichkeit	$p = 0.175, \rho = 0.14$	$p < 0.001, \rho = 0.51^{**}$
Zufriedenheit	$p = 0.043, \rho = 0.21^*$	$p < 0.001, \rho = 0.60^{**}$
Einfachheit der Nutzung	$p < 0.001, \rho = 0.41^{**}$	$p < 0.001, \rho = 0.42^{**}$
Einfachheit des Lernens	$p < 0.001, \rho = 0.45^{**}$	$p = 0.003, \rho = 0.32^{**}$

### 5.3.4 Weitere Betrachtungen der Usability von HyperDocSystems

Im Folgenden werden weitere Aspekte der Usability von HyperDocSystems betrachtet und deskriptiv ausgewertet. Dabei werden die Schulformen im Falle der Mittelstufe zusammen erfasst. Als konkrete Variablen werden hierbei die Usability der Textfelder (Tastatureingabe und handschriftliche Eingabe) und die Erkennbarkeit der Darstellung der HyperDocs und Hilfen bewertet. Die Mittelstufe und Oberstufe werden aufgrund des Altersunterschieds und den damit verbundenen Ansprüchen an ein Lernsystem getrennt voneinander bewertet. Fehlende Werte wurden nicht importiert, da die Schüler:innen lediglich eine Angabe machen sollten, falls sie die jeweilige Texteingabe genutzt haben.

## **Textfelder**

Für die Bewertung der Usability der Textfelder wurde je ein Item aus den verschiedenen Skalen des USE verwendet, um einen Gesamtusabilityscore zu bilden. Die Fragen beziehen sich ausschließlich auf die Nutzung der Texteingabe zur Lösung der Aufgaben.

### *Mittelstufe*

Sowohl nach der erstmaligen Nutzung als auch nach der mehrmaligen Nutzung bewerten Schüler:innen die Usability der Texteingabe und der handschriftlichen Eingabe als hoch. Im Falle der handschriftlichen Eingabe bewerten die Lernenden das System nach der mehrmaligen Nutzung besser. Beide Eingabemethoden werden nahezu gleich oft genutzt (vgl. Tabelle 102).

*Tabelle 102: Deskriptive Statistik der Usability der Eingabeform in der Mittelstufe. n = Stichprobengröße, MZP = Messzeitpunkt.*

<i>Eingabeform</i>	<i>M<sub>MZP1</sub>(SD)</i>	<i>M<sub>MZP2</sub>(SD)</i>
Texteingabe	4.68 (0.53), n = 244	4.68 (0.54), n = 272
Handschriftliche Eingabe	4.40 (0.77), n = 254	4.45 (1.16), n = 272

### *Oberstufe*

Auch in der Oberstufe kann die Usability der Texteingabe und der handschriftlichen Eingabe als hoch eingestuft werden. Hierbei nimmt jedoch die Usability nach mehrmaliger Nutzung stärker ab als in der Mittelstufe (vgl. Tabelle 103). Tendenziell wird die Eingabe mit einem Stylus häufiger verwendet als die Eingabe mit der virtuellen Tastatur. Dabei nimmt jedoch auch die Bewertung der handschriftlichen Eingabe über die Messdauer stärker ab.

*Tabelle 103: Deskriptive Statistik der Usability der Eingabeform in der Oberstufe. n = Stichprobengröße, MZP = Messzeitpunkt.*

<i>Eingabeform</i>	<i>M<sub>MZP1</sub>(SD)</i>	<i>M<sub>MZP2</sub>(SD)</i>
Texteingabe	4.68 (0.55), n = 73	4.56 (0.59), n = 85
Handschriftliche Eingabe	4.60 (0.54), n = 81	4.32 (0.72), n = 89

## **Darstellung der Arbeitsblätter**

Für die Beurteilung der geeigneten Darstellungsgröße und Erkennbarkeit der Inhalte auf dem HyperDoc wurde ebenfalls ein Konstrukt, bestehend aus zwei Fragen, gebildet. Die



Auswertung erfolgt ebenfalls deskriptiv unter Einbezug beider Schulformen in der Mittelstufe. Mittel- und Oberstufe werden getrennt voneinander untersucht.

#### *Mittelstufe*

302 Schüler:innen bewerten die Darstellungsgröße mit einem Mittelwert von  $M = 4.63$  (0.52) nach der erstmaligen Nutzung als gut. Auch am Ende der vier Stunden beurteilen sie die Darstellungsgröße und die Erkennbarkeit der Inhalte mit einem Mittelwert von  $M = 4.55$  (0.73) als ausreichend groß. Es kommt zu einer leichten Abnahme, wie bei bisher nahezu allen betrachteten Aspekten der Usability.

#### *Oberstufe*

Bei der Oberstufe werden die Darstellungsgröße und Sichtbarkeit der Inhalte ebenfalls als sehr hoch eingestuft. Nach der erstmaligen Nutzung beträgt der Mittelwert bei 91 Schüler:innen bei  $M = 4.72$  (0.78) und nach der mehrmaligen Nutzung bei  $M = 4.70$  (0.52). Es kommt damit, im Gegensatz zur Mittelstufe, zu keiner Abnahme der Beurteilung der Darstellung.

### **Darstellung der Hilfen**

Analog zur Erkennbarkeit der Inhalte auf dem HyperDoc wurde nach der Darstellung der Hilfen gefragt. Die Auswertung erfolgt erneut deskriptiv unter Einbezug beider Schulformen in der Mittelstufe. Zusätzlich erfolgt die Auswertung getrennt in Mittel- und Oberstufe. Die Fragen sollten nur von Schüler:innen beantwortet werden, die auch eine Hilfe genutzt haben. Die absolute Stichprobengröße unterscheidet sich daher zwischen den Messzeitpunkten.

#### *Mittelstufe*

In der Mittelstufe bewerten 242 Schüler:innen die Erkennbarkeit der Hilfen mit einem Mittelwert von  $M = 4.39$  (0.74) als sehr hoch. Nach der mehrmaligen Nutzung bewerten 268 Schüler:innen die Erkennbarkeit mit  $M = 4.52$  (0.70). Zu beiden Messzeitpunkten besteht eine hohe Standardabweichung.

#### *Oberstufe*

Nach der erstmaligen Nutzung bewerten 69 Lernende die Lesbarkeit der Hilfen mit einem Mittelwert von  $M = 4.37$  (0.82) sehr gut, wobei eine hohe Standardabweichung besteht. Nach der mehrmaligen Nutzung bewerten 83 Schüler:innen die Erkennbarkeit des

Differenzierungsangebots mit einem Mittelwert von  $M = 4.46$  ( $0.78$ ) ebenfalls als sehr gut. Der Wert nimmt insgesamt zu, wobei die hohe Standardabweichung bestehen bleibt.

## 5.4 Lernhilfen und Vertiefungsaufgaben

Die Auswertung des Differenzierungsangebots, bestehend aus den Hilfen und den Vertiefungsaufgaben, erfolgt in diesem Kapitel. Dabei werden die Nutzungshäufigkeiten der Interventions- und Vergleichsgruppe getrennt voneinander analysiert, da keine genaue Nutzung des Differenzierungsangebots in der Vergleichsgruppe erfasst werden konnte beziehungsweise die Nutzung dessen auf einer subjektiven Einschätzung beruht (vgl. Kapitel 4.5.4). Ein Vergleich erfolgt daher im Rahmen einer deskriptiven Analyse. Bei der Auswertung werden sowohl die Mittelstufe als auch die Oberstufe betrachtet, wobei ein direkter Vergleich zwischen den beiden Gruppen aufgrund der unterschiedlichen unterrichtlichen Einzelstunden nicht sinnvoll erscheint. Im ersten Teilkapitel erfolgt die Analyse der Gründe für die Nutzung von Hilfen und Vertiefungsaufgaben. In den nächsten Teilkapiteln werden dann die verschiedenen Fragen in Bezug auf die Nutzungshäufigkeiten des Differenzierungsangebots ausgewertet. Die Anzahl der Rückmeldungen variiert je nach Stunde und Frage, da lediglich Schüler:innen antworten sollten, wenn sie eine Hilfe oder Vertiefungsaufgabe in der Stunde nutzten.

Zunächst erfolgt die Darstellung der Ergebnisse der Nützlichkeit der Hilfen in der Mittel- und Oberstufe (Interventions- und Vergleichsgruppe) über die vier Unterrichtsstunden und die präferierte Form der Hilfedarbietung im Pre-/Post-Vergleich (lediglich Interventionsgruppe).

In Bezug auf die Nützlichkeit der Hilfen zeigt sich über die Stunden hinweg in der Interventionsgruppe der Mittelstufe eine leichte Abnahme. Dennoch liegt das Niveau selbst in der vierten Stunde mit  $M = 3.86$  noch auf einem hohen Niveau (vgl. Abbildung 46 & Anhang 9.3.5).

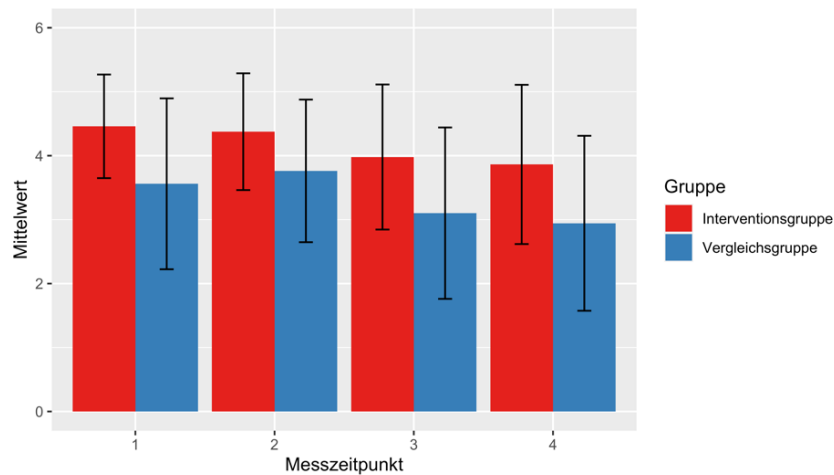


Abbildung 46: Mittelwerte mit Standardabweichung der Nützlichkeit der Lernhilfen, aufgeteilt nach Gruppe in den einzelnen Stunden in der Mittelstufe.

Die Vergleichsgruppe bewertet die Hilfen als deutlich weniger nützlich. Auch in dieser Gruppe ist ab der dritten Stunde eine Abnahme der Bewertung der Nützlichkeit zu beobachten.

In der Interventionsgruppe der Oberstufe liegt die Bewertung der Nützlichkeit der Lernhilfen ebenfalls auf einem hohen Niveau. Dabei kommt es allerdings zu einer Zunahme der Bewertung der Nützlichkeit zur dritten und vierten Stunde (vgl. Abbildung 47 & Anhang 9.3.6).

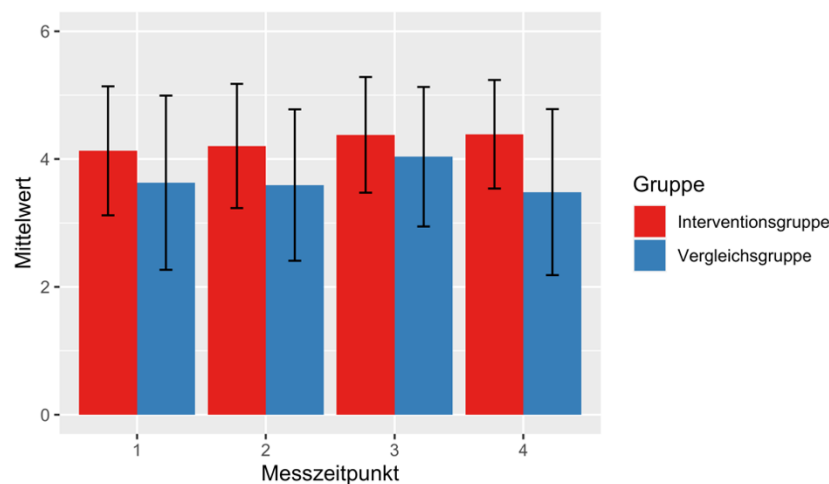


Abbildung 47: Mittelwerte mit Standardabweichung der Nützlichkeit der Lernhilfen, aufgeteilt nach Gruppe in den einzelnen Stunden in der Oberstufe.

In der Vergleichsgruppe werden in der Oberstufe, wie auch schon in der Mittelstufe, die Hilfen als weniger nützlich bewertet. Auch dort erfährt die Nützlichkeit der Hilfen in der dritten Stunde eine Zunahme.

In Bezug auf die Angebotsform, digitale Lernhilfen oder analoge Lernhilfen, präferieren die Schüler:innen der Mittel- und Oberstufe der Interventionsgruppe ebenfalls stark die Lernhilfen in HyperDocs, wie Abbildung 48 verdeutlicht. Dabei kommt es auch zu keiner wahrnehmbaren Veränderung über die Unterrichtsreihe. In der Vergleichsgruppe wurde diese Frage nicht erhoben.

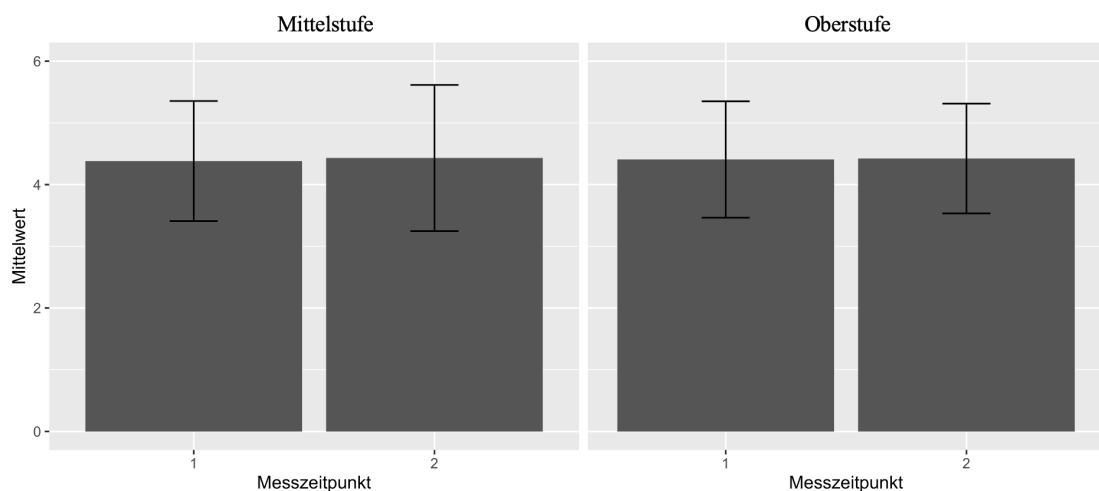


Abbildung 48: Mittelwerte mit Standardabweichung der präferierten Angebotsform (digital) der Lernhilfen in der Mittel- und Oberstufe im Pre- und Posttest (MZP 1 und MZP 2).

#### 5.4.1 Gründe für die Nutzung des Differenzierungsangebots

Im nachfolgenden Teil werden die Gründe für die Nutzung der Lernhilfen und Vertiefungsaufgaben in den verschiedenen Gruppen (Mittelstufe / Oberstufe, Interventionsgruppe / Vergleichsgruppe) deskriptiv analysiert. Die Auswertung erfolgt mittels Säulendiagrammen. Dabei soll durch das geschlossene Antwortformat und die Befragung nach jeder Stunde eine Veränderung über die vier Unterrichtsstunden betrachtet werden. Da Schüler:innen lediglich antworten sollten, wenn sie die Differenzierung nutzten, variiert die Zahl der Antworten in den einzelnen Stunden. Fehlende Werte wurden daher nicht ersetzt.

##### Mittelstufe

###### *Interventionsgruppe*

In der Interventionsgruppe der Mittelstufe kommt es zu einer sichtbaren Veränderung der Gründe für die Nutzung der Lernhilfen (vgl. Abbildung 49 & Anhang 9.3.5). Die Nutzung aus Neugier (H3) und Interesse (H6) nimmt vor allem in den ersten beiden Stunden ab. Gleichzeitig steigt die Nutzung wegen des tatsächlichen Hilfebedarfs (H4) deutlich an. Auch die Verwendung der Hilfen aus Langeweile (H5) steigt über den Messzeitpunkt

kontinuierlich an, wie in Abbildung 49 ersichtlich wird. Unter Betrachtung des Betrags der Zustimmung sind alle Gründe bis auf H5 in etwa gleichermaßen stark. Ab der zweiten Stunde überwiegt jedoch leicht die Nutzung aus dem tatsächlichen Hilfebedarf (H4).

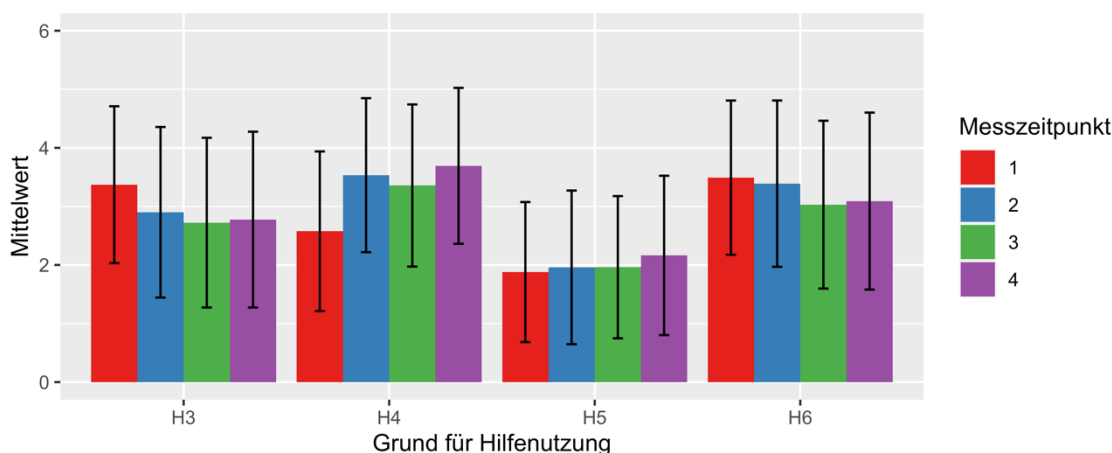


Abbildung 49: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Lernhilfen in der Interventionsgruppe der Mittelstufe nach Messzeitpunkt (Stunde). Nutzungsgrund: H3: Neugier, H4: Hilfebedarf, H5: Langeweile, H6: Interesse.

Bei der Nutzung der Vertiefungsaufgaben zeigen sich ähnliche Trends. Das Öffnen einer Vertiefungsaufgabe aus Neugier (V1) und Interesse (V2) nimmt kontinuierlich in den Stunden ab. Gleichzeitig nimmt die Betrachtung einer Vertiefungsaufgabe aus Langeweile zu (V3). Die Nutzung einer Vertiefungsaufgabe, um sein eigenes Fachwissen zu erweitern (V4), nimmt ebenfalls tendenziell ab der dritten Stunde ab. Als zeitlichen Puffer nutzen die Schüler:innen Vertiefungsaufgaben stetig seltener (V5). Insgesamt verwenden Schüler:innen die Vertiefungsaufgaben besonders aus Interesse und als zeitlichen Puffer, wobei in der letzten Stunde, mit Ausnahme der Langeweile, allen Gründen gleichermaßen stark zugestimmt wird (vgl. Abbildung 50 & Anhang 9.3.5).

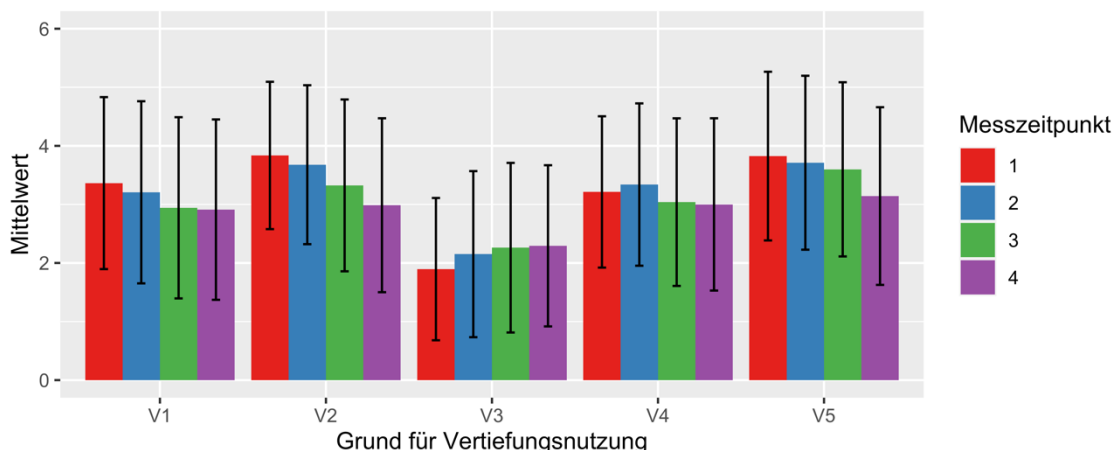


Abbildung 50: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Vertiefungsaufgaben in der Interventionsgruppe der Mittelstufe nach Messzeitpunkt (Stunde). Nutzungsgrund: V1: Neugier, V2: Interesse, V3: Langeweile, V4: Fachwissen serweiterung, V5: Zeitlicher Puffer.

Neben den verschiedenen Gründen für die Nutzung des Differenzierungsangebots wurde auch von den Schüler:innen eine Einschätzung bezüglich der zur Verfügung stehenden Zeit eingeholt. Demnach sollten die Lernenden angeben, ob sie genügend Zeit hatten, um die HyperDocs zu bearbeiten (Z1) und ob sie sich mehr Zeit gewünscht hätten (Z2 - invertiert). Die Ergebnisse sind in Abbildung 51 und Anhang 9.3.5 dargestellt.

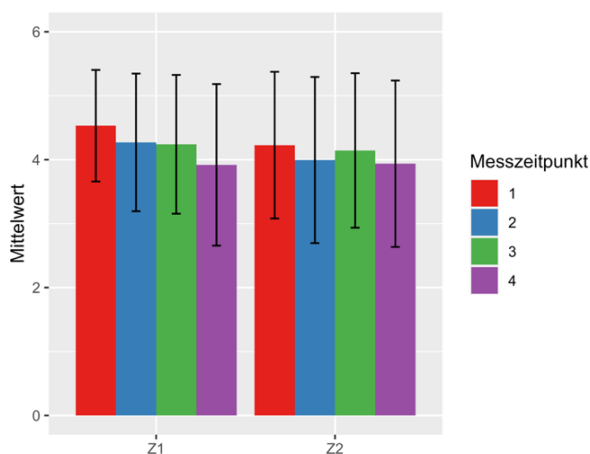


Abbildung 51: Mittelwerte mit Standardabweichung der Beurteilung der zeitlichen Komponente bei der Bearbeitung der HyperDocs in der Mittelstufe (Interventionsgruppe) nach Messzeitpunkt (Stunde). Ein niedriger Wert von Z1 und Z2 signalisiert Zeitdruck.

Während insbesondere im Vergleich zwischen der ersten und der vierten Stunde die verfügbare Zeit zunehmend als kürzer bewertet wird, wünschten sich die Schüler:innen jedoch auch nicht zwangsläufig mehr Zeit, um die Aufgaben zu bearbeiten. Insgesamt bewerten die Jugendlichen die zur Verfügung stehende Zeit als ausreichend.

### Vergleichsgruppe

In der Vergleichsgruppe zeigt sich ein ähnliches Bild wie in der Interventionsgruppe (vgl. Abbildung 52 & Anhang 9.3.5). Demnach nimmt die Nutzung der Lernhilfen aus Neugier (H3) und dem Interesse (H6) über die Stunden tendenziell ab. Gleichzeitig nimmt die Verwendung aus dem tatsächlichen Hilfebedarf heraus stark zu (H4). Die Nutzung aus Langeweile (H5) zeigt im Vergleich zur Interventionsgruppe eher einen gegenteiligen Trend, wobei auch hier die Unterschiede zwischen den Stunden sehr gering sind. Bei Betrachtung des Betrags der Zustimmung, ist die Nutzung aus Neugier und Interesse deutlich geringer als in der Interventionsgruppe. In der Vergleichsgruppe überwiegt demnach die Nutzung aus dem tatsächlichen Hilfebedarf (H4).

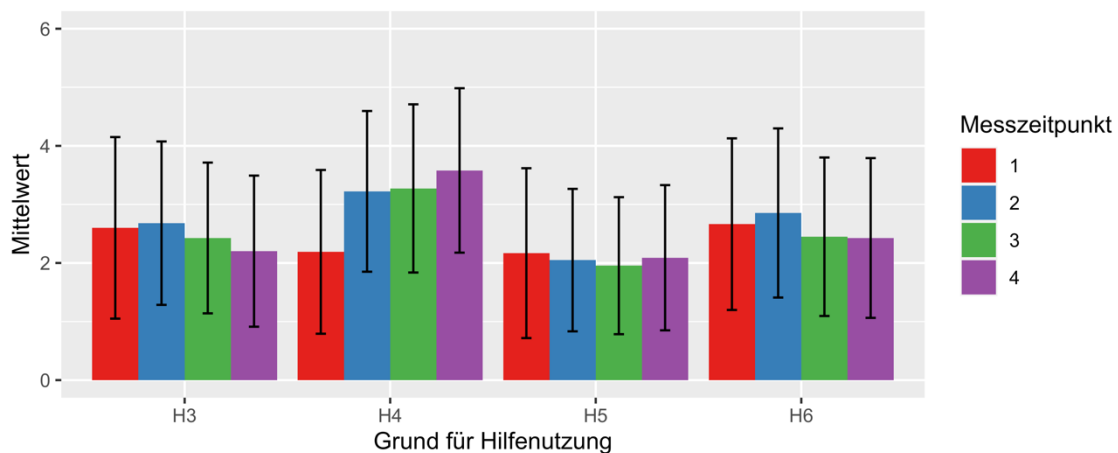


Abbildung 52: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Lernhilfen in der Vergleichsgruppe der Mittelstufe nach Messzeitpunkt (Stunde). Nutzungsgrund: H3: Neugier, H4: Hilfebedarf, H5: Langeweile, H6: Interesse.

In Bezug auf die Gründe für die Nutzung der Vertiefungsaufgaben zeigt sich ebenfalls ein ähnliches Bild wie in der Interventionsgruppe (vgl. Abbildung 53 & Anhang 9.3.5). Demnach nimmt die Nutzung aus Neugier (V1) und Interesse (V2) sowie als zeitlicher Puffer (V5) über die Einsatzdauer ab. Dabei bleiben die weiteren Gründe „Langeweile“ (V3) und „Fachwissen“ (V4) nahezu konstant. Bezüglich des Grades der Zustimmung sind Unterschiede insbesondere bei der „Langeweile“ (V3), „Zeitlicher Puffer“ (V5) und „Fachwissen“ (V4) erkennbar. Während die ersten beiden genannten Gründe in der Vergleichsgruppe überwiegen, wird letzterem in der Interventionsgruppe stärker zugestimmt.

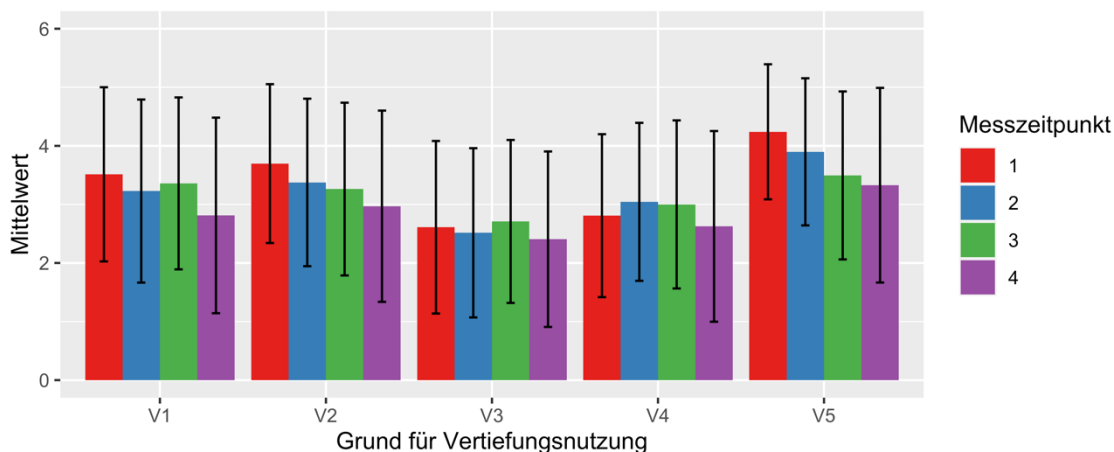


Abbildung 53: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Vertiefungsaufgaben in der Vergleichsgruppe der Mittelstufe nach Messzeitpunkt (Stunde).  
 Nutzungsgrund: V1: Neugier, V2: Interesse, V3: Langeweile, V4: Fachwissenserweiterung, V5: Zeitlicher Puffer.

In der Vergleichsgruppe wird die zur Verfügung stehende Zeit als ausreichend eingestuft. Wie auch schon in der Interventionsgruppe, geben die Schüler:innen zunehmend häufiger an, weniger Zeit zur Verfügung zu haben. Gleichzeitig wünschen sie sich mehr Zeit, um die Aufgaben zu bearbeiten (vgl. Abbildung 54 & Anhang 9.3.5).

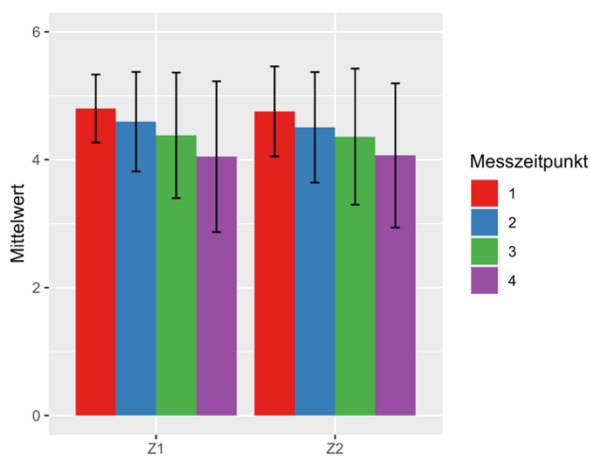


Abbildung 54: Mittelwert mit Standardabweichung der Beurteilung der zeitlichen Komponente bei der Bearbeitung der Arbeitsblätter in der Mittelstufe (Vergleichsgruppe) nach Messzeitpunkt (Stunde). Ein niedriger Wert von Z1 und Z2 signalisiert Zeitdruck.



## Oberstufe

### *Interventionsgruppe*

Die Gründe für die Nutzung der Lernhilfen zeigen ein ähnliches Bild wie in der Mittelstufe (vgl. Abbildung 55 & Anhang 9.3.6). Die Verwendung der Hilfen aus Neugier (H3) nimmt dabei leicht ab. Die Nutzung aus Interesse (H6) bleibt hingegen konstant. Eine deutliche Zunahme der Zustimmung über die Stunden erfolgt für die Nutzung aus dem tatsächlichen Hilfebedarf heraus (H4). Aus Langeweile (H5) rufen die Lernenden eher selten eine Hilfe auf. Ansonsten werden die Hilfen in etwa gleich häufig wegen der Neugier, des Interesses und des Hilfebedarfs genutzt.

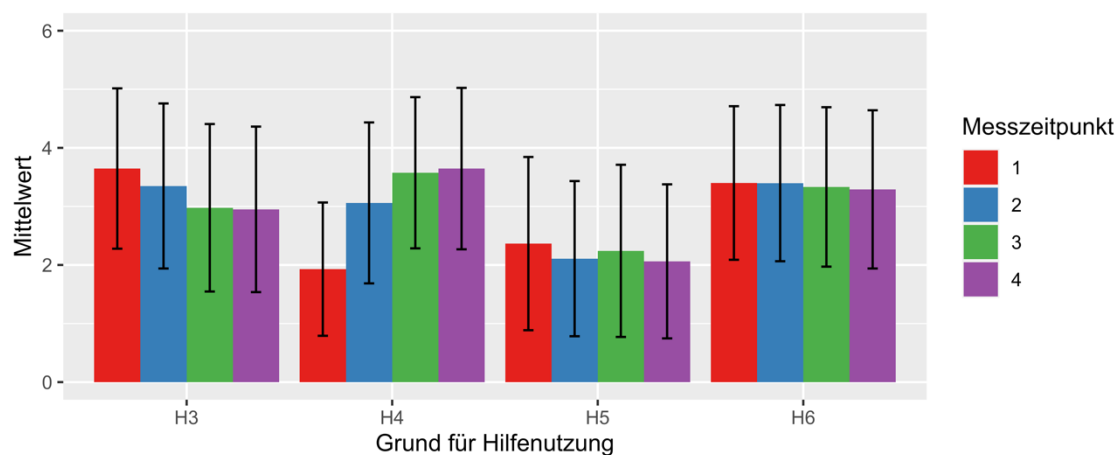


Abbildung 55: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Lernhilfen in der Interventionsgruppe in der Oberstufe nach Messzeitpunkt (Stunde). Nutzungsgrund: H3: Neugier, H4: Hilfebedarf, H5: Langeweile, H6: Interesse.

Bei Betrachtung der Gründe für die Nutzung der Vertiefungsaufgaben in der Oberstufe zeigt sich insbesondere eine Abnahme der Nutzung aus Interesse (V2) und als zeitlicher Puffer (V5). Auch die Verwendung aus Neugier (V1) nimmt ab der dritten Stunde leicht ab. Die anderen beiden Gründe „Fachwissen vertiefen“ (V4) und „Langeweile“ (V3) zeigen keinen eindeutigen Trend. In Bezug auf die Stärke der Zustimmung zeigen die Gründe „Neugier“, „Interesse“ und „Zeitlicher Puffer“ die meiste Befürwortung (vgl. Abbildung 56 & Anhang 9.3.6).

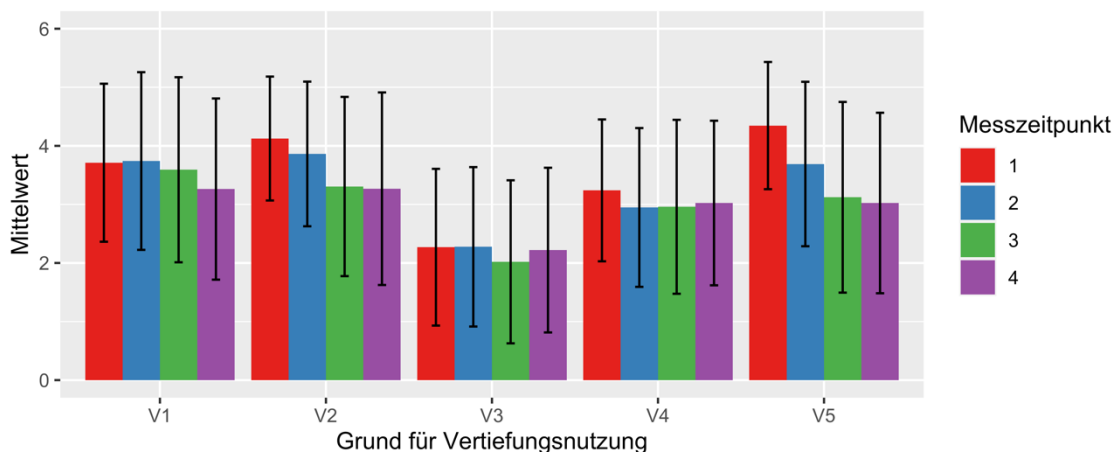


Abbildung 56: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Vertiefungsaufgaben in der Interventionsgruppe der Oberstufe nach Messzeitpunkt (Stunde). Nutzungsgrund: V1: Neugier, V2: Interesse, V3: Langeweile, V4: Fachwissenserweiterung, V5: Zeitlicher Puffer.

Auch in der Oberstufe wurde von den Schüler:innen eine Selbsteinschätzung bezüglich der zur Verfügung stehenden Zeit eingeholt. Auch sie bewerten die Zeitkomponente als ausreichend, oftmals sogar höher als die Mittelstufenschüler:innen (vgl. Abbildung 57 & Anhang 9.3.6).

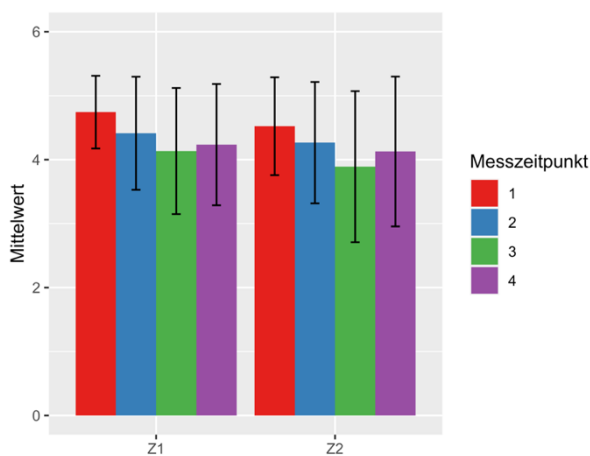


Abbildung 57: Mittelwert mit Standardabweichung der Beurteilung der zeitlichen Komponente bei der Bearbeitung der HyperDocs in der Oberstufe (Interventionsgruppe) nach Messzeitpunkt (Stunde). Ein niedriger Wert von Z1 und Z2 signalisiert Zeitdruck.

Auch in der Oberstufe schätzen die Schüler:innen zunehmend die zur Verfügung stehende Zeit als kürzer ein. Im gleichen Moment wünschen sie sich auch zunehmend mehr Zeit, um die Aufgaben zu bearbeiten.

### Vergleichsgruppe

In der Vergleichsgruppe nimmt die Nutzung der Hilfen aus Neugier (H3) und Interesse (H6) über die Stunden ab (vgl. Abbildung 58 & Anhang 9.3.6). Die Verwendung aus dem Hilfebedarf heraus (H4) nimmt dabei bis zur dritten Stunde zu. Aus Langeweile öffnen die Schüler:innen insbesondere in der ersten Stunde die Lernhilfen. In den weiteren Stunden bleibt der Wert jedoch konstant niedrig. Im Vergleich zur Interventionsgruppe zeigt sich bei nahezu allen Gründen eine geringere Zustimmung. Am häufigsten Nutzen die Lernenden die Hilfe aus dem Hilfebedarf heraus (H4).

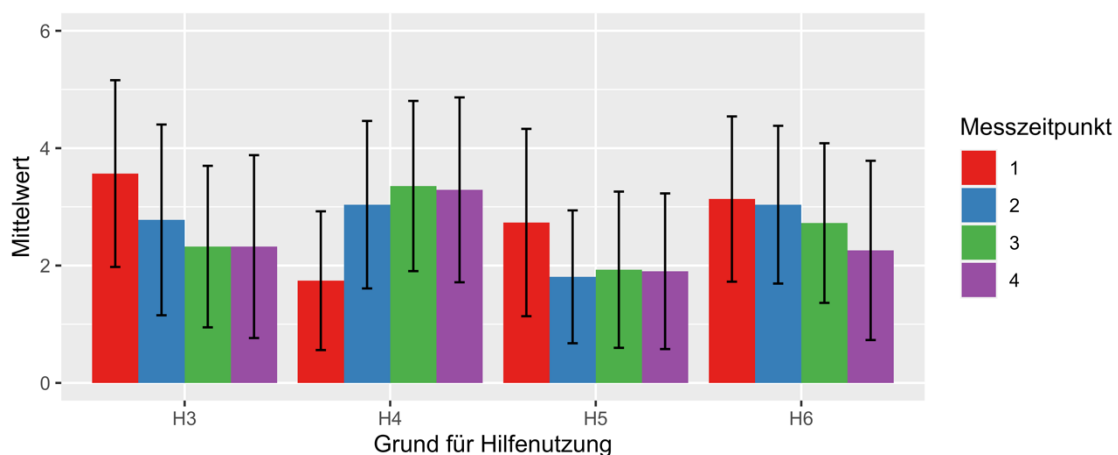


Abbildung 58: Mittelwerte mit Standardabweichung der Gründe für die Nutzung der Lernhilfen in der Vergleichsgruppe in der Oberstufe nach Messzeitpunkt (Stunde). Nutzungsgrund: H3: Neugier, H4: Hilfebedarf, H5: Langeweile, H6: Interesse.

Im Verlauf und im Betrag der Zustimmung ähneln die Antworten der Vergleichsgruppe denen der Interventionsgruppe (vgl. Abbildung 59 & Anhang 9.3.6). Demnach nimmt die Nutzung der Vertiefungsaufgaben als zeitlicher Puffer (V5) mit den Stunden ab. Gleiches gilt für die Öffnung aus Neugier (V2) oder um Fachwissen zu erschließen (V4). Auch die Verwendung aus Neugier (V1) nimmt ab der dritten Stunde ab. In Bezug auf die absolute Zustimmung verwenden die Schüler:innen die Vertiefungsaufgaben am häufigsten aus Neugier (V1), Interesse (V2) und als zeitlichen Puffer (V5).

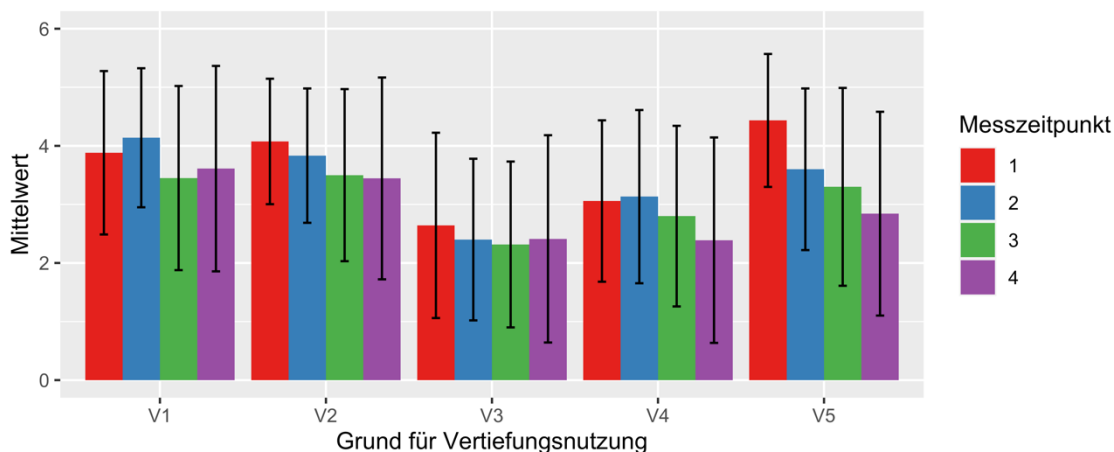


Abbildung 59: Mittelwerte mit Standardabweichung der Antworten der Gründe für die Nutzung der Vertiefungsaufgaben in der Vergleichsgruppe der Oberstufe nach Messzeitpunkt (Stunde). Nutzungsgrund: V1: Neugier, V2: Interesse, V3: Langeweile, V4: Fachwissenserweiterung, V5: Zeitlicher Puffer.

In der Vergleichsgruppe der Oberstufe fällt die Beurteilung der zur Verfügung stehenden Zeit differenzierter aus. Insbesondere in der ersten Stunde hatten die Lernenden ausreichend Zeit, um die Aufgaben zu bearbeiten. Sie hätten sich nicht mehr Zeit gewünscht. Ab der zweiten Stunde nimmt diese Einschätzung stark ab, wobei die Schüler:innen in der vierten Stunde nach eigenen Angaben wieder mehr Zeit zur Verfügung hatten (vgl. Abbildung 60 & Anhang 9.3.6). Es bestehen damit leichte Unterschiede zur Interventionsgruppe. Dort wurde die zur Verfügung stehende Zeit als höher eingestuft.

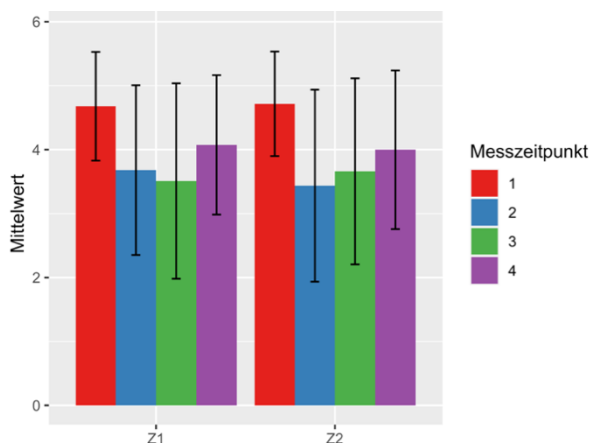


Abbildung 60: Mittelwert mit Standardabweichung der Beurteilung der zeitlichen Komponente bei der Bearbeitung der Arbeitsblätter in der Oberstufe (Vergleichsgruppe) nach Messzeitpunkt (Stunde). Ein niedriger Wert von Z1 und Z2 signalisiert Zeitdruck.

### 5.4.2 Bevorzugtes Medium der Hilfen

In diesem Teilkapitel wird die bevorzugte Art der Hilfen in der Interventions- und Vergleichsgruppe in der Mittel- und Oberstufe, unabhängig von der Schulform oder dem Geschlecht, analysiert. Zu jeder Aufgabe wurden, sofern möglich, Hilfen in Form eines Textes, Bildes, Sprachaufnahme (Audio) oder Video angeboten (vgl. Kapitel 4.3.3). Die nachfolgenden Nutzungen berücksichtigen in der Interventionsgruppe sowohl Hilfen, die als „gelesen“ angenommen werden können, als auch die Rohdaten. Zudem beinhalten die Nutzungsdaten einen Korrekturfaktor, der die Gesamtanzahl einer Hilfeart pro Arbeitsblatt berücksichtigt. Beispielsweise wurden Hilfen in Form eines Videos auf den HyperDocs seltener angeboten als Hilfen in Textform. Die Anzahl der Nutzungen wurde daher durch die Anzahl der zur Verfügung stehenden Hilfen geteilt. Aus den jeweiligen Mittelwerten wurde anschließend die prozentuale Verteilung berechnet und dargestellt. In der Vergleichsgruppe basieren die Daten auf einer Selbstangabe. Dabei werden keine absoluten Verteilungen, sondern Tendenzen erfasst (vgl. Kapitel 4.2.2). Fehlende Werte, durch zum Beispiel eine fehlende Nutzung durch die Lernenden, wurden nicht ersetzt. Für die Auswertung wurde auch die Hilfenutzung bezüglich des Experiments berücksichtigt (vgl. Kapitel 4.3.3).

#### **Interventionsgruppe**

##### *Mittelstufe*

*Tabelle 104: Mittelwerte mit Standardabweichung (in Klammern) der genormten Nutzung der Lernhilfen nach Typ, bezogen auf gelesene Hilfen in der Mittelstufe (n = 331).*

	<i>Stunde 1</i>	<i>Stunde 2</i>	<i>Stunde 3</i>	<i>Stunde 4</i>
Mittelwert Text	0.17 (0.38)	0.47 (0.55)	0.12 (0.19)	0.33 (0.43)
Mittelwert Audio	0.06 (0.20)	0.07 (0.21)	0.02 (0.09)	0.02 (0.08)
Mittelwert Bild	0.36 (0.60)	0.39 (0.58)	0.15 (0.24)	0.29 (0.41)
Mittelwert Video	0.02 (0.11)	0.08 (0.19)	0.01 (0.05)	0.04 (0.12)

Über alle Stunden hinweg favorisieren die Schüler:innen Hilfen in Form von Bildern oder Texten. Video und Audiohilfen werden selten verwendet (vgl. Tabelle 104). Der

gemeinsame Anteil der Text- und Bildhilfen liegt bei über 80 Prozent (vgl. Abbildung 61). Eine Entwicklung im Verlauf der Zeit ist nicht zu erkennen.

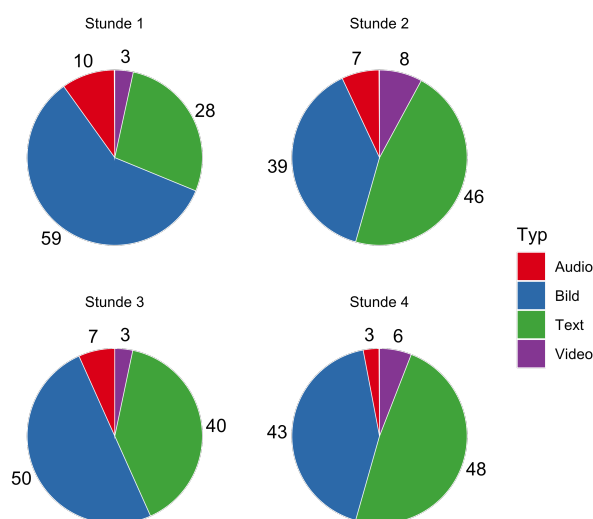


Abbildung 61: Prozentuale Verteilung der Hilfenutzung nach Typ der Hilfe (gelesen) in der Mittelstufe (n = 331).

Bei Betrachtung der Rohwerte, ohne vorherige Bereinigung der Hilfen, finden sich ähnliche Tendenzen wieder, wobei erwartungsgemäß die prozentuale Nutzung von Audio- und Videohilfen höher liegt (vgl. Abbildung 62). Dafür nimmt die Nutzung der Bildhilfen ab, während die Verwendung der Texthilfen stabil bleibt.

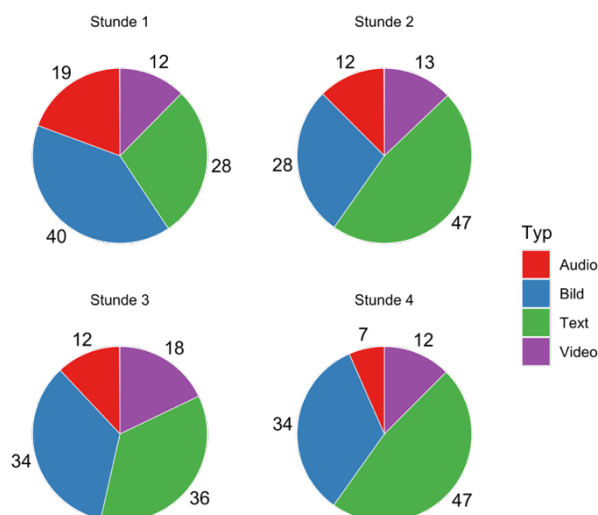


Abbildung 62: Prozentuale Verteilung der Hilfenutzung nach Typ der Hilfe (Rohdaten) in der Mittelstufe (n = 331). Aufgrund von gerundeten Werten ergibt die Prozentsumme nicht bei allen Stunden 100.

## Oberstufe

Tabelle 105: Mittelwerte und Standardabweichung (in Klammern) der genormten Nutzung der Lernhilfen nach Typ, bezogen auf gelesene Hilfen in der Oberstufe ( $n = 110$ , Ausnahme Stunde 2 & 3: 109).

	Stunde 1	Stunde 2	Stunde 3	Stunde 4
Mittelwert Text	0.19 (0.34)	0.18 (0.25)	1.26 (1.24)	0.37 (0.36)
Mittelwert Audio	0.04 (0.15)	0.01 (0.03)	0.04 (0.14)	< 0.01 (0.27)
Mittelwert Bild	0.31 (0.47)	0.33 (0.50)	0.69 (1.14)	0.23 (0.37)
Mittelwert Video	0.03 (0.11)	0.06 (0.25)	0.06 (0.24)	0.02 (0.08)

In der Oberstufe zeigt sich ein ähnliches Bild wie in der Mittelstufe (vgl. Tabelle 105). Demnach werden die Hilfen in Form eines Textes oder Bildes am häufigsten von den Schüler:innen herangezogen. Insgesamt werden die Audio- und Videohilfen seltener verwendet. Während in den ersten beiden Stunden die Nutzung der Bildhilfen überwiegt, nutzen die Lernenden in den letzten beiden Stunden vermehrt die Texthilfen (vgl. Abbildung 63).

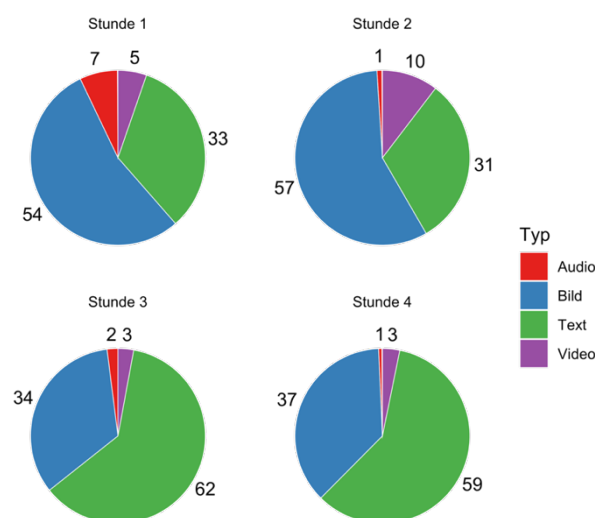


Abbildung 63: Prozentuale Verteilung der Hilfenutzung nach Typ der Hilfe (gelesen) in der Oberstufe ( $n = 110$ , Ausnahme Stunde 2 & 3: 109). Aufgrund von gerundeten Werten ergibt die Prozentsumme nicht bei allen Stunden 100.

Auch bei der Oberstufe verschiebt sich die prozentuale Nutzung bei Betrachtung der Rohdaten zugunsten der Audio- und Videohilfen, wobei die Zunahme nicht so stark ausfällt wie in der Mittelstufe (vgl. Abbildung 64).

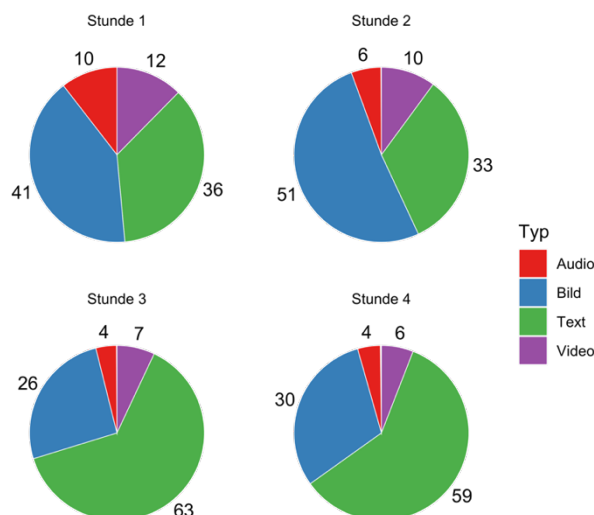


Abbildung 64: Prozentuale Verteilung der Hilfenutzung nach Typ der Hilfe (Rohdaten) in der Oberstufe (n = 110, Ausnahme Stunde 2 & 3: 109). Aufgrund von gerundeten Werten ergibt die Prozentsumme nicht bei allen Stunden 100.

**Vergleichsgruppe**

*Mittelstufe*

In der Vergleichsgruppe wurde für jede Aufgabe gefragt, welche Art von Hilfe (5-stufige Skala vom Typ Likert mit den Polen „viel häufiger visuell“ bis „viel häufiger verbal“) genutzt wurde. Für die jeweiligen Stunden wurden die Daten zusammengefasst und als Mittelwert ausgegeben (vgl. Tabelle 106). Die Daten basieren auf den Angaben von 151 Schüler:innen, wovon jedoch nicht jede:r eine Hilfe zu den jeweiligen Zeitpunkten nutzte.

Tabelle 106: Mittelwerte mit Standardabweichung (in Klammern) bezüglich der Art der Hilfenutzung in der Mittelstufe (n = 151 Schüler:innen). Werte über 3 bedeuten eine häufigere Nutzung der verbalen Hilfen (Text). Werte kleiner als 3 bedeuten eine häufigere Nutzung der visuellen Hilfen (Bild).

	<i>Stunde 1</i>	<i>Stunde 2</i>	<i>Stunde 3</i>	<i>Stunde 4</i>
Mittelwert bevorzugte Nutzung	2.25 (1.05)	3.02 (1.05)	2.71 (0.98)	2.76 (1.09)

Aus der Tabelle geht hervor, dass die Lernenden in der Vergleichsgruppe in der ersten Stunde tendenziell die visuellen Lernhilfen nutzten, während in den weiteren Stunden die Verteilung ausgeglichener erscheint. Insgesamt bevorzugen die Lernenden die visuellen Lernhilfen.



### Oberstufe

Die Ergebnisse werden analog zur Mittelstufe dargestellt. Die Stichprobe umfasst Daten von 53 Schüler:innen, wovon erneut nicht alle eine Lernhilfe in den jeweiligen Stunden nutzten.

*Tabelle 107: Mittelwerte mit Standardabweichung (in Klammern) bezüglich der Art der Hilfe-nutzung in der Oberstufe (n = 53 Schüler:innen). Werte über 3 bedeuten eine häufigere Nutzung der verbalen Hilfen (Text). Werte kleiner als 3 bedeuten eine häufigere Nutzung der visuellen Hilfen (Bild).*

	<i>Stunde 1</i>	<i>Stunde 2</i>	<i>Stunde 3</i>	<i>Stunde 4</i>
Mittelwert bevor-zugte Nutzung	2.30 (0.97)	2.44 (1.00)	3.00 (1.43)	2.73 (1.23)

In der Oberstufe zeigen sich ähnliche Ergebnisse wie in der Mittelstufe (vgl. Tabelle 107). In der ersten Stunde werden bevorzugt die visuellen Hilfen verwendet. In der dritten Stunde ist die Nutzung ausgeglichen. In den anderen beiden Stunden werden tendenziell erneut die visuellen Lernhilfen bevorzugt.

#### 5.4.3 Einfluss der Einsatzdauer, des Geschlechts und der Schulform in der Mittelstufe

Die Analyse zwischen Vergleichs- und Interventionsgruppe erfolgt aufgrund der verschiedenen Erhebungsmethoden und des verschiedenen Gesamtangebots an Hilfen getrennt. Die Nutzungshäufigkeiten der Hilfen wurden durch die zur Verfügung stehende Bearbeitungszeit in Minuten (nur Interventionsgruppe) und die Anzahl der jeweiligen Hilfen in den betrachteten Schulstunden geteilt (vgl. Kapitel 4.1.2). Zunächst werden die drei in der Überschrift genannten Faktoren für die Interventionsgruppe analysiert. Anschließend erfolgt die Analyse der Vergleichsgruppe unter Ausschluss des Faktors „Schulform“, da hierfür keine Daten für die Gesamtschule gewonnen werden konnten. Für die Auswertung unter Verwendung einer gemischten Varianzanalyse werden zunächst die Voraussetzungen geprüft. Die Auswertung erfolgt analog zu den vorherigen Kapiteln 5.2 und 5.3.

**Hilfenutzung***Mittelstufe*

Da eine Verletzung der Sphärizität ( $p < 0.001$ ), der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk*:  $p < 0.001$ ), der Varianzhomogenität in der dritten Stunde ( $p = 0.047$ ) sowie der Homogenität der Kovarianzmatrizen für den Faktor „Gruppe“ und „Geschlecht“ vorliegt, wurde eine robuste gemischte Varianzanalyse mit Bootstrapping eingesetzt. Bestärkt wurde diese Entscheidung durch das Vorhandensein von 46 Ausreißern, wovon fünf extrem sind. Die deskriptive Statistik ist Tabelle 108 zu entnehmen.

*Tabelle 108: Deskriptive Statistik der Hilfenutzung im Schulformvergleich in der Mittelstufe.*

	<b>Gymnasium</b>							
	weiblich				männlich			
	1	2	3	4	1	2	3	4
Messzeitpunkt								
Stichprobenumfang	107				98			
Mittelwert	0.010	0.013	0.004	0.007	0.011	0.010	0.005	0.006
(Standardabweichung)	(0.013)	(0.014)	(0.005)	(0.006)	(0.014)	(0.009)	(0.006)	(0.006)
Median	0.000	0.009	0.003	0.006	0.008	0.007	0.003	0.004
	<b>Gesamtschule</b>							
	weiblich				männlich			
	1	2	3	4	1	2	3	4
Messzeitpunkt								
Stichprobenumfang	52				63			
Mittelwert	0.008	0.018	0.006	0.009	0.011	0.011	0.007	0.005
(Standardabweichung)	(0.013)	(0.013)	(0.008)	(0.008)	(0.012)	(0.011)	(0.012)	(0.006)
Median	0.000	0.013	0.003	0.008	0.008	0.009	0.000	0.004

Die robuste gemischte Varianzanalyse zeigt eine signifikante Zweiwege-Interaktion. Zudem ist ein signifikanter Haupteffekt „Messzeitpunkt“ vorhanden (vgl. Tabelle 109). Der Haupteffekt wird jedoch aufgrund der Zweiwege-Interaktionen nicht näher analysiert.

Tabelle 109: Ergebnisse der robusten Varianzanalyse der Hilfenutzung im Schulformvergleich in der Mittelstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap Verfahren.

Effekt	$Q$	$p$
Geschlecht	1.91	0.179
Schulform	1.40	0.237
Messzeitpunkt	25.44**	< 0.001
Geschlecht*Schulform	0.51	0.446
Geschlecht*Messzeitpunkt	4.01**	< 0.001
Schulform*Messzeitpunkt	0.47	0.265
Geschlecht*Gruppe*Messzeitpunkt	0.50	0.326

Die signifikante Interaktion ist in Abbildung 65 zu erkennen. Die Geschlechter unterscheiden sich hinsichtlich der Nutzung der Hilfen zum Teil deutlich zwischen den einzelnen Schulstunden, unabhängig von der Schulform. Demnach nutzen die Schülerinnen in der zweiten und vierten Stunde häufiger eine Hilfe als Schüler. Bei den Gesamtschülerinnen ist dieser Effekt am größten.

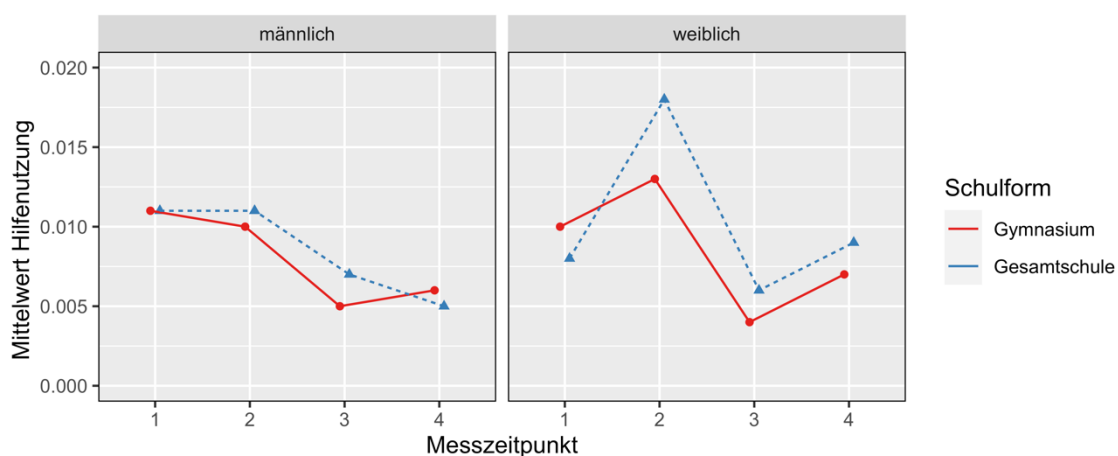


Abbildung 65: Genormte Mittelwerte für die Hilfenutzung, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Interventionsgruppe in der Mittelstufe. Die Standardabweichung ist aus Gründen der Lesbarkeit nicht gezeigt.

Die genauen Unterschiede werden über die einfachen Haupteffekte bestimmt. Für die Interaktion „Geschlecht\*Messzeitpunkt“ zeigt sich ein signifikanter Zwischensubjektfaktor für die zweite und vierte Stunde (vgl. Tabelle 110).

Tabelle 110: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Geschlecht“ der Interaktion „Geschlecht\*Messzeitpunkt“ (Lernhilfen) in der Mittelstufe.

Messzeitpunkt	<i>t</i>	<i>df</i>	<i>p</i>	$\hat{\xi}$	$M_m (SD)$	$M_w (SD)$
1	1.42	184.06	0.156	0.13	0.011 (0.014)	0.009 (0.013)
2	2.31*	183.10	0.022	0.20	0.011 (0.010)	0.015 (0.014)
3	0.09	191.93	0.932	< 0.01	0.006 (0.009)	0.005 (0.006)
4	2.84**	188.86	0.005	0.24	0.006 (0.006)	0.008 (0.007)

Der Innersubjektfaktor „Messzeitpunkt“ ist für die Schülerinnen (*robust*  $F(2.52, 242.17) = 27.16, p < 0.001$ ) und für die Schüler (*robust*  $F(2.27, 217.62) = 12.48, p < 0.001$ ) signifikant. Dabei sind bei den weiblichen Lernenden zwischen allen Stunden signifikante Unterschiede mit  $\delta_t = 0.08$  (*alle Stunden*) und  $\delta_t = 0.01$  (*erste – vierte Stunde*) feststellbar - bei den männlichen Lernenden zwischen der zweiten und dritten (vgl. Anhang 9.3.5). Die Effektstärken sind  $\delta_t = 0.21$  (*alle Stunden*) und  $\delta_t = 0.28$  (*erste – vierte Stunde*) größer als bei den Schülerinnen. Insgesamt nimmt bei den Schülern die Nutzung daher stärker ab.

#### Vergleichsgruppe

Als Datengrundlage für die Hilfenutzung dienen die Angaben aus dem Fragebogen. Die absolute Anzahl wurde anschließend durch die Anzahl der generell zur Verfügung stehenden Hilfen geteilt. Aufgrund der Verletzung der Kovarianzmatrizen des Faktors „Geschlecht“ ( $p = 0.028$ ) und der hohen Anzahl an Ausreißern (46, 14 extrem) wurde eine robuste gemischte Varianzanalyse mit Bootstrapping durchgeführt. Die deskriptive Statistik ist Tabelle 111 zu entnehmen.

Tabelle 111: Deskriptive Statistik der Hilfenutzung in der Vergleichsgruppe in der Mittelstufe.

Messzeitpunkt	Gymnasium							
	weiblich				männlich			
	1	2	3	4	1	2	3	4
Stichprobenumfang	71				78			
Mittelwert (Standardabweichung)	0.25 (0.39)	0.27 (0.32)	0.13 (0.21)	0.18 (0.24)	0.17 (0.32)	0.24 (0.37)	0.13 (0.28)	0.17 (0.38)
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Die Auswertung der Varianzanalyse zeigt zwei signifikante Haupteffekte. Die Interaktion zwischen „Geschlecht“ und „Messzeitpunkt“ ist nicht signifikant (vgl. Tabelle 112).

Tabelle 112: Ergebnisse der robusten Varianzanalyse der Hilfenutzung in der Vergleichsgruppe der Mittelstufe. Die *Q*-Werte stammen aus dem non-bootstrap Verfahren.

Effekt	<i>Q</i>	<i>p</i>
Geschlecht	4.05*	0.037
Messzeitpunkt	5.29**	0.003
Geschlecht*Messzeitpunkt	1.02	0.381

In Abbildung 66 zeigen sich die beiden Haupteffekte. Die Schülerinnen nutzen in drei der vier Stunden häufiger eine Hilfe als die Schüler.

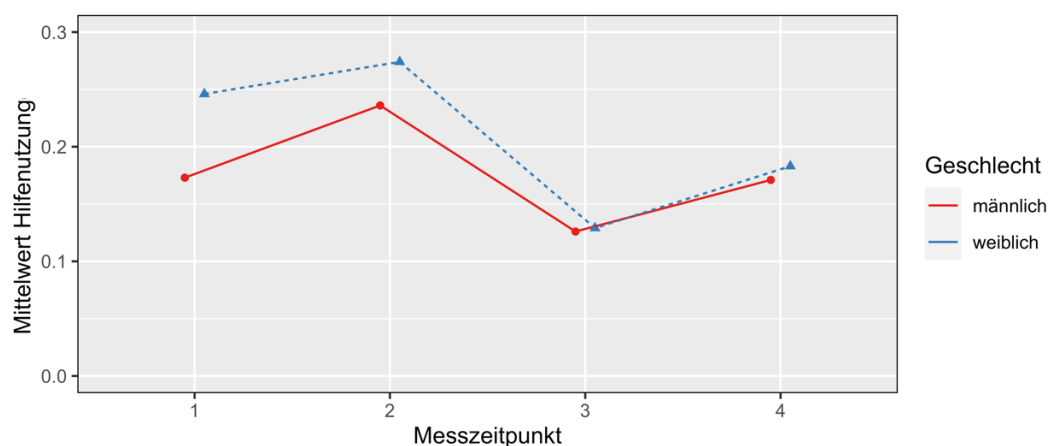


Abbildung 66: Genormte Mittelwerte für die Hilfenutzung, aufgeteilt nach Geschlecht im zeitlichen Verlauf in der Vergleichsgruppe in der Mittelstufe. Die Standardabweichung ist aus Gründen der Lesbarkeit nicht gezeigt.

Gleichzeitig nimmt insbesondere in der zweiten Stunde die Hilfenutzung zu, während sie dann wieder abfällt. Dieser Verlauf konnte bereits bei der Interventionsgruppe beobachtet werden. Die Überprüfung des signifikanten Haupteffektes „Messzeitpunkt“ zeigt mit  $\delta_t = 0.04$  (alle Stunden) und  $\delta_t = 0.04$  (erste – vierte Stunde) keine starken Unterschiede zwischen den Messzeitpunkten. Der Haupteffekt „Geschlecht“ zeigt eine mittlere Effektstärke (robust  $t(88.97) = 2.20, p = 0.030, \xi = 0.26; M_m = 0.18 \pm 0.27, M_w = 0.21 \pm 0.21$ ).

### **Vertiefungsnutzung**

#### *Interventionsgruppe*

Die Nutzung der Vertiefungsaufgaben wird lediglich in der Interventionsgruppe untersucht. Für die Analyse der Vergleichsgruppe stehen keine valide Daten zur Verfügung. Die Daten für die Nutzungshäufigkeiten der Vertiefungsaufgaben wurden, wie auch bereits die der Hilfen, aus HyperDocSystems gewonnen und anschließend aufbereitet. Die absoluten Nutzungszahlen wurden anschließend mit der zur Verfügung stehende Anzahl an Bearbeitungsminuten verrechnet, um einen besseren Vergleich zwischen den einzelnen Klassen und Stunden zu erreichen. Aufgrund der großen Streuung der Daten und der damit verbundenen hohen Abweichung von der Normalverteilung sowie dem Vorhandensein von 104 Ausreißern, wovon der überwiegende Teil als extrem anzusehen ist, wurde auf eine Inferenzanalyse verzichtet und die Daten lediglich deskriptiv analysiert. Die deskriptive Statistik ist Tabelle 113 zu entnehmen.

Tabelle 113: Deskriptive Statistik der Vertiefungsnutzung in der Interventionsgruppe der Mittelstufe.

		<b>Gymnasium</b>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		107				97			
Mittelwert		0.1	0.05	0.04	0.01	0.13	0.07	0.05	0.01
(Standardabweichung)		(0.08)	(0.05)	(0.05)	(0.03)	(0.12)	(0.06)	(0.06)	(0.04)
Median		0.09	0.05	0.04	0.00	0.09	0.06	0.04	0.00
		<b>Gesamtschule</b>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		52				63			
Mittelwert		0.12	0.04	0.05	0.01	0.09	0.05	0.06	0.01
(Standardabweichung)		(0.11)	(0.05)	(0.08)	(0.03)	(0.09)	(0.05)	(0.10)	(0.02)
Median		0.12	0.00	0.00	0.00	0.06	0.05	0.00	0.00

Einen Überblick über den Verlauf der Nutzung der Vertiefungsaufgaben und mögliche Unterschiede zwischen den Geschlechtern und Schulformen liefert Abbildung 67.

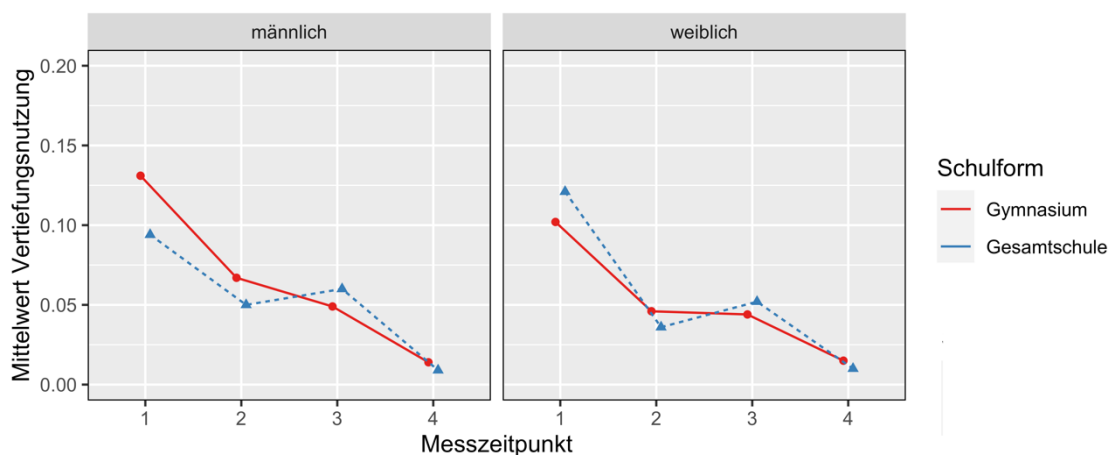


Abbildung 67: Genormte Mittelwerte für die Vertiefungsnutzung, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Interventionsgruppe in der Mittelstufe. Die Standardabweichung ist aus Gründen der Lesbarkeit nicht gezeigt.

Es ist ein deutlicher Einfluss des Messzeitpunktes zu erkennen. Demnach nimmt über beide Schulformen und Geschlechter die Nutzung der Vertiefungsaufgaben ab.

Tendenziell nutzen Schülerinnen, insbesondere am Gymnasium, die Aufgaben weniger häufig als die Schüler. Auch zwischen den Schulformen zeigen sich teilweise Unterschiede, die jedoch keinem einheitlichen Trend folgen.

#### 5.4.4 Einfluss der Einsatzdauer und des Geschlechts in der Oberstufe

##### Hilfenutzung

###### *Interventionsgruppe*

Da für die Oberstufe nicht genug Datenpunkte für einen Vergleich von Gymnasium und Gesamtschule vorhanden sind, erfolgt die Analyse der Hilfenutzung lediglich unter Beachtung des Geschlechts und der Einsatzdauer im Gymnasium. Dazu wird eine gemischte Varianzanalyse verwendet. Die Prüfung der Voraussetzungen zeigt eine Verletzung der Sphärizität ( $p < 0.001$ ) und der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ). Es sind 17 Ausreißer vorhanden, zwei davon extrem. Es wurde daher eine robuste Varianzanalyse durchgeführt. Die deskriptive Statistik ist Tabelle 114 zu entnehmen.

*Tabelle 114: Deskriptive Statistik der Hilfenutzung in der Interventionsgruppe der Oberstufe.*

Messzeitpunkt	Gymnasium							
	weiblich				männlich			
	1	2	3	4	1	2	3	4
Stichprobenumfang	46				55			
Mittelwert	0.009	0.005	0.015	0.007	0.010	0.005	0.014	0.006
(Standardabweichung)	(0.015)	(0.007)	(0.014)	(0.008)	(0.015)	(0.007)	(0.013)	(0.006)
Median	0.000	0.004	0.012	0.004	0.000	0.003	0.011	0.004

Die mixed ANOVA zeigt keine signifikante Interaktion, jedoch einen signifikanten Haupteffekt „Messzeitpunkt“ (vgl. Tabelle 115).



Tabelle 115: Ergebnisse der robusten Varianzanalyse der Hilfenutzung in der Interventionsgruppe der Oberstufe. Die Q-Werte stammen aus dem non-bootstrap Verfahren.

Effekt	Q	p
Geschlecht	0.01	0.914
Messzeitpunkt	10.00**	< 0.001
Geschlecht*Messzeitpunkt	0.14	0.948

Der Effekt ist in Abbildung 68 deutlich zu erkennen. Demnach unterscheiden sich die Stunden im Verlauf teilweise stark. Eine klare Abnahme der Hilfenutzung ist nicht feststellbar. Geringe Unterschiede zwischen den Geschlechtern sind lediglich in den letzten beiden Stunden erkennbar. Dabei verwenden Schülerinnen häufiger eine Hilfe als Schüler.

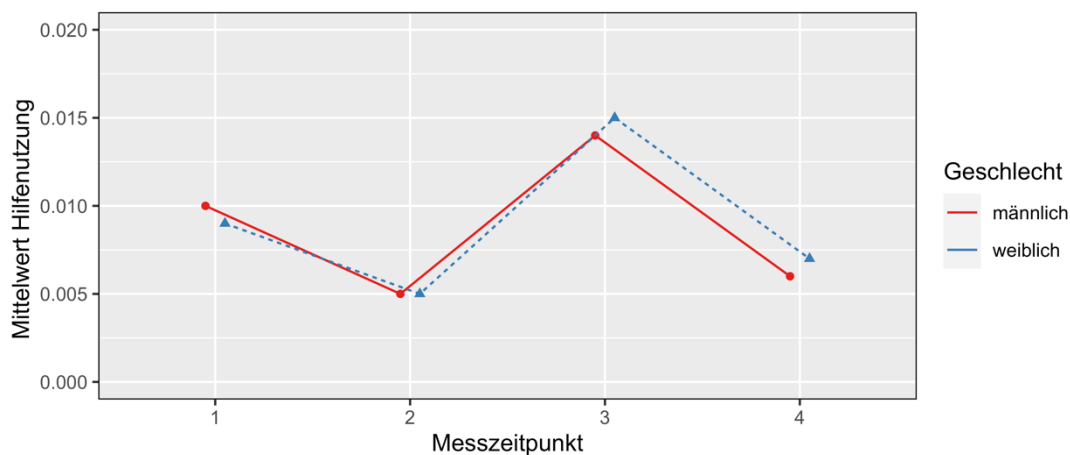


Abbildung 68: Genormte Mittelwerte für die Hilfenutzung, aufgeteilt nach Geschlecht im zeitlichen Verlauf in der Interventionsgruppe in der Oberstufe. Die Standardabweichung ist aus Gründen der Lesbarkeit nicht gezeigt.

Die Überprüfung des signifikanten Haupteffektes „Messzeitpunkt“ zeigt mit  $\delta_t = 0.06$  (alle Stunden) und  $\delta_t = 0.01$  (erste – vierte Stunde) lediglich einen schwachen Effekt zwischen der ersten und vierten Stunde.

### Vergleichsgruppe

Wie bereits in der Mittelstufe werden für die Nutzungen die Angaben aus dem Fragebogen verwendet. Danach erfolgt eine Verrechnung mit der Anzahl der zur Verfügung stehenden Hilfen in den einzelnen Stunden. Aufgrund der Verletzung der Sphärizität ( $p = 0.004$ ), der Normalverteilung (QQ-Plot, Shapiro-Wilk-Test:  $p < 0.001$ ) und der Homogenität der Kovarianzmatrizen ( $p < 0.001$ ) wurde eine robuste gemischte Varianzanalyse durchgeführt. Verstärkt wird diese Annahme durch 21 Ausreißer, wovon elf als extrem

einzustufen sind. Zwei Ausreißer wurden aufgrund der sehr extremen Werte aus der weiteren Analyse ausgeschlossen. Die deskriptive Statistik ist in Tabelle 116 aufgeführt.

*Tabelle 116: Deskriptive Statistik der Hilfenutzung in der Vergleichsgruppe der Oberstufe.*

Messzeitpunkt	<i>Gymnasium</i>							
	weiblich				männlich			
	1	2	3	4	1	2	3	4
Stichprobenumfang	24				24			
Mittelwert	0.10	0.08	0.11	0.10	0.21	0.11	0.17	0.09
Standardabweichung	(0.23)	(0.18)	(0.14)	(0.17)	(0.32)	(0.15)	(0.30)	(0.19)
Median	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Die Analyse zeigt keine signifikanten Effekte (vgl. Tabelle 117). Aufgrund der vielen Nullwerte bei der Hilfenutzung konnte kein Bootstrapping angewandt werden.

*Tabelle 117: Ergebnisse der robusten Varianzanalyse in der Vergleichsgruppe der Oberstufe.*

<i>Effekt</i>	<i>Q</i>	<i>p</i>
Geschlecht	0.20	0.655
Messzeitpunkt	1.84	0.173
Geschlecht*Messzeitpunkt	0.86	0.479

Die graphische Darstellung der Mittelwerte zeigt einen ähnlichen Verlauf wie auch die der Interventionsgruppe (vgl. Abbildung 69). Demnach nimmt die Hilfenutzung tendenziell ab, wobei es in der dritten Stunde zu einer starken Nutzung kommt. Insgesamt nutzen die Schüler in der ersten, zweiten und dritten Stunde häufiger eine Hilfe als die Schülerinnen.

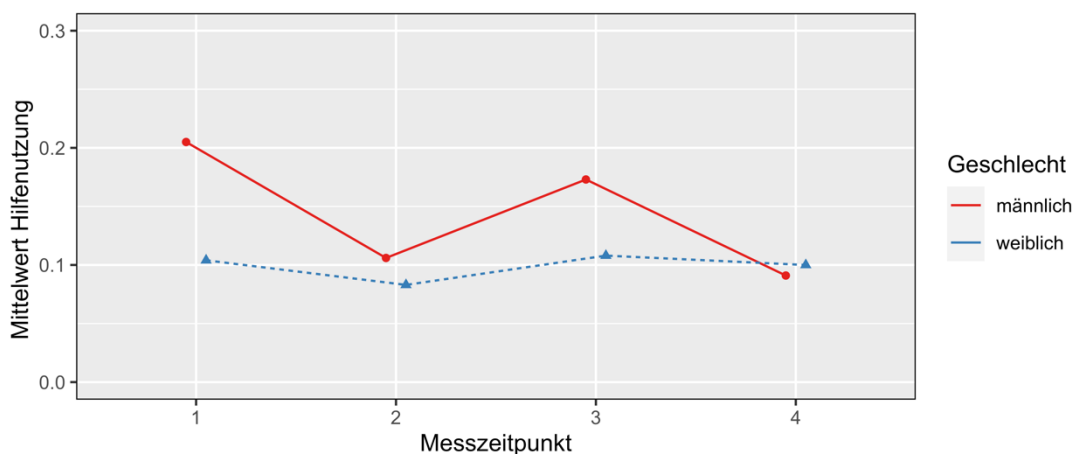


Abbildung 69: Genormte Mittelwerte für die Hilfenutzung, aufgeteilt nach Geschlecht im zeitlichen Verlauf in der Vergleichsgruppe in der Oberstufe. Die Standardabweichung ist aus Gründen der Lesbarkeit nicht gezeigt.

## Vertiefungsnutzung

### *Interventionsgruppe*

Wie bereits in der Mittelstufe wird in der Oberstufe aufgrund der zur Verfügung stehenden Daten lediglich die Interventionsgruppe betrachtet. Die Nutzungshäufigkeiten der Vertiefungsaufgaben wurden ebenfalls aus HyperDocSystems gewonnen und anschließend ausgewertet. Die absolute Zahl der Nutzungen wurde durch die verfügbare Bearbeitungszeit geteilt, um einen besseren Vergleich zwischen den Stunden zu ermöglichen. Da auch in der Stichprobe der Oberstufe etwa ein Drittel der Werte als Ausreißer eingestuft werden, erfolgt die Analyse lediglich deskriptiv (vgl. Tabelle 118). Hierbei werden nur die Lerngruppen des Gymnasiums betrachtet.

Tabelle 118: Deskriptive Statistik der Vertiefungsnutzung in der Oberstufe.

Messzeitpunkt	<i>Gymnasium</i>							
	weiblich				männlich			
	1	2	3	4	1	2	3	4
Stichprobenumfang	46				56			
Mittelwert	0.11	0.02	0.01	0.02	0.16	0.03	0.01	0.03
Standardabweichung	(0.11)	(0.04)	(0.01)	(0.03)	(0.12)	(0.04)	(0.02)	(0.05)
Median	0.08	0.000	0.000	0.000	0.14	0.00	0.00	0.00

Abbildung 70 zeigt die Mittelwerte, aufgeteilt nach Geschlecht, im zeitlichen Verlauf. Zu sehen ist eine Abnahme der Nutzungen mit der Zeit, wobei die Schüler:innen in der

vierten Stunde wieder etwas häufiger eine Aufgabe aufrufen. Insgesamt liegen die Nutzungen der Schülerinnen unter denen der Schüler, wobei die Unterschiede mit Ausnahme von der ersten Stunde sehr gering sind.

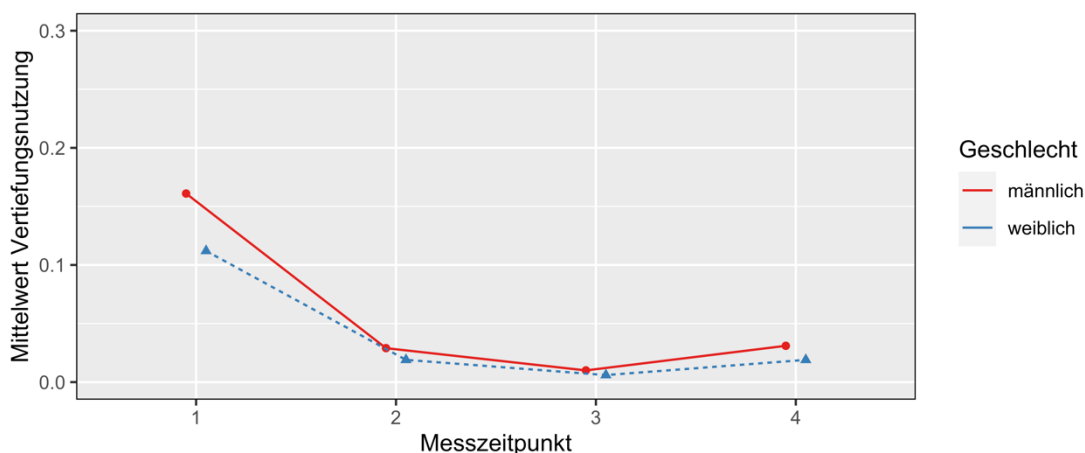


Abbildung 70: Genormte Mittelwerte für die Vertiefungsnutzung, aufgeteilt nach Geschlecht im zeitlichen Verlauf in der Interventionsgruppe in der Oberstufe. Die Standardabweichung ist aus Gründen der Lesbarkeit nicht gezeigt.

#### 5.4.5 Zusammenhangsanalysen zu weiteren Variablen

In diesem Kapitel erfolgen die Zusammenhangsanalysen zu weiteren Variablen. Das betrifft die Variablen zu Motivation nach dem Intrinsic Motivation Inventory, die kognitive Last und den Leistungsstand, der zu Beginn der Erhebung über eine Selbsteinschätzung erfragt wurde. Die Auswertung betrifft lediglich die Interventionsgruppe, da hierbei die Nutzung im Rahmen von HyperDocSystems im Vordergrund steht. Die Korrelationen zwischen den Variablen wurde für die erste Stunde und die vierte Stunde berechnet, um mögliche Veränderungen über den Messzeitpunkt feststellen zu können. Zwischen den Schulformen wird keine Unterscheidung getroffen. Es erfolgt lediglich eine Unterteilung in Mittel- und Oberstufe. Die Nutzungsstatistik für die Hilfen umfasst die Einberechnung der zur Verfügung stehenden Zeit pro Lerngruppe. Die Zusammenhänge wurden mittels Spearmans rho analysiert, da die Daten Ausreißer beinhalten. Für die Nutzung der Vertiefungsaufgaben wurden keine Korrelationen berechnet, da die Nutzungshäufigkeit sehr gering ist und damit keine genauen Aussagen gemacht werden können.

## Motivation und Interesse

### *Mittelstufe*

In der Mittelstufe zeigen sich keine signifikanten Korrelationen zwischen den Prädikatoren der intrinsischen Motivation und der Nutzung von Lernhilfen. Die Korrelationskoeffizienten sind ebenfalls sehr gering (vgl. Tabelle 119).

*Tabelle 119: Korrelationen (Spearman's rho) zwischen den Faktoren der intrinsischen Motivation und der Nutzung der Lernhilfen in der Mittelstufe (n = 316) in der ersten und vierten Stunde.*

	<i>Stunde 1</i>	<i>Stunde 4</i>
Interesse	$p = 0.278, \rho = -0.06$	$p = 0.747, \rho = 0.01$
Kompetenzerleben	$p = 0.896, \rho = 0.01$	$p = 0.479, \rho = 0.04$
Wahrgenommene Wahlfreiheit	$p = 0.097, \rho = -0.09$	$p = 0.266, \rho = 0.06$
Druck	$p = 0.814, \rho = -0.01$	$p = 0.711, \rho = -0.02$

### *Oberstufe*

In der Oberstufe sind ebenfalls nahezu alle Korrelationen nicht-signifikant. Die einzige signifikante Korrelation ist zwischen dem Druck und der Nutzung der Lernhilfen. Die Stärke ist mit  $\rho = 0.23$  als schwach einzustufen (vgl. Tabelle 120).

*Tabelle 120: Korrelationen (Spearman's rho) zwischen den Faktoren der intrinsischen Motivation und der Nutzung der Hilfen in der Oberstufe (n = 109) in der ersten und vierten Stunde.*

	<i>Stunde 1</i>	<i>Stunde 4</i>
Interesse	$p = 0.725, \rho = -0.03$	$p = 0.464, \rho = 0.07$
Kompetenzerleben	$p = 0.368, \rho = 0.09$	$p = 0.504, \rho = -0.06$
Wahrgenommene Wahlfreiheit	$p = 0.151, \rho = 0.14$	$p = 0.753, \rho = -0.03$
Druck	$p = 0.015, \rho = 0.23^*$	$p = 0.476, \rho = 0.07$

## Cognitive Load

### *Mittelstufe*

Bei Betrachtung der einzelnen Arten des Cognitive Load und des Mental Effort zeigt sich lediglich in der ersten Stunde beim Germane Load ( $\rho = -0.12$ ) und in der vierten Stunde bei der mentalen Anstrengung eine signifikante Korrelation mit einer schwachen Stärke von  $\rho = 0.23$  (vgl. Tabelle 121).

Tabelle 121: Korrelationen (Spearman's rho) zwischen den Faktoren des Cognitive Load und der Nutzung der Hilfen in der Mittelstufe ( $n = 312$ ) in der ersten und vierten Stunde.

	Stunde 1	Stunde 4
Intrinsic Load	$p = 0.622, \rho = 0.03$	$p = 0.530, \rho = 0.04$
Extraneous Load	$p = 0.744, \rho = 0.02$	$p = 0.966, \rho < 0.01$
Germane Load	$p = 0.033, \rho = -0.12^*$	$p = 0.183, \rho = 0.08$
Mental Effort	$p = 0.737, \rho = -0.02$	$p < 0.001, \rho = 0.23^{**}$

### Oberstufe

In der Oberstufe liegen mehrere signifikante Zusammenhänge zwischen der Nutzung der Lernhilfen und den verschiedenen Loadtypen vor. In der ersten Stunde gibt es einen signifikanten negativen Zusammenhang zum Intrinsic Load ( $\rho = -0.28$ ) und zur mentalen Anstrengung ( $\rho = -0.23$ ). In der vierten Stunde bleiben die Korrelationen signifikant, allerdings werden die Zusammenhänge positiv (vgl. Tabelle 122).

Tabelle 122: Korrelationen (Spearman's rho) zwischen den Faktoren des Cognitive Load und der Nutzung der Hilfen in der Oberstufe ( $n = 109$ ) in der ersten und vierten Stunde.

	Stunde 1	Stunde 4
Intrinsic Load	$p = 0.003, \rho = -0.28^{**}$	$p = 0.032, \rho = 0.20^*$
Extraneous Load	$p = 0.607, \rho = -0.05$	$p = 0.415, \rho = 0.07$
Germane Load	$p = 0.567, \rho = -0.06$	$p = 0.112, \rho = 0.15$
Mental Effort	$p = 0.018, \rho = -0.23^*$	$p < 0.001, \rho = 0.35^{**}$

## Leistungsstand

### Mittelstufe

In der Mittelstufe sind keine signifikanten Zusammenhänge zwischen der Nutzung von Lernhilfen und der Selbsteinschätzung bezüglich der eigenen Fähigkeiten im Fach Chemie, den Naturwissenschaften und der schulischen Leistungen vorhanden (vgl. Tabelle 123).

*Tabelle 123: Korrelationen (Spearman's rho) zwischen der Selbsteinschätzung der Schulleistungen und der Nutzung der Hilfen in der Mittelstufe (n = 320) in der ersten und vierten Stunde.*

	<i>Stunde 1</i>	<i>Stunde 4</i>
Fähigkeiten Chemie	$p = 0.419, \rho = -0.05$	$p = 0.713, \rho = 0.02$
Fähigkeiten Naturwissenschaften	$p = 0.342, \rho = -0.05$	$p = 0.510, \rho = -0.04$
Fähigkeiten Schule	$p = 0.902, \rho < 0.01$	$p = 0.725, \rho = -0.02$

### *Oberstufe*

In der Oberstufe sind alle Korrelationen zwischen den Fähigkeiten in den Fächern und der Verwendung von Hilfen nicht signifikant. Die Korrelation zwischen den Fähigkeiten im Fach Chemie und der Nutzung der Hilfen ist in der ersten Stunde nur knapp nicht signifikant. Die Stärke des Zusammenhangs ist jedoch schwach (vgl. Tabelle 124).

*Tabelle 124: Korrelationen (Spearman's rho) zwischen der Selbsteinschätzung der Schulleistungen und der Nutzung der Hilfen in der Oberstufe (n = 110) in der ersten und vierten Stunde.*

	<i>Stunde 1</i>	<i>Stunde 4</i>
Fähigkeiten Chemie	$p = 0.081, \rho = 0.17$	$p = 0.825, \rho = 0.02$
Fähigkeiten Naturwissenschaften	$p = 0.832, \rho = -0.02$	$p = 0.174, \rho = -0.13$
Fähigkeiten Schule	$p = 0.456, \rho = 0.07$	$p = 0.659, \rho = -0.04$

## **5.5 Variablen in Zusammenhang mit digitalem Lernen**

In dem nachfolgenden Kapitel werden die mit digitalen Medien verbundenen Variablen untersucht. Dazu gehören der Cognitive Load und die Variablen im Umgang mit dem Lernen mit Tablets.

### **5.5.1 Cognitive Load und Mental Effort im Gruppenvergleich**

Die Auswertung der kognitiven Last erfolgt nach den drei Grundtypen Intrinsische Load, Extraneous Load und Germane Load. Die Auswertung bezieht sich auf den Vergleich zwischen der Interventions- und Vergleichsgruppe und umfasst daher nur gymnasiale Schulklassen. Die Berücksichtigung des Faktors „Schulform“ findet im Kapitel 5.5.3 statt. Zusätzlich wird der Faktor „Geschlecht“ in einer gemischten Varianzanalyse betrachtet. Bei jedem Loadtyp erfolgt zunächst die Prüfung der Voraussetzungen dieser Analyseverfahren.

**Intrinsic Load: Geschlecht, Gruppe, Messzeitpunkt***Mittelstufe*

Die Prüfung der Voraussetzungen ergibt eine Verletzung der Sphärizität ( $p < 0.001$ ), der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ) und der Homogenität der Kovarianzmatrizen für beide Faktoren (Gruppe:  $p = 0.034$ , Geschlecht:  $p = 0.033$ ). Zudem sind vier normale Ausreißer vorhanden. Aus diesen Gründen wurde eine robuste mixed ANOVA mit Bootstrapping eingesetzt. Die deskriptive Statistik ist in Tabelle 125 dargestellt.

*Tabelle 125: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Intrinsic Load in der Mittelstufe.*

	<i>Interventionsgruppe</i>								
	weiblich				männlich				
	Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang		107				97			
Mittelwert (Standardabweichung)		1.53 (0.52)	1.97 (0.78)	2.26 (0.81)	2.45 (0.92)	1.58 (0.57)	1.93 (0.76)	2.05 (0.78)	2.15 (0.92)
Median		1.50	2.00	2.00	2.50	1.50	2.00	2.00	2.00
	<i>Vergleichsgruppe</i>								
	weiblich				männlich				
	Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang		70				76			
Mittelwert (Standardabweichung)		1.52 (0.57)	2.29 (0.87)	2.14 (0.90)	2.77 (0.94)	1.53 (0.62)	2.09 (0.81)	2.12 (0.79)	2.36 (0.90)
Median		1.50	2.00	2.00	3.00	1.50	2.00	2.00	2.50

Die gemischte Varianzanalyse zeigt zwei signifikante Zweifache-Interaktionen und zwei signifikante Haupteffekte (vgl. Tabelle 126). Im Folgenden werden daher nun die zwei Interaktionen näher analysiert.



Tabelle 126: Ergebnisse der robusten Varianzanalyse des Intrinsic Load in der Mittelstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap Verfahren.

Effekt	$Q$	$p$
Geschlecht	3.50	0.070
Gruppe	2.08	0.137
Messzeitpunkt	71.33**	< 0.001
Geschlecht*Gruppe	< 0.01	0.930
Geschlecht*Messzeitpunkt	2.90**	0.003
Gruppe*Messzeitpunkt	5.16**	< 0.001
Geschlecht*Gruppe*Messzeitpunkt	1.22	0.152

Die beiden Zweifache-Interaktionen zeigen sich zunächst deutlich in Abbildung 71.

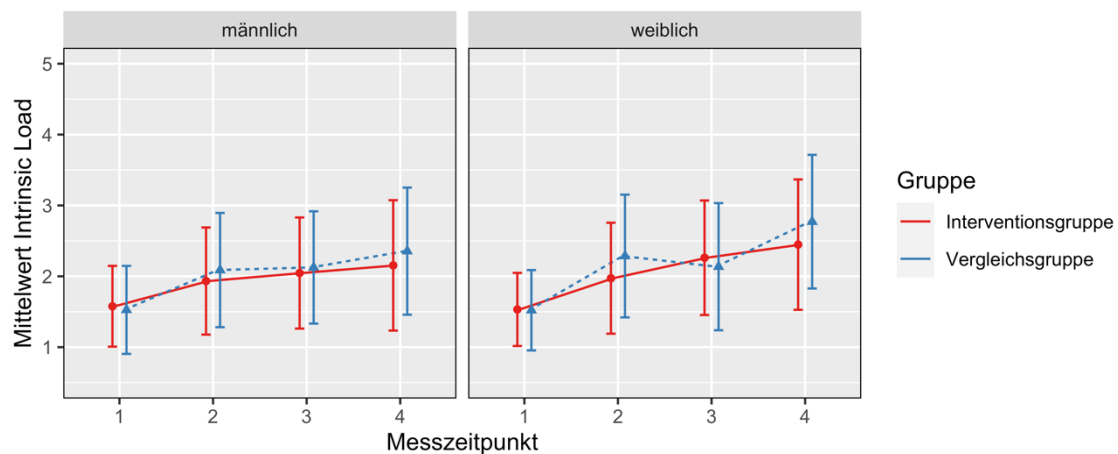


Abbildung 71: Mittelwerte mit Standardabweichung des Intrinsic Load beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.

Demnach liegt bei beiden Geschlechtern der Intrinsic Load bei der Vergleichsgruppe in mindestens zwei der vier Stunden über dem der Interventionsgruppe, wobei bei beiden Gruppen die Last über die Stunden nahezu kontinuierlich ansteigt. Eine Analyse des Zwischensubjektfaktors zeigt signifikante Unterschiede in der zweiten und vierten Stunde (vgl. Tabelle 127).

Tabelle 127: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Gruppe“ der Interaktion „Gruppe\*Messzeitpunkt“ (Intrinsic Load) in der Mittelstufe.

Messzeitpunkt	<i>t</i>	<i>df</i>	<i>p</i>	$\hat{\xi}$	<i>M<sub>I</sub></i> (SD)	<i>M<sub>V</sub></i> (SD)
1	1.03	184.43	0.305	0.08	1.55 (0.54)	1.52 (0.59)
2	2.77**	184.98	0.006	0.19	1.95 (0.77)	2.18 (0.84)
3	0.44	158.22	0.657	0.05	2.16 (0.80)	2.13 (0.84)
4	2.06*	165.14	0.041	0.18	2.31 (0.93)	2.56 (0.94)

Die Auswertung des Innersubjektfaktors zeigt sowohl für die Interventionsgruppe (*robust*  $F(2.92, 359.49) = 50.41, p < 0.001$ ) als auch für die Vergleichsgruppe (*robust*  $F(2.72, 236.69) = 54.49, p < 0.001$ ) einen signifikanten Effekt. In beiden Gruppen liegen, mit Ausnahme von der dritten Stunde in der Vergleichsgruppe, signifikante Unterschiede vor (vgl. Anhang 9.3.9). Die Effektstärken sind in der Interventionsgruppe mit  $\delta_t = 0.42$  (alle Stunden) und  $\delta_t = 0.75$  (erste – vierte Stunde) und in der Vergleichsgruppe mit  $\delta_t = 0.50$  (alle Stunden) und  $\delta_t = 0.88$  (erste – vierte Stunde) als mittel bis stark einzuordnen. Dabei nimmt in der Vergleichsgruppe der Intrinsic Load stärker zu.

Aus Abbildung 71 geht außerdem hervor, dass Schülerinnen ab der dritten Stunde eine höhere kognitive Last empfinden als Schüler. Dieser Effekt zeigt sich in der Vergleichsgruppe weniger linear als in der Interventionsgruppe. Die Auswertung der Interaktion „Geschlecht\*Messzeitpunkt“ zeigt für den Zwischensubjektfaktor lediglich für die letzte Stunde einen signifikanten Unterschied (vgl. Tabelle 128).

Tabelle 128: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Geschlecht“ der Interaktion „Geschlecht\*Messzeitpunkt“ (Intrinsic Load) in der Mittelstufe.

Messzeitpunkt	<i>t</i>	<i>df</i>	<i>p</i>	$\hat{\xi}$	<i>M<sub>w</sub></i> (SD)	<i>M<sub>m</sub></i> (SD)
1	< 0.01	209.43	0.996	< 0.01	1.53 (0.54)	1.56 (0.59)
2	0.82	209.84	0.413	0.05	2.10 (0.83)	2.00 (0.78)
3	0.95	199.69	0.341	0.09	2.21 (0.84)	2.08 (0.79)
4	3.29**	209.04	0.001	0.26	2.58 (0.94)	2.24 (0.91)

Der Innersubjektfaktor „Messzeitpunkt“ ist für weibliche (*robust*  $F(2.88, 305.69) = 71.70, p < 0.001$ ) und männliche Lernende (*robust*  $F(2.8, 291.21) = 35.12, p < 0.001$ ) signifikant. Es liegen dabei insbesondere zwischen den ersten beiden Stunden signifikante Unterschiede vor (vgl. Anhang 9.3.9). Eine Analyse der Haupteffekte erfolgt aufgrund der signifikanten Zweifache-Interaktionen nicht.

Die Effektstärken sind für Schüler mit  $\delta_i = 0.37$  (*alle Stunden*) und  $\delta_i = 0.68$  (*erste – vierte Stunde*) und für Schülerinnen mit  $\delta_i = 0.55$  (*alle Stunden*) und  $\delta_i = 0.90$  (*erste – vierte Stunde*) erneut als mittel bis stark einzuordnen. Bei Schülerinnen nimmt der Intrinsic Load stärker zu als bei den männlichen Lernenden.

### *Oberstufe*

Auch für die Oberstufe wurden zunächst die Voraussetzungen der gemischten Varianzanalyse geprüft. Es ergibt sich eine Verletzung der Sphärizität ( $p < 0.001$ ) und der Normalverteilung (*Shapiro-Wilk-Test*:  $p < 0.001$ ). In der Stichprobe sind zwölf normale Ausreißer vorhanden, weshalb eine robuste gemischte ANOVA durchgeführt wurde. Die deskriptive Statistik ist in Tabelle 129 aufgeführt.

*Tabelle 129: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Intrinsic Load in der Oberstufe.*

	<i>Interventionsgruppe</i>							
	weiblich				männlich			
	1	2	3	4	1	2	3	4
Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang	45				56			
Mittelwert	1.48	2.00	2.81	2.28	1.35	1.77	2.49	2.00
(Standardabweichung)	(0.58)	(0.81)	(0.89)	(0.67)	(0.46)	(0.62)	(0.77)	(0.75)
Median	1.50	2.00	3.00	2.50	1.00	1.50	2.50	2.00
	<i>Vergleichsgruppe</i>							
	weiblich				männlich			
	1	2	3	4	1	2	3	4
Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang	23				24			
Mittelwert	1.37	2.11	2.63	2.3	1.46	2.1	2.67	2.02
(Standardabweichung)	(0.43)	(0.83)	(0.80)	(0.73)	(0.53)	(0.68)	(0.94)	(0.73)
Median	1.50	2.00	3.00	2.00	1.50	2.00	2.50	2.00

Aufgrund der geringen Teilstichproben wurde die ANOVA mit Bootstrapping durchgeführt. Die Ergebnisse sind in Tabelle 130 aufgeführt.

Tabelle 130: Ergebnisse der robusten Varianzanalyse des Intrinsic Load in der Oberstufe. Die Q-Werte stammen aus dem non-bootstrap Verfahren.

Effekt	Q	p
Geschlecht	2.08	0.155
Gruppe	0.40	0.487
Messzeitpunkt	47.74**	< 0.001
Geschlecht*Gruppe	0.99	0.301
Geschlecht*Messzeitpunkt	0.57	0.247
Gruppe*Messzeitpunkt	0.46	0.332
Geschlecht*Gruppe*Messzeitpunkt	0.24	0.603

Es sind keine signifikanten Interaktionen vorhanden. Die Haupteffekte zeigen lediglich den Effekt „Messzeitpunkt“ als signifikant. In Abbildung 72 sind die Mittelwerte der Gruppen aufgetragen.

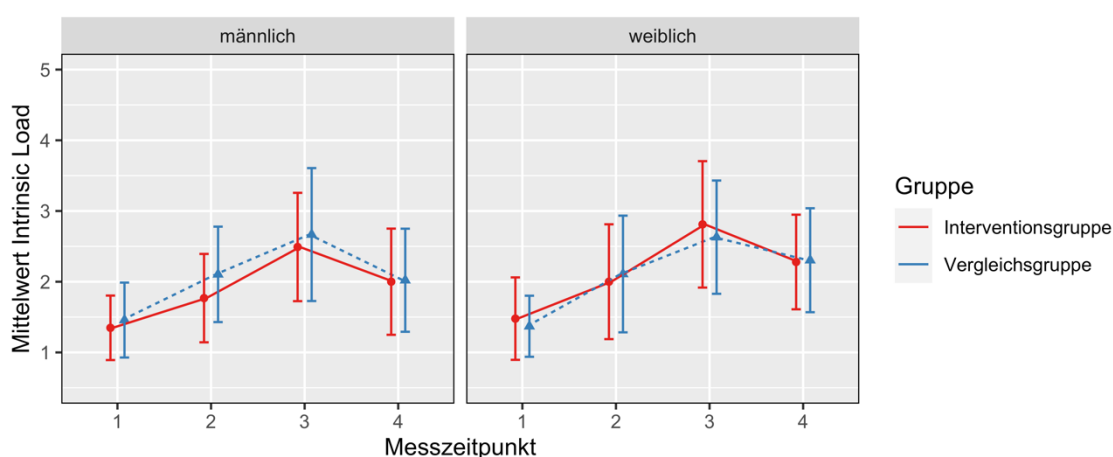


Abbildung 72: Mittelwerte mit Standardabweichung des Intrinsic Load beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe.

Demnach steigt der Intrinsic Load kontinuierlich an, bis er in der dritten Stunde einen Höhepunkt erreicht und wieder leicht abfällt. Dieser Verlauf ist bei beiden Geschlechtern und beiden Gruppen beobachtbar. Bei den Schülern liegen die Werte der Vergleichsgruppe in der zweiten und dritten Stunde über denen der Interventionsgruppe, bei den Schülerinnen ist dieser Unterschied nicht zu beobachten. Der Haupteffekt „Messzeitpunkt“ zeigt mit  $\delta_t = 0.52$  (alle Stunden) und  $\delta_t = 0.75$  (erste – vierte Stunde) mittlere bis starke Effektstärken.

**Extraneous Load: Geschlecht, Gruppe, Messzeitpunkt***Mittelstufe*

Der Extraneous Load wurde ebenfalls über eine robuste gemischte Varianzanalyse ausgewertet, da eine Verletzung der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ), der Varianzhomogenität in der zweiten ( $p = 0.002$ ) und vierten ( $p = 0.021$ ) Stunde und der Homogenität der Kovarianzmatrizen für den Faktor *Gruppe* ( $p < 0.001$ ) vorliegt. Zudem sind vierzehn normale Ausreißer vorhanden. Die deskriptive Analyse ist Tabelle 131 zu entnehmen.

*Tabelle 131: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Extraneous Load in der Mittelstufe.*

	<i>Interventionsgruppe</i>								
	weiblich				männlich				
	Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang		107				97			
Mittelwert (Standardabweichung)		1.54 (0.63)	1.71 (0.64)	2.13 (0.88)	2.05 (0.86)	1.64 (0.68)	1.77 (0.84)	1.75 (0.80)	1.97 (0.87)
Median		1.50	1.50	2.00	2.00	1.50	1.50	1.50	2.00
	<i>Vergleichsgruppe</i>								
	weiblich				männlich				
	Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang		70				76			
Mittelwert (Standardabweichung)		1.72 (0.88)	2.01 (0.95)	2.12 (0.91)	2.41 (1.08)	1.60 (0.75)	2.05 (1.02)	2.00 (0.88)	2.22 (0.88)
Median		1.50	2.00	2.00	2.50	1.00	2.00	2.00	2.25

Die anschließende Varianzanalyse zeigt, wie auch bereits beim Intrinsic Load, zwei signifikante Zweifache-Interaktionen (vgl. Tabelle 132). Es wurden daher die einfachen Haupteffekte dieser Interaktionen geprüft.

Tabelle 132: Ergebnisse der robusten Varianzanalyse des Extraneous Load in der Mittelstufe.  
Die Q-Werte stammen aus dem non-bootstrap Verfahren.

Effekt	Q	p
Geschlecht	1.41	0.247
Gruppe	5.40*	0.020
Messzeitpunkt	27.39**	< 0.001
Geschlecht*Gruppe	0.05	0.835
Geschlecht*Messzeitpunkt	2.27*	0.020
Gruppe*Messzeitpunkt	2.56*	0.012
Geschlecht*Gruppe*Messzeitpunkt	0.82	0.298

Aus Abbildung 73 lassen sich die Interaktionen ablesen. Es fällt auf, dass die Mittelwerte insgesamt niedriger sind als beim Intrinsic Load. In der Abbildung liegen die Mittelwerte der Vergleichsgruppe in drei der vier Stunden über denen der Interventionsgruppe. Zudem erfahren auch hier die Schülerinnen eine höhere kognitive Last als Schüler.

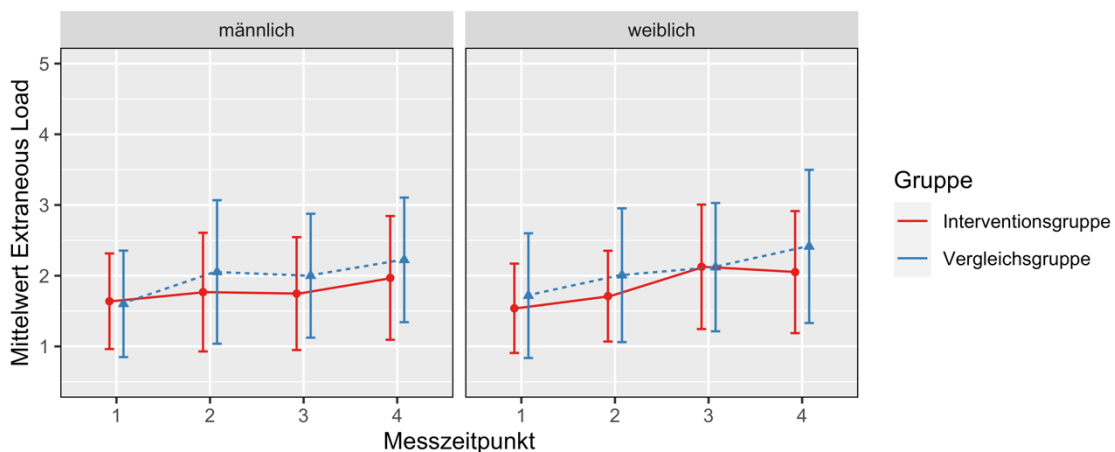


Abbildung 73: Mittelwerte mit Standardabweichung des Extraneous Load beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.

Zunächst wurde die Interaktion „Geschlecht\*Messzeitpunkt“ überprüft. Eine Analyse des Zwischensubjekteffekts zeigt lediglich für die dritte Stunde einen signifikanten Unterschied (vgl. Tabelle 133).

Tabelle 133: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Geschlecht“ der Interaktion „Geschlecht\*Messzeitpunkt“ (Extraneous Load) in der Mittelstufe.

Messzeitpunkt	<i>t</i>	<i>df</i>	<i>p</i>	$\hat{\xi}$	<i>M<sub>w</sub></i> ( <i>SD</i> )	<i>M<sub>m</sub></i> ( <i>SD</i> )
1	0.52	191.64	0.60	0.05	1.61 (0.74)	1.62 (0.71)
2	0.16	195.60	0.87	0.02	1.83 (0.79)	1.89 (0.93)
3	3.05	204.11	0.003**	0.23	2.12 (0.89)	1.86 (0.84)
4	0.61	209.14	0.540	0.05	2.19 (0.97)	2.08 (0.88)

Unter Betrachtung des Innersubjektfaktors zeigt sich ein signifikantes Ergebnis für das weibliche (*robust F*(2.85, 302.38) = 31.10, *p* < 0.001) und das männliche (*robust F*(2.95, 307.02) = 13.52, *p* < 0.001) Geschlecht. Bei den Schülerinnen liegen die Unterschiede insbesondere zwischen den ersten drei Stunden, während es bei den Schülern lediglich zwischen der dritten und vierten Stunde Unterschiede gibt (vgl. Anhang 9.3.9). Die Effektstärken lauten wie folgt: Schülerinnen -  $\delta_t = 0.35$  (alle Stunden);  $\delta_t = 0.47$  (erste – vierte Stunde) und Schüler -  $\delta_t = 0.22$  (alle Stunden);  $\delta_t = 0.50$  (erste – vierte Stunde). Sie sind insgesamt als schwach bis mittelstark anzusehen.

Anschließend wurde die Interaktion „Gruppe\*Messzeitpunkt“ analysiert. Hierbei zeigen sich in der zweiten und vierten Stunde signifikante Unterschiede zwischen der Interventions- und Vergleichsgruppe (vgl. Tabelle 134). Die Prüfung des Innersubjektfaktors zeigt für beide Gruppen ein signifikantes Ergebnis (Interventionsgruppe: *robust F*(2.77, 340.75) = 16.02, *p* < 0.001,  $\delta_t = 0.26$  (alle Stunden);  $\delta_t = 0.45$  (erste – vierte Stunde)), Untersuchungsgruppe: *robust F*(2.89, 251.70) = 24.81, *p* < 0.001,  $\delta_t = 0.34$  (alle Stunden);  $\delta_t = 0.57$  (erste – vierte Stunde)). Dabei steigt bei der Vergleichsgruppe der Extraneous Load über den Einsatz stärker an, was sich auch in den Effektstärken widerspiegelt.

Tabelle 134: Untersuchung des einfachen Haupteffektes des Zwischensubjektfaktors bezüglich der Interaktion „Gruppe\*Messzeitpunkt“ (Extraneous Load) in der Mittelstufe mittels robusten *t*-Tests.

Messzeitpunkt	<i>t</i>	<i>df</i>	<i>p</i>	$\hat{\xi}$	<i>M<sub>I</sub></i> ( <i>SD</i> )	<i>M<sub>V</sub></i> ( <i>SD</i> )
1	0.09	147.02	0.932	0.02	1.59 (0.65)	1.66 (0.82)
2	2.15	149.53	0.033*	0.21	1.74 (0.74)	2.03 (0.98)
3	1.26	158.51	0.208	0.10	1.95 (0.86)	2.06 (0.89)
4	3.12	200.77	0.002**	0.23	2.01 (0.87)	2.32 (0.98)

*Oberstufe*

Die Daten der Oberstufe werden ebenfalls über eine gemischte Varianzanalyse ausgewertet. Dazu wurden zunächst die Voraussetzungen geprüft. Es zeigt sich eine Verletzung der Sphärizität ( $p < 0.001$ ) und der Normalverteilung (*QQ-Plot, Shapiro-Wilk-Test:  $p < 0.001$* ). Zudem sind 21 Ausreißer vorhanden, von denen vier als extrem einzustufen sind. Die Varianzanalyse wurde daher robust mit Bootstrapping durchgeführt. Die deskriptive Statistik ist Tabelle 135 zu entnehmen.

*Tabelle 135: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Extraneous Load in der Oberstufe.*

	<i>Interventionsgruppe</i>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	45				56			
Mittelwert (Standardabweichung)	1.36 (0.52)	2.01 (0.84)	2.5 (1.08)	1.89 (0.83)	1.39 (0.53)	1.71 (0.72)	2.14 (0.87)	2.03 (0.91)
Median	1.00	2.00	2.50	2.00	1.00	1.50	2.00	2.00
	<i>Vergleichsgruppe</i>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	23				24			
Mittelwert (Standardabweichung)	1.33 (0.68)	1.98 (0.80)	2.41 (0.79)	1.89 (0.77)	1.44 (0.71)	2.19 (0.92)	2.58 (1.05)	1.92 (0.75)
Median	1.00	2.00	2.50	1.50	1.00	2.00	2.50	2.00

Die mixed ANOVA zeigt keine signifikanten Interaktionen (vgl. Tabelle 136). Wie bereits bei dem Intrinsic Load zeigt sich ein signifikanter Haupteffekt des Messzeitpunkts.



Tabelle 136: Ergebnisse der robusten Varianzanalyse des Extraneous Load in der Oberstufe  
Die Q-Werte stammen aus dem non-bootstrap Verfahren.

Effekt	Q	p
Geschlecht	0.02	0.858
Gruppe	0.84	0.349
Messzeitpunkt	24.52**	< 0.001
Geschlecht*Gruppe	1.37	0.282
Geschlecht*Messzeitpunkt	0.36	0.476
Gruppe*Messzeitpunkt	0.71	0.174
Geschlecht*Gruppe*Messzeitpunkt	0.38	0.469

In Abbildung 74 sind die Mittelwerte über die Einsatzdauer aufgeführt.

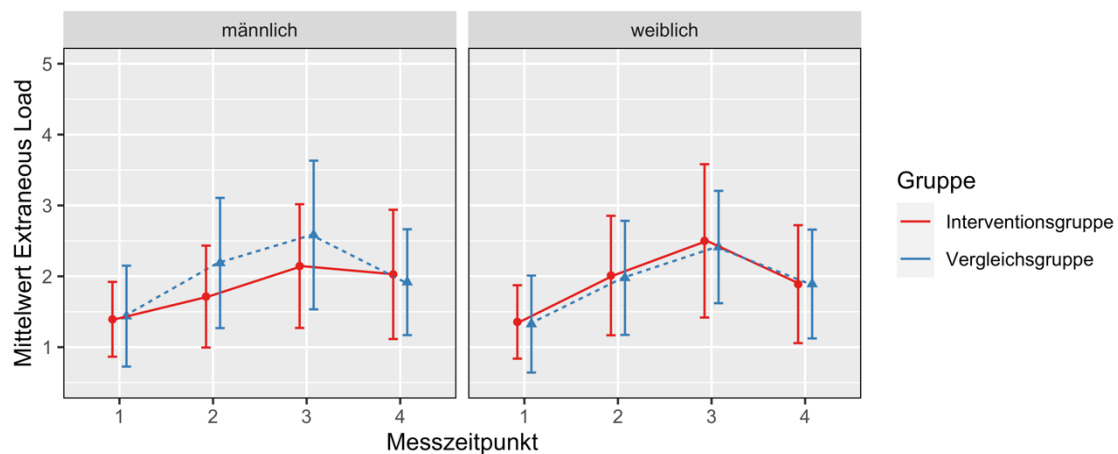


Abbildung 74: Mittelwerte mit Standardabweichung des Extraneous Load beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe.

Die Zunahme des Extraneous Load über die ersten drei Stunden ist deutlich erkennbar. Der Effekt fällt bei den Schülerinnen erneut stärker aus. Die Werte der Vergleichsgruppe liegen bei den Schülern in der zweiten und dritten Stunde über denen der Interventionsgruppe. Die Effektstärke des Haupteffektes „Messzeitpunkt“ ist mit  $\delta_t = 0.30$  (alle Stunden) und  $\delta_t = 0.48$  (erste – vierte Stunde) als schwach beziehungsweise mittelstark anzusehen.

**Germane Load: Geschlecht, Gruppe, Messzeitpunkt***Mittelstufe*

Zuletzt wird der Germane Load in Abhängigkeit von der Zeit und dem Geschlecht zwischen der Interventions- und Vergleichsgruppe untersucht. Eine Prüfung der Voraussetzungen ergibt eine Verletzung der Sphärizität ( $p < 0.001$ ), der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ) und der Varianzhomogenität in der letzten Stunde ( $p = 0.015$ ). Zusätzlich sind 36 Ausreißer vorhanden, wovon vier als extrem einzustufen sind. Daher wurde eine robuste gemischte Varianzanalyse durchgeführt. Die deskriptive Statistik ist in Tabelle 137 zusammengefasst.

*Tabelle 137: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Germane Load in der Mittelstufe.*

	<i>Interventionsgruppe</i>								
	weiblich				männlich				
	Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang		107				97			
Mittelwert	3.92	3.96	3.81	3.69	3.8	3.87	3.81	3.56	
(Standardabweichung)	(0.77)	(0.79)	(0.80)	(0.85)	(0.84)	(0.83)	(0.90)	(0.97)	
Median	4.00	4.00	3.67	3.67	4.00	4.00	4.00	3.67	
	<i>Vergleichsgruppe</i>								
	weiblich				männlich				
	Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang		70				76			
Mittelwert	3.43	3.48	3.41	3.20	3.59	3.69	3.51	3.48	
(Standardabweichung)	(0.95)	(0.97)	(1.02)	(1.05)	(0.89)	(0.83)	(0.75)	(0.71)	
Median	3.33	3.50	3.33	3.17	3.67	3.67	3.33	3.33	

Tabelle 138 zeigt die Ergebnisse der robusten mixed ANOVA. Es sind keine signifikanten Interaktionen vorhanden. Die Haupteffekte „Gruppe“ und „Messzeitpunkt“ sind signifikant.

Tabelle 138: Ergebnisse der robusten Varianzanalyse des Germane Load in der Mittelstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap Verfahren.

Effekt	$Q$	$p$
Geschlecht	0.62	0.431
Gruppe	22.12**	< 0.001
Messzeitpunkt	9.51**	< 0.001
Geschlecht*Gruppe	1.96	0.172
Geschlecht*Messzeitpunkt	0.016	0.995
Gruppe*Messzeitpunkt	0.20	0.813
Geschlecht*Gruppe*Messzeitpunkt	1.62	0.070

Abbildung 75 zeigt eine Abnahme des Germane Load ab der dritten Stunde für beide Geschlechter und beide Gruppen. Der Load liegt bei der Interventionsgruppe sowohl bei den Schülerinnen als auch bei den Schülern über dem der Vergleichsgruppe. Von besonderem Interesse ist der Haupteffekt „Gruppe“, weshalb dieser weiter untersucht wird. Nach einem robusten t-Test unterscheiden sich die Gruppen signifikant voneinander mit einer mittelstarken Effektgröße ( $\text{robust } t(163.46) = 4.43, p < 0.001, \xi = 0.37, M_I = 3.80 \pm 0.69, M_V = 3.48 \pm 0.79$ ).

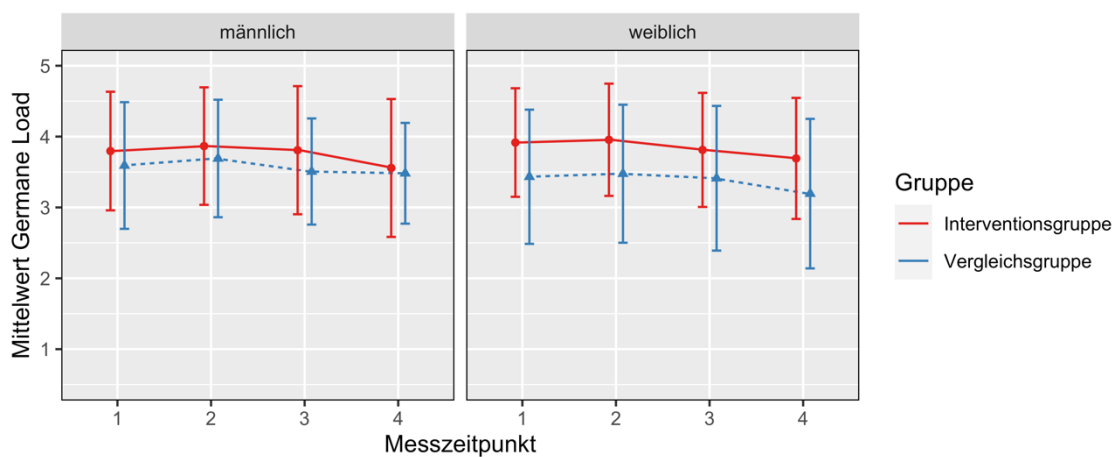


Abbildung 75: Mittelwerte mit Standardabweichung des Germane Load beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.

Die Effektstärken des Haupteffektes „Messzeitpunkt“ sind mit  $\delta_i = 0.17$  (alle Stunden) und  $\delta_i = 0.34$  (erste – vierte Stunde) schwach.

*Oberstufe*

Auch für die Oberstufe wurden zunächst die Voraussetzungen der gemischten Varianzanalyse geprüft. Es zeigt sich eine Verletzung der Normalverteilung (*QQ-Plot*, *Shapiro-Wilk-Test*:  $p < 0.001$ ), der Varianzhomogenität in der ersten Stunde ( $p = 0.014$ ) und der Homogenität der Kovarianzmatrizen des Faktors „Gruppe“. Zudem sind 25 Ausreißer, zwei davon extrem, vorhanden. Die Varianzanalyse wurde daher robust mit Bootstrapping durchgeführt. Die deskriptive Statistik ist Tabelle 139 zu entnehmen.

*Tabelle 139: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Germane Load in der Oberstufe.*

	<i>Interventionsgruppe</i>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	45				56			
Mittelwert (Standardabweichung)	3.89 (0.60)	3.78 (0.65)	3.79 (0.79)	3.89 (0.75)	3.6 (0.69)	3.64 (0.75)	3.6 (0.87)	3.56 (0.86)
Median	4.00	4.00	4.00	4.00	3.67	3.67	3.67	3.67
	<i>Vergleichsgruppe</i>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	23				24			
Mittelwert (Standardabweichung)	3.81 (0.89)	3.72 (0.78)	3.51 (0.97)	3.65 (0.99)	3.54 (0.99)	3.39 (0.80)	3.43 (0.93)	3.6 (0.77)
Median	4.00	3.67	3.33	4.00	3.67	3.50	3.33	3.50

Die robuste gemischte Varianzanalyse zeigt keine signifikanten Effekte. Die Ergebnisse sind in Tabelle 140 abgebildet.

Tabelle 140: Ergebnisse der robusten Varianzanalyse des Germane Load in der Oberstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap Verfahren.

Effekt	$Q$	$p$
Geschlecht	2.75	0.105
Gruppe	1.26	0.262
Messzeitpunkt	0.31	0.551
Geschlecht*Gruppe	0.17	0.661
Geschlecht*Messzeitpunkt	0.42	0.377
Gruppe*Messzeitpunkt	0.38	0.369
Geschlecht*Gruppe*Messzeitpunkt	0.19	0.732

In Abbildung 76 sind die Mittelwerte des Germane Load über die Zeit aufgetragen. Dort sind geringe Unterschiede zwischen den Geschlechtern erkennbar, was durch den relativ niedrigen  $p$ -Wert der Varianzanalyse für den Haupteffekt „Geschlecht“ ersichtlich wird. Schülerinnen erfahren einen höheren Germane Load als Schüler. In einzelnen Stunden liegt zudem die Vergleichsgruppe unter der Interventionsgruppe.

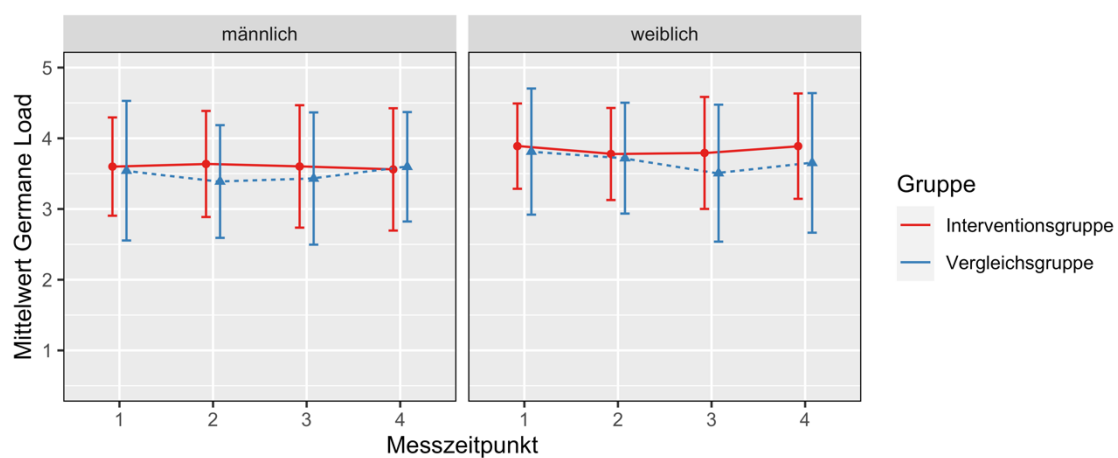


Abbildung 76: Mittelwerte mit Standardabweichung des Germane Load beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe.

**Mental Effort***Mittelstufe*

Der Mental Effort wurde lediglich über ein Item vom Likert-Typ erhoben, wobei die Skalierung als intervallskaliert angenommen werden kann (z. B. Norman, 2010). Die mixed ANOVA wurde als robuste Variante durchgeführt. Als Alternative hätte eine rangbasierte mixed ANOVA durchgeführt werden können (Wilcox, 2021, S. 526). Die Funktion unterstützt jedoch nur einen between-Faktor, womit das Geschlecht nicht hätte in die Auswertung mit aufgenommen werden können. Die deskriptive Statistik ist Tabelle 141 zu entnehmen.

*Tabelle 141: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Mental Effort in der Mittelstufe.*

	<i>Interventionsgruppe</i>							
	weiblich				männlich			
	1	2	3	4	1	2	3	4
Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang	107				97			
Mittelwert	3.07	3.67	3.85	3.90	3.09	3.2	3.44	3.41
(Standardabweichung)	(1.57)	(1.58)	(1.48)	(1.45)	(1.58)	1.57)	(1.51)	(1.66)
Median	3	4	4	4	3	3	4	4
	<i>Vergleichsgruppe</i>							
	weiblich				männlich			
	1	2	3	4	1	2	3	4
Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang	70				76			
Mittelwert	2.56	3.44	3.67	3.71	2.64	3.13	3.31	3.56
(Standardabweichung)	(1.26)	(1.21)	(1.34)	(1.53)	(1.59)	(1.52)	(1.69)	(1.68)
Median	2	3	4	4	2	3	3	4

Die ANOVA zeigt einen signifikanten Haupteffekt der Zeit. Die Teststatistik ist in Tabelle 142 aufgeführt.

Tabelle 142: Ergebnisse der robusten Varianzanalyse des Mental Effort in der Mittelstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap Verfahren.

Effekt	$Q$	$p$
Geschlecht	2.68	0.095
Gruppe	2.26	0.135
Messzeitpunkt	19.18**	< 0.001
Geschlecht*Gruppe	< 0.01	0.950
Geschlecht*Messzeitpunkt	1.20	0.152
Gruppe*Messzeitpunkt	1.10	0.170
Geschlecht*Gruppe*Messzeitpunkt	0.44	0.526

In Abbildung 77 sind die Mittelwerte des Mental Effort nach Gruppen aufgetragen. Dort zeigt sich, dass der Mental Effort der Interventionsgruppe über dem der Vergleichsgruppe liegt. Die Jungen weisen zur ersten Stunde in etwa den gleichen Mittelwert wie die Mädchen auf, wobei zu Stunde vier bei diesen der Wert schneller ansteigt und die Werte der Schülerinnen ab der zweiten Stunde über denen der Schüler liegen. Dies zeigt sich für beide Gruppen.

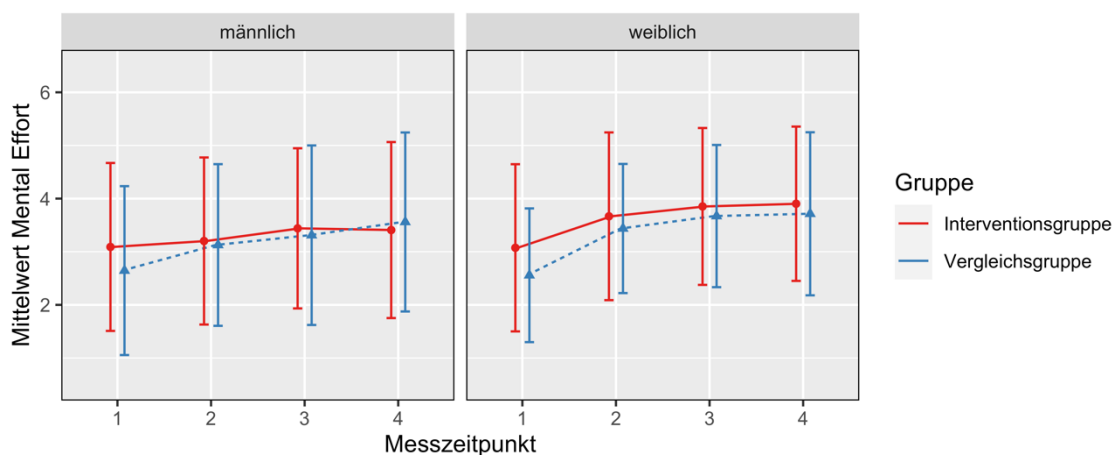


Abbildung 77: Mittelwerte mit Standardabweichung des Mental Effort beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Mittelstufe.

Die Überprüfung des Haupteffektes „Messzeitpunkt“ zeigt signifikante Unterschiede zwischen allen Stunden mit Ausnahme des Vergleichs zwischen der dritten und vierten Stunde. Die paarweisen Vergleiche befinden sich in Anhang 9.3.9. Die Effektstärken sind mit  $\delta_i = 0.29$  (alle Stunden) und  $\delta_i = 0.51$  (erste – vierte Stunde) als schwach bis mittelstark anzusehen.

*Oberstufe*

Der Mental Effort wird wie in der Mittelstufe durch eine robuste gemischte Varianzanalyse ausgewertet. In Tabelle 143 ist die deskriptive Statistik aufgeführt.

*Tabelle 143: Deskriptive Statistik der Interventions- und Vergleichsgruppe in Bezug auf den Mental Effort in der Oberstufe.*

	<i>Interventionsgruppe</i>								
	weiblich				männlich				
	Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang		53				56			
Mittelwert (Standardabweichung)		2.81 (1.32)	3.51 (1.44)	3.89 (1.44)	3.52 (1.56)	2.52 (1.40)	3.01 (1.23)	3.61 (1.47)	3.11 (1.36)
Median		3.00	4.00	4.00	4.00	2.00	3.00	4.00	3.57
	<i>Vergleichsgruppe</i>								
	weiblich				männlich				
	Messzeitpunkt	1	2	3	4	1	2	3	4
Stichprobenumfang		24				25			
Mittelwert (Standardabweichung)		2.33 (1.20)	3.25 (1.26)	3.79 (1.14)	3.29 (1.40)	2.08 (1.22)	3.6 (1.29)	3.91 (1.46)	3.29 (1.38)
Median		2.00	3.00	4.00	3.50	2.00	3.00	4.00	3.00

Die ANOVA zeigt eine signifikante Zweifache-Interaktion „Gruppe\*Messzeitpunkt“, womit sich die Gruppen zu den Messzeitpunkten unterschiedlich entwickeln (vgl. Tabelle 144). Der Haupteffekt „Messzeitpunkt“ ist ebenfalls signifikant, wird jedoch aufgrund der Interaktion nicht weiter analysiert.



Tabelle 144: Ergebnisse der robusten Varianzanalyse des Mental Effort in der Oberstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap Verfahren.

Effekt	$Q$	$p$
Geschlecht	1.12	0.292
Gruppe	0.21	0.649
Messzeitpunkt	24.73**	< 0.001
Geschlecht*Gruppe	1.42	0.244
Geschlecht*Messzeitpunkt	0.61	0.240
Gruppe*Messzeitpunkt	2.34**	0.003
Geschlecht*Gruppe*Messzeitpunkt	0.61	0.284

Zu sehen ist diese Entwicklung in Abbildung 78. So beginnen beide Geschlechter in der Vergleichsgruppe auf einem niedrigeren Niveau, das jedoch im Verlauf der Zeit stärker als in der Interventionsgruppe ansteigt. Damit liegt der Mental Effort der Jungen in der Vergleichsgruppe am Ende der Unterrichtsreihe über dem der Interventionsgruppe. Bei den Mädchen erreicht der Mental Effort in beiden Gruppen in etwa das gleiche Niveau.

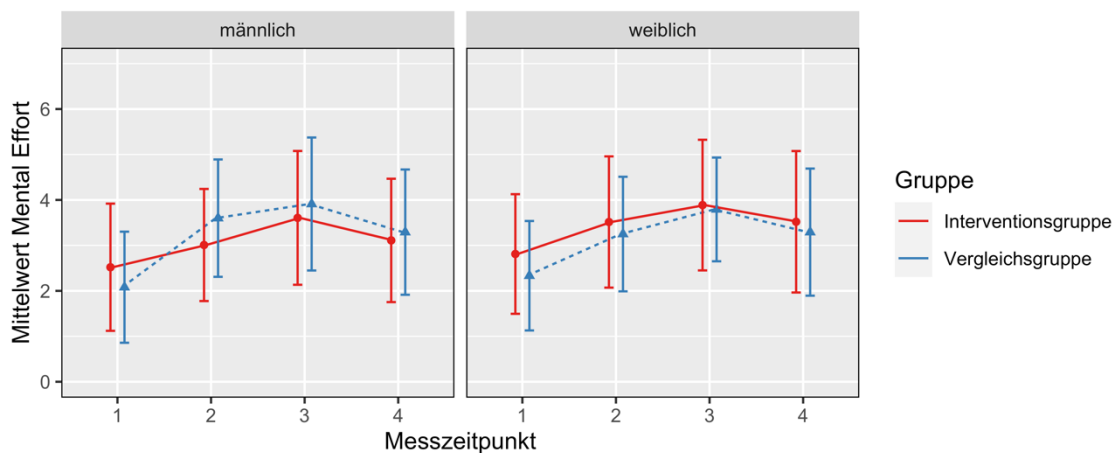


Abbildung 78: Mittelwerte mit Standardabweichung des Mental Effort beim Lernen mit Arbeitsblättern, aufgeteilt nach Geschlecht und Gruppe im zeitlichen Verlauf in der Oberstufe.

Anschließend wurde die Interaktion „Gruppe\*Messzeitpunkt“ analysiert. Der einfache Haupteffekt des Zwischensubjektfaktors zeigt lediglich zum ersten Messzeitpunkt einen signifikanten Unterschied mit mittlerer Effektstärke (vgl. Tabelle 145). Der Innersubjektfaktor zeigt für beide Gruppen ein signifikantes Ergebnis (Interventionsgruppe: *robust*  $F(2.76, 182.43) = 19.81, p < 0.001, \delta_t = 0.26$  (alle Stunden),  $\delta_t = 0.44$  (erste – vierte Stunde)), Vergleichsgruppe: *robust*  $F(2.74, 82.08) = 23.11, p < 0.001, \delta_t = 0.45$  (alle Stunden),  $\delta_t = 0.93$  (erste – vierte Stunde)). In der Interventions- und Vergleichsgruppe

liegt zwischen allen aufeinanderfolgenden Stunden ein signifikanter Unterschied vor. Die genaue Aufschlüsselung ist in Anhang 9.3.10 zu finden.

*Tabelle 145: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Gruppe“ der Interaktion „Gruppe\*Messzeitpunkt“ (Mental Effort) in der Oberstufe.*

<i>Messzeitpunkt</i>	<i>t</i>	<i>df</i>	<i>p</i>	$\hat{\xi}$	<i>M<sub>I</sub>(SD)</i>	<i>M<sub>V</sub>(SD)</i>
1	2.35	80.36	0.021*	0.25	2.66 (1.36)	2.20 (1.21)
2	0.77	62.70	0.442	0.10	3.26 (1.36)	3.43 (1.27)
3	0.42	83.95	0.680	0.09	3.74 (1.46)	3.85 (1.30)
4	0.18	58.48	0.856	0.05	3.31 (1.46)	3.29 (1.37)

### 5.5.2 Cognitive Load und Mental Effort im Schulformvergleich

Wie bereits bei der Motivation wird der Cognitive Load nun in Abhängigkeit von der Schulform, vom Geschlecht und vom Messzeitpunkt analysiert. Da die bisherigen Analysen ebenfalls als robuste Verfahren durchgeführt wurden, findet auch hier eine robuste gemischte Varianzanalyse Anwendung. Die Voraussetzungen werden daher nicht erneut geprüft.

#### Intrinsic Load

Die deskriptive Statistik ist in Tabelle 146 dargestellt.

*Tabelle 146: Deskriptive Statistik des Intrinsic Load im Schulformvergleich in der Mittelstufe.*

	<b>Gymnasium</b>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	107				97			
Mittelwert (Standardabweichung)	1.53 (0.52)	1.97 (0.78)	2.26 (0.81)	2.45 (0.92)	1.58 (0.57)	1.93 (0.76)	2.05 (0.78)	2.15 (0.92)
Median	1.50	2.00	2.00	2.50	1.50	2.00	2.00	2.00
	<b>Gesamtschule</b>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	53				64			
Mittelwert (Standardabweichung)	1.54 (0.53)	2.27 (0.94)	2.05 (0.75)	2.26 (0.91)	1.60 (0.59)	2.30 (0.89)	2.21 (0.71)	2.41 (0.93)
Median	1.50	2.50	2.00	2.00	1.50	2.00	2.50	2.50

Die gemischte Varianzanalyse zeigt den erwarteten signifikanten Haupteffekt „Messzeitpunkt“ und eine signifikante Zweifache-Interaktion „Schulform\*Messzeitpunkt“ (vgl. Tabelle 147). Demnach unterscheiden sich die beiden Schulformen im Laufe der Unterrichtsreihe.

Tabelle 147: Ergebnisse der robusten Varianzanalyse des Intrinsic Load im Schulformvergleich in der Mittelstufe. Die Q-Werte stammen aus dem non-bootstrap Verfahren.

<i>Effekt</i>	<i>Q</i>	<i>p</i>
Geschlecht	< 0.01	0.928
Schulform	2.10	0.150
Messzeitpunkt	46.07**	< 0.001
Geschlecht*Schulform	3.13	0.077
Geschlecht*Messzeitpunkt	0.17	0.831
Schulform*Messzeitpunkt	2.45*	0.012
Geschlecht*Schulform*Messzeitpunkt	1.26	0.104

In Abbildung 79 zeigen beiden Schulformen in der ersten, zweiten und vierten Stunde den gleichen Verlauf, wobei die Schüler:innen der Gesamtschule in der zweiten Stunde eine stärkere Zunahme des Intrinsic Load erfahren. In der dritten Stunde nimmt der Intrinsic Load in der Gesamtschule bei beiden Geschlechtern ab, während er bei den Gymnasiast:innen weiter ansteigt. Dadurch kann die signifikante Zweiwege-Interaktion erklärt werden. Zusätzlich sind Unterschiede zwischen den Geschlechtern in Abhängigkeit von der Schulform festzustellen, wodurch sich die entsprechende Zweiwege-Interaktion „Geschlecht\*Schulform“ dem Signifikanzniveau annähert. Während die Werte der Jungen der Gesamtschule ab der zweiten Stunde konsequent über denen des Gymnasiums liegen, kehrt sich der Trend bei den Mädchen ab der dritten Stunde um.

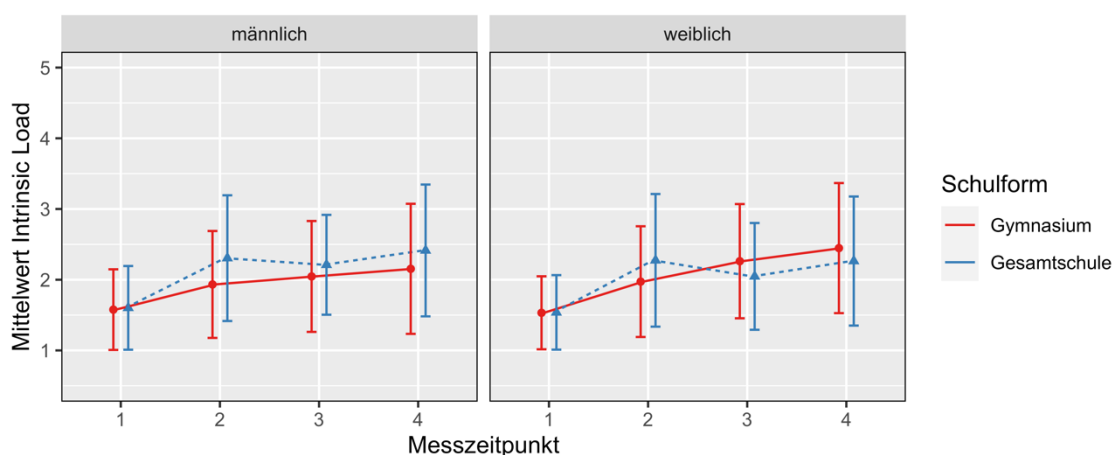


Abbildung 79: Mittelwerte mit Standardabweichung des Intrinsic Load beim Lernen mit Hyper-Docs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.

Die Auswertung der einfachen Haupteffekte der signifikanten Zweifache-Interaktion ist in Tabelle 148 dargestellt. Demnach unterscheiden sich die Schulformen, unabhängig vom Geschlecht, lediglich zum zweiten Messzeitpunkt voneinander.

*Tabelle 148: Ergebnisse der Untersuchung des Zwischensubjektfaktors „Schulform“ der Interaktion „Schulform\*Messzeitpunkt“ in der Mittelstufe.*

<i>Messzeitpunkt</i>	<i>t</i>	<i>df</i>	<i>p</i>	$\hat{\xi}$	<i>M<sub>Gym</sub>(SD)</i>	<i>M<sub>Ges</sub>(SD)</i>
1	0.36	142.52	0.718	0.04	1.55 (0.54)	1.57 (0.56)
2	3.24**	145.18	0.001	0.25	1.95 (0.77)	2.29 (0.91)
3	0.22	179.74	0.826	0.04	2.16 (0.80)	2.14 (0.73)
4	0.62	145.50	0.537	0.06	2.31 (0.93)	2.35 (0.92)

Der Innersubjektfaktor zeigt für beide Schulformen ein signifikantes Ergebnis (Gymnasium: *robust*  $F(2.92, 359.49) = 50.41, p < 0.001, \delta_t = 0.42$  (alle Stunden),  $\delta_t = 0.75$  (erste – vierte Stunde)), Gesamtschule: *robust*  $F(2.78, 194.40) = 36.40, p < 0.001, \delta_t = 0.39$  (alle Stunden),  $\delta_t = 0.71$  (erste – vierte Stunde)). Bei dem Gymnasium unterscheiden sich alle Stunden signifikant voneinander (vgl. Anhang 9.3.11). Bei der Gesamtschule liegt zwischen der zweiten und dritten Stunde kein signifikantes Ergebnis vor.

**Extraneous Load**

Die deskriptive Statistik ist Tabelle 149 zu entnehmen.

*Tabelle 149: Deskriptive Statistik des Extraneous Load im Schulformvergleich in der Mittelstufe.*

		<b>Gymnasium</b>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		107				97			
Mittelwert		1.54	1.71	2.13	2.05	1.64	1.77	1.75	1.97
(Standardabweichung)		(0.63)	(0.64)	(0.88)	(0.86)	(0.69)	(0.84)	(0.80)	(0.86)
Median		1.50	1.50	2.00	2.00	1.50	1.50	1.50	2.00
		<b>Gesamtschule</b>							
		weiblich				männlich			
Messzeitpunkt		1	2	3	4	1	2	3	4
Stichprobenumfang		53				64			
Mittelwert		1.97	2.29	2.28	2.31	1.72	2.19	2.11	2.29
(Standardabweichung)		(0.76)	(0.85)	(0.89)	(0.99)	(0.73)	(0.90)	(0.84)	(0.86)
Median		2.00	2.50	2.00	2.00	1.50	2.00	2.12	2.50

Die Ergebnisse der gemischten Varianzanalyse sind in Tabelle 150 dargestellt. Demnach liegen eine signifikante Dreiwege-Interaktion, eine signifikante Zweiwege-Interaktion „Schulform\*Messzeitpunkt“ und zwei signifikante Haupteffekte vor.

Tabelle 150: Ergebnisse der robusten Varianzanalyse des Extraneous Load im Schulformvergleich in der Mittelstufe. Die  $Q$ -Werte stammen aus dem non-bootstrap Verfahren.

Effekt	$Q$	$p$
Geschlecht	2.67	0.098
Schulform	20.76**	< 0.001
Messzeitpunkt	14.21**	< 0.001
Geschlecht*Schulform	0.05	0.843
Geschlecht*Messzeitpunkt	1.19	0.157
Schulform*Messzeitpunkt	1.85*	0.032
Geschlecht*Schulform*Messzeitpunkt	2.18*	0.018

In Abbildung 80 sind die Mittelwerte mit Standardabweichung dargestellt. Es ist zu erkennen, dass sich die Geschlechter nach Schulform und Messzeitpunkt unterschiedlich entwickeln. Generell liegt der Extraneous Load der Schüler:innen der Gesamtschule über dem der Gymnasiast:innen. Dabei erfahren die Schülerinnen des Gymnasiums jedoch in der dritten Stunde eine deutliche Zunahme des Extraneous Load.

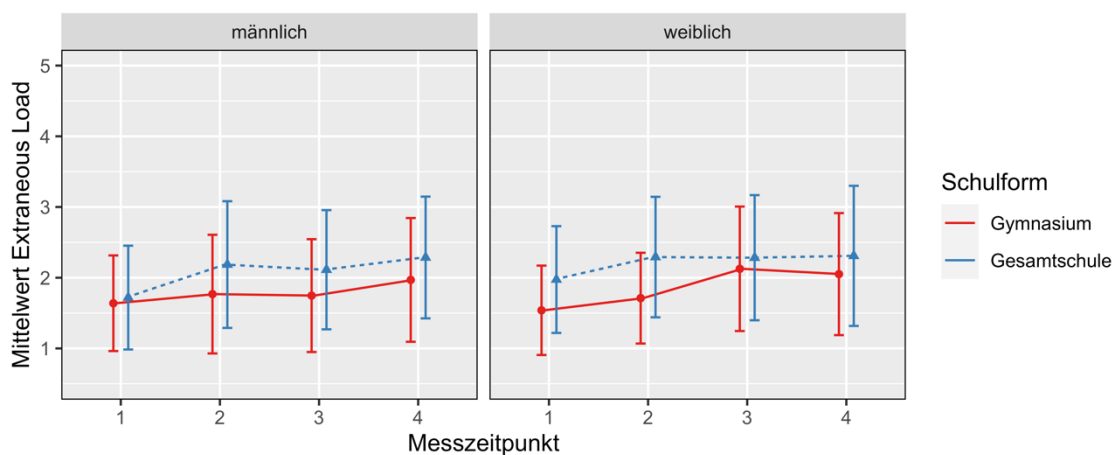


Abbildung 80: Mittelwerte mit Standardabweichung des Extraneous Load beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.

Die Auswertung der signifikanten Dreifach-Interaktion erfolgte durch eine zweifaktorielle robuste Varianzanalyse mit Bootstrapping mit den Faktoren „Geschlecht“ und „Schulform“ zu den einzelnen Messzeitpunkten. Die Ergebnisse sind in Tabelle 151 aufgeführt. Zusätzlich erreicht der Haupteffekt „Schulform“ zu allen Messzeitpunkten das Signifikanzniveau ( $p < 0.01$ ).

Tabelle 151: Teststatistik der jeweiligen Untersuchung der Interaktion „Geschlecht\*Schulform“ zu den einzelnen Messzeitpunkten in der Mittelstufe.

Messzeitpunkt	WJ	df <sub>N</sub>	df <sub>D</sub>	p
1	5.38*	1	95.85	0.025
2	0.57	1	132.70	0.570
3	1.85	1	117.40	0.179
4	0.15	1	121.50	0.697

Nach Tabelle 151 liegt eine signifikante Zweifache-Interaktion zum ersten Messzeitpunkt vor. Daraus kann geschlossen werden, dass sich die Geschlechter in Abhängigkeit von der Schulform voneinander unterscheiden. Die Ursache dafür zeigt sich in Abbildung 80. In der ersten Stunde zeigen die Schüler beider Schulformen den gleichen Load, während die Schülerinnen der Gesamtschule einen höheren Load erfahren als die Gymnasiastinnen. Bei Analyse des einfachen Haupteffekts nach dem Faktor „Schulform“ zeigt sich das erwartete Ergebnis, wonach sich die Schülerinnen in der Gesamtschule von denen des Gymnasiums signifikant voneinander unterscheiden (*robust*  $t(3.28) = 47.89$ ,  $p = 0.002$ ,  $\hat{\xi} = 0.40$ ). Der einfache Haupteffekt des Faktors „Geschlecht“ zeigt signifikante Unterschiede zwischen den Geschlechtern der Gesamtschule (*robust*  $t(2.14) = 54.75$ ,  $p = 0.037$ ,  $\hat{\xi} = 0.28$ ). Die Effektstärken des Innersubjektfaktors „Messzeitpunkt“ wurden für beide Schulformen ermittelt. Gesamtschule:  $\delta_t = 0.22$  (alle Stunden);  $\delta_t = 0.45$  (erste – vierte Stunde) und Gymnasium:  $\delta_t = 0.26$  (alle Stunden);  $\delta_t = 0.45$  (erste – vierte Stunde).



## Germane Load

Die deskriptive Statistik bezüglich des Germane Load ist in Tabelle 152 dargestellt.

*Tabelle 152: Deskriptive Statistik des Germane Load im Schulformvergleich in der Mittelstufe.*

	<i>Gymnasium</i>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	107				97			
Mittelwert (Standardabweichung)	3.92 (0.77)	3.96 (0.79)	3.81 (0.80)	3.69 (0.85)	3.80 (0.84)	3.87 (0.83)	3.81 (0.90)	3.56 (0.97)
Median	4.00	4.00	3.67	3.67	4.00	4.00	4.00	3.67
	<i>Gesamtschule</i>							
	weiblich				männlich			
	Messzeitpunkt	1	2	3	4	1	2	3
Stichprobenumfang	53				64			
Mittelwert (Standardabweichung)	3.67 (0.87)	3.82 (0.81)	3.47 (1.03)	3.32 (1.02)	3.81 (0.79)	3.94 (0.67)	3.76 (0.79)	3.48 (0.94)
Median	3.67	4.00	3.67	3.00	4.00	4.00	3.83	3.67

Nach Durchführung der gemischten Varianzanalyse zeigen sich lediglich zwei signifikante Haupteffekte „Schulform“ und „Messzeitpunkt“ (vgl. Tabelle 153).

*Tabelle 153: Ergebnisse der robusten Varianzanalyse des Germane Load im Schulformvergleich in der Mittelstufe. Die Q-Werte stammen aus dem non-bootstrap Verfahren.*

<i>Effekt</i>	<i>Q</i>	<i>p</i>
Geschlecht	0.53	0.467
Schulform	4.39*	0.035
Messzeitpunkt	11.79**	< 0.001
Geschlecht*Schulform	1.69	0.197
Geschlecht*Messzeitpunkt	0.72	0.292
Schulform*Messzeitpunkt	0.76	0.287
Geschlecht*Schulform*Messzeitpunkt	0.13	0.907

Die Abnahme des Germane Load zeigt sich in Abbildung 81. Dabei ist insbesondere ab der dritten und vierten Stunde eine Reduktion zu verzeichnen. Die Schülerinnen der Gesamtschule zeigen in diesen Stunden den niedrigsten Germane Load und liegen auch unterhalb der Schülerinnen der Interventionsgruppe.

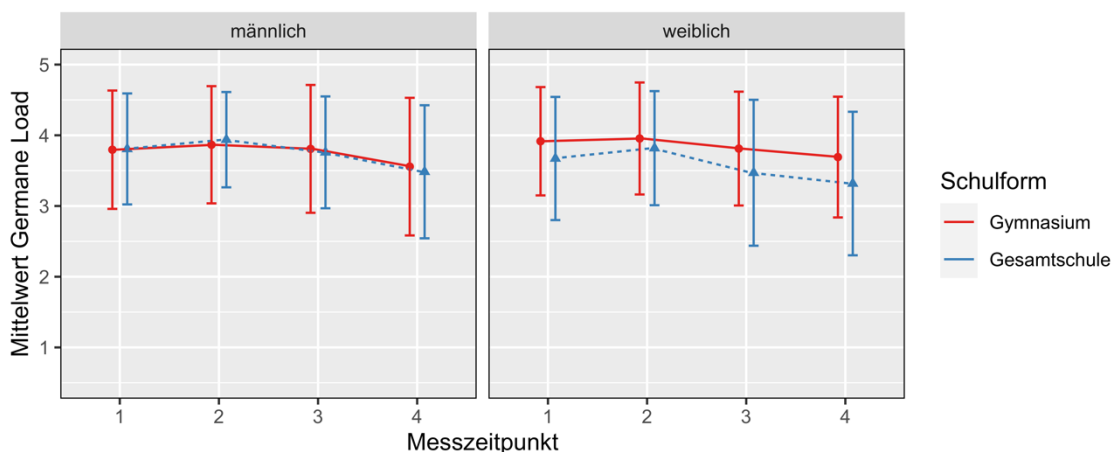


Abbildung 81: Mittelwerte mit Standardabweichung des Germane Load beim Lernen mit HyperDocs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.

Zur Abschätzung der Effektstärke in Bezug auf die Unterschiede zwischen den Schulformen wurde ein einfacher robuster t-Test gerechnet ( $t(140.93) = 2.03, p = 0.045, \hat{\xi} = 0.16$ ). Die Unterschiede zwischen den Gymnasiast:innen ( $M = 3.80 \pm 0.69$ ) und den Gesamtschüler:innen ( $M = 3.67 \pm 0.71$ ) sind daher nur gering. Die Effektstärken des Messzeitpunktes sind  $\delta_i = 0.17$  (alle Stunden) beziehungsweise  $\delta_i = 0.31$  (erste – vierte Stunde).

## Mental Effort

Auch beim Mental Effort zeigen sich bereits deskriptive Unterschiede, die in Tabelle 154 dargestellt sind.

Tabelle 154: Deskriptive Statistik des Mental Effort im Schulformvergleich in der Mittelstufe.

Messzeitpunkt	<b>Gymnasium</b>							
	weiblich				männlich			
	1	2	3	4	1	2	3	4
Stichprobenumfang	107				97			
Mittelwert	3.07	3.67	3.85	3.90	3.09	3.20	3.44	3.41
(Standardabweichung)	(1.57)	(1.58)	(1.48)	(1.45)	(1.58)	(1.57)	(1.51)	(1.66)
Median	3.00	4.00	4.00	4.00	3.00	3.00	4.00	4.00
Messzeitpunkt	<b>Gesamtschule</b>							
	weiblich				männlich			
	1	2	3	4	1	2	3	4
Stichprobenumfang	53				64			
Mittelwert	3.85	4.44	4.15	4.12	2.98	3.29	3.48	3.39
(Standardabweichung)	(1.58)	(1.58)	(1.46)	(1.45)	(1.68)	(1.64)	(1.62)	(1.73)
Median	4.00	4.00	4.00	4.00	3.00	3.00	4.00	3.59

Nach Tabelle 155 zeigt die gemischte Varianzanalyse lediglich zwei signifikante Haupteffekte.

Tabelle 155: Ergebnisse der robusten Varianzanalyse des Mental Effort im Schulformvergleich in der Mittelstufe. Die *Q*-Werte stammen aus dem non-bootstrap Verfahren.

<b>Effekt</b>	<b>Q</b>	<b>p</b>
Geschlecht	12.16**	0.002
Schulform	1.34	0.262
Messzeitpunkt	7.86**	< 0.001
Geschlecht*Schulform	3.83	0.60ß
Geschlecht*Messzeitpunkt	1.40	0.096
Schulform*Messzeitpunkt	0.83	0.274
Geschlecht*Schulform*Messzeitpunkt	1.00	0.219

Die Auswertung des Haupteffektes „Messzeitpunkt“ zeigt lediglich zwischen der ersten und zweiten Stunde einen interessanten signifikanten Unterschied (vgl. Anhang 9.3.11), der sich durch den sprunghaften Anstieg des Mental Effort bei den Schülerinnen in Abbildung 82 zeigt. Der Einfache Haupteffekt „Geschlecht“ zeigt ein signifikantes Ergebnis (*robust t*(189.61) = 3.12,  $p = 0.002$ ,  $\hat{\xi} = 0.25$ ), wonach Schülerinnen ( $M = 3.67 \pm 1.40$ ) einen höheren Mental Effort zeigen als Schüler ( $M = 3.46 \pm 1.32$ ). Die Effektstärken des Haupteffektes „Messzeitpunkt“ sind  $\delta_t = 0.21$  (alle Stunden) beziehungsweise  $\delta_t = 0.34$  (erste – vierte Stunde).

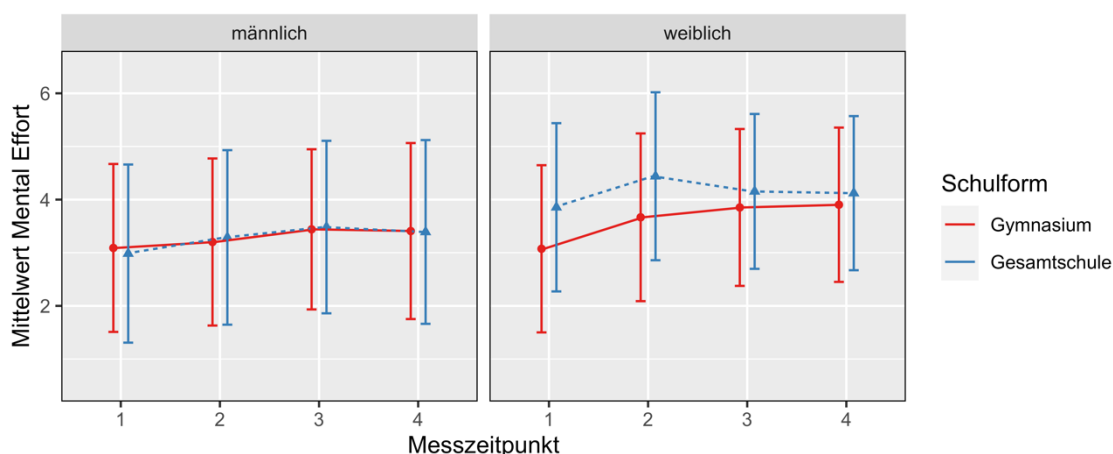


Abbildung 82: Mittelwerte mit Standardabweichung des Mental Effort beim Lernen mit Hyper-Docs, aufgeteilt nach Geschlecht und Schulform im zeitlichen Verlauf in der Mittelstufe.

### 5.5.3 Zusammenhangsanalyse Cognitive Load und intrinsische Motivation

Aufgrund der bisherigen Ergebnisse wurde ein Bezug zwischen dem Cognitive Load und der intrinsischen Motivation vermutet. Es wurden daher Korrelationen zwischen dem Cognitive Load und der intrinsischen Motivation für einen vereinten Datensatz für die Interventions- und Vergleichsgruppe jeweils für die Mittel- und Oberstufe berechnet.

#### Mittelstufe

##### Stunde 1

Zwischen nahezu allen Variablen bestehen signifikante Zusammenhänge auf einem schwachen bis mittelstarken Niveau (vgl. Tabelle 156). Die Zusammenhänge zum Mental Effort sind am schwächsten, gefolgt vom Intrinsic Load. Der Germane Load und der Extraneous Load zeigen einen hohen Zusammenhang zu den Variablen der intrinsischen Motivation. Die Korrelationen entsprechen den erwarteten Richtungen, sodass mit

steigendem Intrinsic und Extraneous Load die Motivation abnimmt. Bei steigendem Germane Load nimmt sie zu. Der Mental Effort zeigt keine klare Tendenz.

*Tabelle 156: Korrelationen zwischen den Faktoren der intrinsischen Motivation und des Cognitive Load in der Mittelstufe (n = 466) in der ersten Stunde.*

	<i>Intrinsic Load</i>	<i>Extraneous Load</i>	<i>Germane Load</i>	<i>Mental Effort</i>
Interesse	$p = 0.055,$ $\rho = -0.09$	$p = 0.002,$ $\rho = -0.14^{**}$	$p < 0.001,$ $\rho = 0.42^{**}$	$p = 0.001,$ $\rho = 0.15^{**}$
Kompetenzerleben	$p < 0.001,$ $\rho = -0.29^{**}$	$p < 0.001,$ $\rho = -0.27^{**}$	$p < 0.001,$ $\rho = 0.27^{**}$	$p = 0.002,$ $\rho = -0.14^{**}$
Wahrgenommene Wahlfreiheit	$p < 0.001,$ $\rho = -0.16^{**}$	$p < 0.001,$ $\rho = -0.24^{**}$	$p < 0.001,$ $\rho = 0.36^{**}$	$p = 0.725,$ $\rho = -0.02$
Druck	$p < 0.001,$ $\rho = -0.35^{**}$	$p < 0.001,$ $\rho = -0.30^{**}$	$p = 0.008,$ $\rho = 0.12^{**}$	$p < 0.001,$ $\rho = -0.27^{**}$

#### *Stunde 4*

In der vierten Stunde nimmt die Stärke der Zusammenhänge tendenziell zu (vgl. Tabelle 157). Der Mental Effort zeigt nach wie vor nur geringe Zusammenhänge zu den Variablen der intrinsischen Motivation. Der Germane Load und der Extraneous Load zeigen zu allen Variablen mittelstarke bis starke Korrelationen.

*Tabelle 157: Korrelationen zwischen den Faktoren der intrinsischen Motivation und des Cognitive Load in der Mittelstufe (n = 466) in der vierten Stunde.*

	<i>Intrinsic Load</i>	<i>Extraneous Load</i>	<i>Germane Load</i>	<i>Mental Effort</i>
Interesse	$p < 0.001,$ $\rho = -0.16^{**}$	$p < 0.001,$ $\rho = -0.35^{**}$	$p < 0.001,$ $\rho = 0.50^{**}$	$p = 0.005,$ $\rho = 0.13^{**}$
Kompetenzerleben	$p < 0.001,$ $\rho = -0.37^{**}$	$p < 0.001,$ $\rho = -0.44^{**}$	$p < 0.001,$ $\rho = 0.47^{**}$	$p = 0.630,$ $\rho = -0.02$
Wahrgenommene Wahlfreiheit	$p < 0.001,$ $\rho = -0.22^{**}$	$p < 0.001,$ $\rho = -0.34^{**}$	$p < 0.001,$ $\rho = 0.32^{**}$	$p = 0.182,$ $\rho = 0.06$
Druck	$p < 0.001,$ $\rho = -0.36^{**}$	$p < 0.001,$ $\rho = -0.42^{**}$	$p < 0.001,$ $\rho = 0.27^{**}$	$p < 0.001,$ $\rho = -0.20^{**}$

**Oberstufe***Stunde 1*

In der Oberstufe bestehen in der ersten Stunde weniger signifikante Zusammenhänge als in der Mittelstufe (vgl. Tabelle 158). Der Germane und Extraneous Load zeigen erneut die höchsten Korrelationen zu den Variablen der intrinsischen Motivation. Die Richtung der Zusammenhänge bleibt bestehen, sodass ein erhöhter Intrinsic oder Extraneous Load die Motivation senkt.

*Tabelle 158: Korrelationen zwischen den Faktoren der intrinsischen Motivation und des Cognitive Load in der Oberstufe (n = 146) in der ersten Stunde.*

	<i>Intrinsic Load</i>	<i>Extraneous Load</i>	<i>Germane Load</i>	<i>Mental Effort</i>
Interesse	$p = 0.124,$ $\rho = 0.128$	$p = 0.130,$ $\rho = -0.13$	$p < 0.001,$ $\rho = 0.41^{**}$	$p = 0.004,$ $\rho = 0.24^{**}$
Kompetenzerleben	$p = 0.007,$ $\rho = -0.22^{**}$	$p < 0.001,$ $\rho = -0.30^{**}$	$p < 0.001,$ $\rho = 0.32^{**}$	$p = 0.060,$ $\rho = -0.16$
Wahrgenommene Wahlfreiheit	$p = 0.840,$ $\rho = 0.02$	$p = 0.002,$ $\rho = -0.25^{**}$	$p = 0.007,$ $\rho = 0.22^{**}$	$p = 0.159,$ $\rho = 0.12$
Druck	$p = 0.039,$ $\rho = -0.17^*$	$p = 0.033,$ $\rho = -0.18^*$	$p = 0.052,$ $\rho = 0.16$	$p < 0.001,$ $\rho = -0.29^{**}$

*Stunde 4*

In der vierten Stunde nehmen die Zusammenhänge zu (vgl. Tabelle 159). Nach wie vor liegen mittelstarke bis starke Zusammenhänge zwischen dem Extraneous und Germane Load und den Variablen der intrinsischen Motivation vor. Das Kompetenzerleben und der Druck zeigen zu allen Variablen des Cognitive Load signifikante Zusammenhänge auf einem mittelstarken Niveau.

Tabelle 159: Korrelationen zwischen den Faktoren der intrinsischen Motivation und des Cognitive Load in der Oberstufe ( $n = 146$ ) in der vierten Stunde.

	<i>Intrinsic Load</i>	<i>Extraneous Load</i>	<i>Germane Load</i>	<i>Mental Effort</i>
Interesse	$p = 0.059,$ $\rho = -0.16$	$p < 0.001,$ $\rho = -0.31^{**}$	$p < 0.001,$ $\rho = 0.60^{**}$	$p = 0.208,$ $\rho = 0.11,$
Kompetenzerleben	$p < 0.001,$ $\rho = -0.32^{**}$	$p < 0.001,$ $\rho = -0.42^{**}$	$p < 0.001,$ $\rho = 0.43^{**}$	$p = 0.005,$ $\rho = -0.23^{**}$
Wahrgenommene Wahlfreiheit	$p = 0.127,$ $\rho = -0.13$	$p = 0.003,$ $\rho = -0.24^{**}$	$p < 0.001,$ $\rho = 0.33^{**}$	$p = 0.090,$ $\rho = 0.09$
Druck	$p = 0.037,$ $\rho = -0.17^*$	$p = 0.009,$ $\rho = -0.21^{**}$	$p < 0.001,$ $\rho = 0.35^{**}$	$p = 0.002,$ $\rho = -0.25^{**}$

#### 5.5.4 Lernen mit Tablets

Neben dem Cognitive Load wurden auch weitere Variablen erhoben, die physische Auswirkungen der Tabletnutzung abbilden sollen. Das betrifft allgemein die Ermüdung durch die Nutzung der Tablets (TAB1) und spezieller die Ermüdung der Augen (TAB2). Zusätzlich wurde erfragt, ob die Schüler:innen durch andere Anwendungen auf dem Tablet vom Lernen abgehalten wurden (TAB3). Fehlende Werte wurden für die Variable TAB3 nicht ersetzt, da nur ein geringer Teil der Werte (1 %) fehlten. Im zweiten Teil werden die Ergebnisse bezüglich tabletbezogenen Kompetenzen geschildert. Die Ergebnisse werden für die Mittel- und Oberstufe der Interventionsgruppe dargestellt.

##### Einschränkungen durch Tablets

In der Mittelstufe gibt nur ein kleiner Teil der Lernenden an, durch die Tablets irgendeine Form von Ermüdung zu erfahren. Die Beurteilung bleibt über alle vier Stunden konstant niedrig (vgl. Abbildung 83).

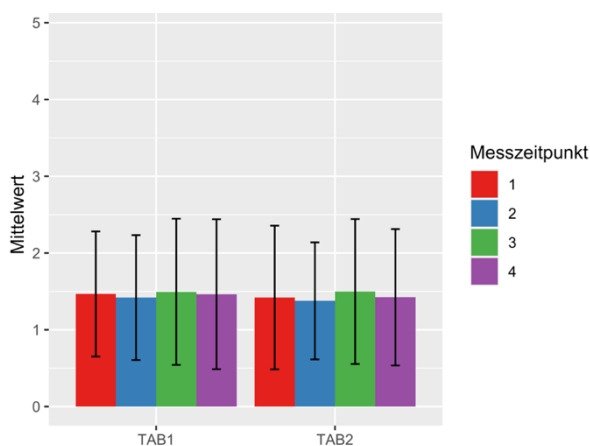


Abbildung 83: Mittelwerte mit Standardabweichung der Variablen TAB1 und TAB2 über die vier Messzeitpunkte (Stunden) in der Mittelstufe.

Das gleiche Bild zeigt sich in der Oberstufe, wobei hier mehr Schüler:innen angeben, durch die Nutzung der Tablets zu ermüden (vgl. Abbildung 84). Die Werte sind jedoch immer noch niedrig und konstant über die Messzeitpunkte. Die hohe Standardabweichung lässt auf eine hohe Streuung der Antworten schließen.

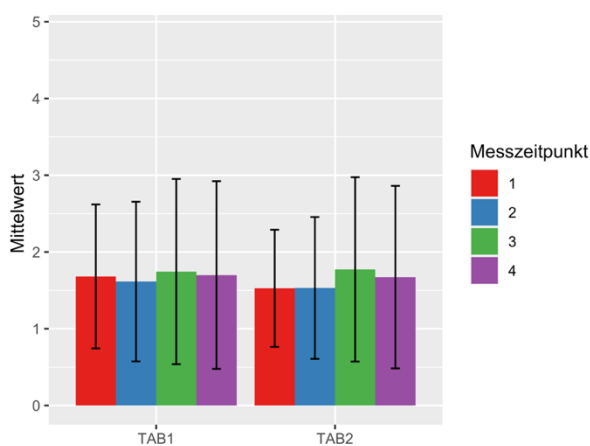


Abbildung 84: Mittelwert mit Standardabweichung der Variablen TAB1 und TAB2 über die vier Messzeitpunkte (Stunden) in der Oberstufe.

In Bezug darauf, ob die Lernenden von Apps vom Lernen abgelenkt wurden, existiert in der Oberstufe erneut eine höhere Zustimmung als in der Mittelstufe. In beiden Gruppen nimmt die Zustimmung über die Unterrichtsreihe zu. Sie verbleibt jedoch auf einem niedrigen Niveau (vgl. Abbildung 85).



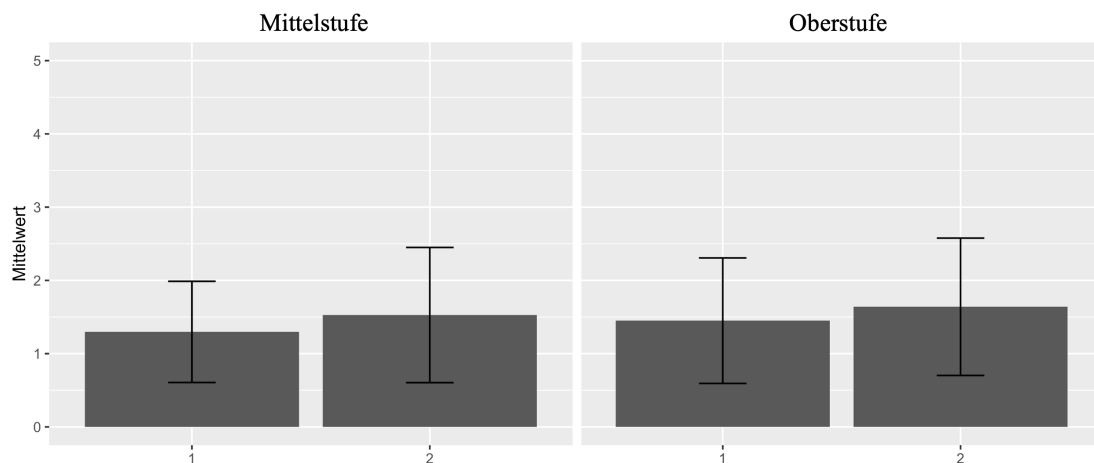


Abbildung 85: Mittelwerte mit Standardabweichung der Variablen TAB3 nach der ersten Stunde (MZP 1) und nach der vierten Stunde (MZP 2) in der Mittel- und Oberstufe.

### **Tabletbezogene Kompetenzen**

Um die Ergebnisse der Usability, des Cognitive Load und der Motivation besser einordnen zu können, werden auch die Daten der Variablen „Tabletbezogene Selbstwirksamkeit“ und „Tablets beim Lernen“ deskriptiv beschrieben.

#### *Mittelstufe*

In Tabelle 160 ist die deskriptive Statistik der Selbstwirksamkeit aufgeführt.

Tabelle 160: Deskriptive Statistik der tabletbezogenen Selbstwirksamkeit im Schulformvergleich in der Mittelstufe.

	<b>Gymnasium</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt	1	2	1	2
Stichprobenumfang	96		89	
Mittelwert (SD)	4.21 (0.67)	4.42 (0.54)	4.43 (0.50)	4.46 (0.61)
Median	4.33	4.52	4.5	4.5
	<b>Gesamtschule</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt	1	2	1	2
Stichprobenumfang	47		55	
Mittelwert (SD)	4.08 (0.70)	4.17 (0.73)	4.3 (0.59)	4.22 (0.75)
Median	4.33	4.33	4.33	4.33

In Abbildung 86 sind die Ergebnisse grafisch verdeutlicht. Demnach liegt die tabletbezogene Selbstwirksamkeit bei allen Teilgruppen auf einem hohen Niveau. Mit Ausnahme der Schüler der Gesamtschule, kommt es zudem bei allen Teilgruppen zu einer Zunahme, die bei den Schülerinnen des Gymnasiums am höchsten ausfällt. Die Schülerinnen erfahren in beiden Schulformen eine geringere Selbstwirksamkeit als die Schüler. Die Differenzen sind jedoch klein.

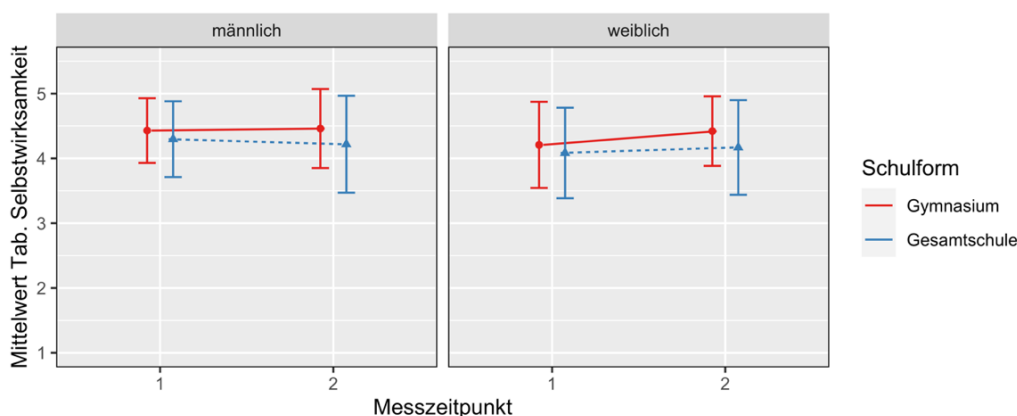


Abbildung 86: Mittelwerte mit Standardabweichung der tabletbezogenen Selbstwirksamkeit, aufgeteilt nach Schulform und Geschlecht im Pre- und Posttest in der Mittelstufe.

Die Nützlichkeit von Tablets beim Lernen wird ebenfalls von den Lernenden aller Schulformen hoch eingestuft (s. Tabelle 161).

Tabelle 161: Deskriptive Statistik von Tablets beim Lernen im Schulformvergleich in der Mittelstufe.

	<b>Gymnasium</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt				
Stichprobenumfang	96		89	
Mittelwert (SD)	4.24 (0.77)	4.36 (0.75)	4.13 (0.88)	4.39 (0.79)
Median	4.33	4.67	4.0	4.67
	<b>Gesamtschule</b>			
	weiblich		männlich	
	1	2	1	2
Messzeitpunkt				
Stichprobenumfang	47		55	
Mittelwert (SD)	4.22 (0.86)	4.2 (0.97)	4.16 (0.75)	4.21 (0.73)
Median	4.67	4.67	4.33	4.33

Bei genauer Betrachtung der Ergebnisse zeigt sich eine leichte Abnahme der Bewertung der Nützlichkeit bei den Schülerinnen der Gesamtschule. Die Bewertungen der Schüler verbleiben auf dem gleichen Niveau. Bei den Schüler:innen des Gymnasiums ist eine Zunahme zu beobachten (vgl. Abbildung 87).

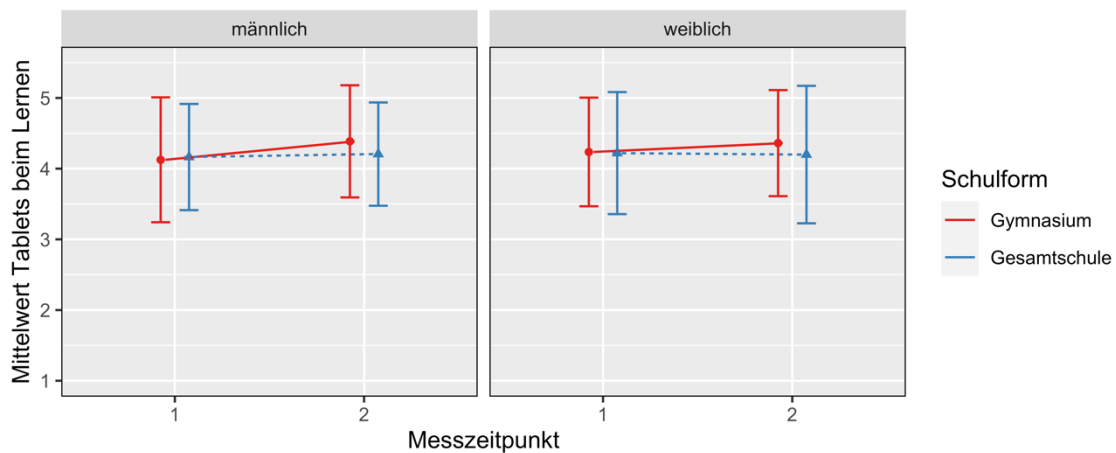


Abbildung 87: Mittelwerte mit Standardabweichung der Variable „Tablets beim Lernen“, aufgeteilt nach Schulform und Geschlecht im Pre- und Posttest in der Mittelstufe.

### Oberstufe

In der Oberstufe liegt ebenfalls eine hohe tabletbezogene Selbstwirksamkeit vor. Die deskriptive Statistik ist Tabelle 162 zu entnehmen.

Tabelle 162: Deskriptive Statistik der tabletbezogenen Selbstwirksamkeit in der Oberstufe.

Messzeitpunkt	<b>Gymnasium</b>			
	weiblich		männlich	
	1	2	1	2
Stichprobenumfang	40		42	
Mittelwert (SD)	4.16 (0.44)	4.25 (0.57)	4.44 (0.51)	4.59 (0.36)
Median	4.17	4.25	4.5	4.67

Sowohl bei den Schülerinnen als auch bei den Schülern ist eine Zunahme der Selbstwirksamkeit durch den Einsatz der HyperDocs erkennbar. Die Werte der Schülerinnen liegen, wie auch in der Mittelstufe, unterhalb derer der Schüler (vgl. Abbildung 88).

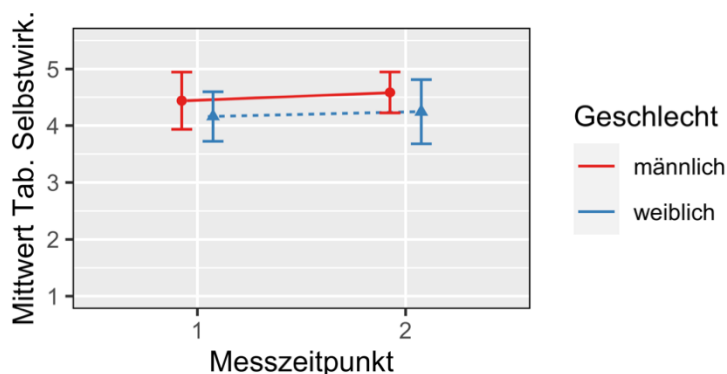


Abbildung 88: Mittelwerte mit Standardabweichung der tabletbezogenen Selbstwirksamkeit, aufgeteilt nach Geschlecht im Pre- und Posttest in der Oberstufe.

Die Nützlichkeit von Tablets beim Lernen erachten die Schüler:innen der Oberstufe als geringer als in der Mittelstufe. Sie ist jedoch ebenfalls auf einem hohen Niveau. Die Nützlichkeit wird von Schülerinnen höher eingestuft (vgl. Tabelle 163 und Abbildung 89).

Tabelle 163: Deskriptive Statistik der Variable „Tablets beim Lernen“ in der Oberstufe.

Messzeitpunkt	Gymnasium			
	weiblich		männlich	
	1	2	1	2
Stichprobenumfang	40		42	
Mittelwert (SD)	3.94 (0.84)	4.00 (0.74)	3.78 (0.84)	3.90 (0.93)
Median	4.00	4.00	3.67	4.00

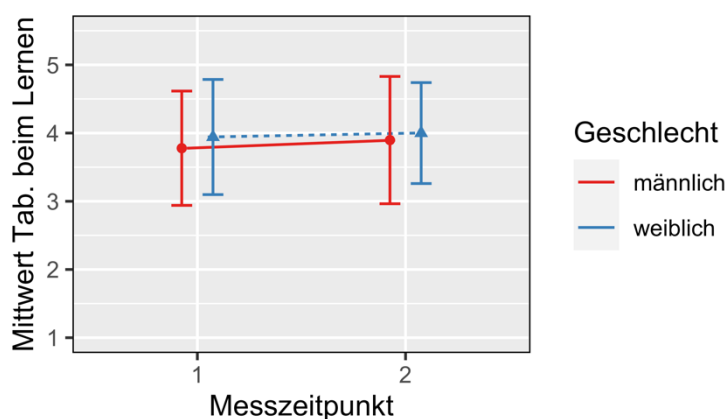


Abbildung 89: Mittelwerte mit Standardabweichung der Variable „Tablets beim Lernen“, aufgeteilt nach Geschlecht im Pre- und Posttest in der Oberstufe.

## 6. Diskussion

In diesem Kapitel werden die Ergebnisse der Studie in einen größeren Zusammenhang und in Beziehung zueinander gesetzt. Dabei erfolgt ein Abgleich mit der bestehenden Literatur. Zudem werden eine ausführliche Methodenkritik dargelegt und Implikationen für das Lernen mit digitalen Arbeitsblättern mit Lernhilfen formuliert. Die letzten Teilkapitel liefern Empfehlungen für die Gestaltung und das Lernen mit digitalen Arbeitsblättern, insbesondere mit HyperDocSystems, und für den weiteren Forschungsbedarf.

### 6.1 Diskussion und Einordnung der Ergebnisse

Die Analyse der Daten, verbunden mit einem Vergleich der bestehenden Literatur, zeigt teilweise überraschende Ergebnisse. Dennoch konnten viele der bisher bestehenden Ergebnisse der Literatur auch für den Kontext des Regelunterrichts im Fach Chemie bestätigt werden. In den folgenden Unterkapiteln werden die Ergebnisse themenbezogen diskutiert und mit der Literatur abgeglichen. Bei der Diskussion wird gegebenenfalls, je nach Forschungsfrage, ausführlich zwischen Mittel- und Oberstufe unterschieden und es werden Detailfragen beantwortet.

Um einen besseren Überblick über die Ergebnisse und Trends in den Daten zu erhalten, wurden zwei Übersichtsgrafiken für die Mittel- und Oberstufe erstellt, die die wichtigsten Ergebnisse im Laufe der Unterrichtsreihe festhalten (vgl. Abbildung 90 & 91).

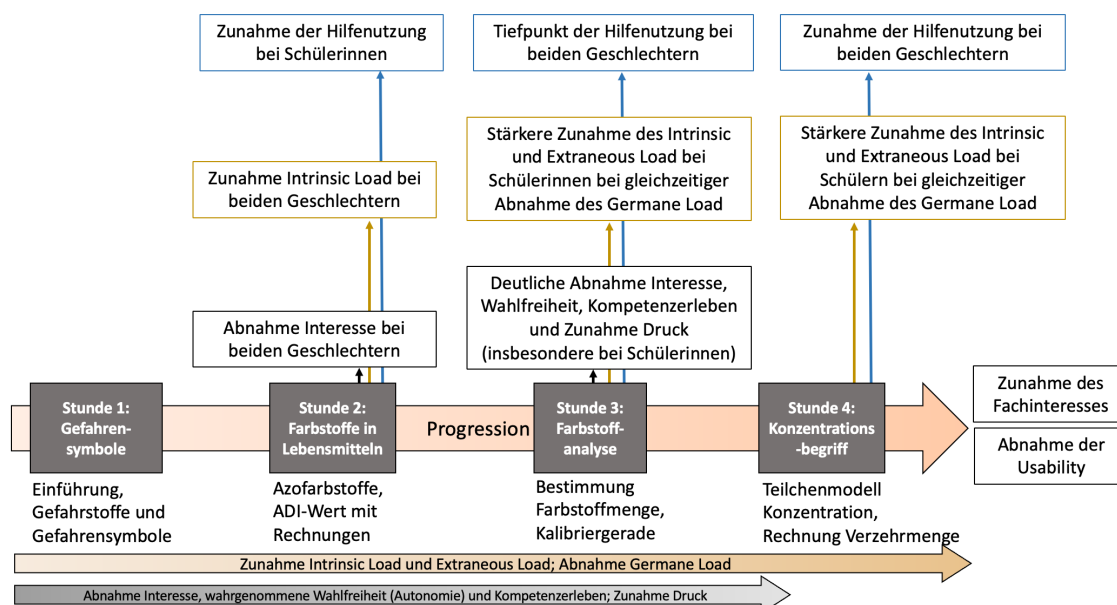


Abbildung 90: Übersicht über die Ergebnisse und Trends der Studie in der Mittelstufe.

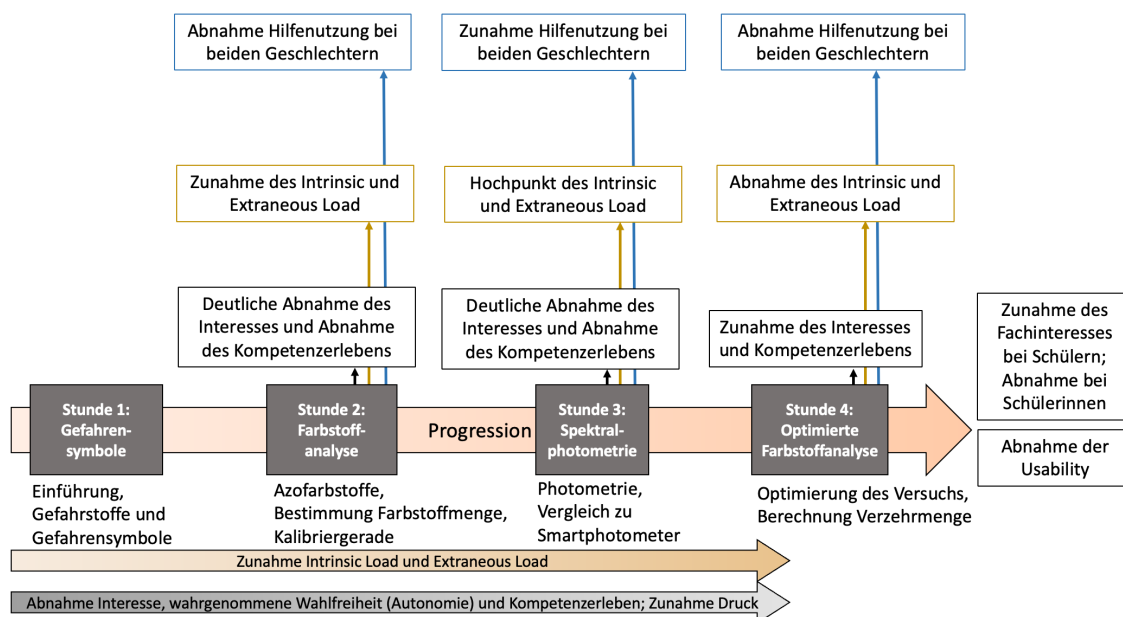


Abbildung 91: Übersicht über die Ergebnisse und Trends der Studie in der Oberstufe.

### 6.1.1 Intrinsische Motivation und Interesse

Die erste Forschungsfrage zum Interesse und der intrinsischen Motivation lautet:

*F\_M1: Inwieweit verändern sich das Fachinteresse Chemie und die intrinsische Motivation beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

#### *Intrinsische Motivation und Interesse F\_M1: Mittelstufe*

In Bezug auf das Fachinteresse Chemie im Pre- und Post-Test zeigen in der Mittelstufe Schülerinnen, unabhängig von der Gruppenzugehörigkeit (Interventions-/Vergleichsgruppe), ein geringeres Interesse als Jungen. Dieses Ergebnis stimmt mit den bisherigen Studien zu diesem Thema überein, wonach Schüler in den Fächern Chemie, Physik und Technik ein höheres Interesse zeigen als Schülerinnen (Osborne et al., 2003, Trumper, 2006b). In der Biologie hingegen findet sich dieser Geschlechterunterschied nicht (Jansen et al., 2013, Trumper, 2006a). Die Unterschiede sind teilweise auf das Selbstkonzept der Individuen zurückzuführen (Jansen et al., 2013, Osborne et al., 2003, Sax et al., 2015), wobei dieses in der vorliegenden Arbeit nicht erhoben wurde. Die Werte der Mädchen der Vergleichsgruppe liegen unterhalb derer der Interventionsgruppe. Bei den Schülern entwickeln sich die Gruppen zum Post-Test auseinander, wodurch eine knappe signifikante ( $p = 0.05$ ) Interaktion „Gruppe\*Messzeitpunkt“ beobachtet werden kann. Die Intervention durch HyperDocs kann damit das Interesse am Fach Chemie, insbesondere für

Jungen, fördern. Ein Grund für die Förderung des Fachinteresses könnten die digitalen Hilfen darstellen. Schüler:innen der Interventionsgruppe bewerten die Hilfen als nützlicher und rufen sie häufiger aus Neugier und Interesse auf als die Vergleichsgruppe. Außerdem schätzen die Lernenden der Vergleichsgruppe, insbesondere die Schülerinnen, ihre Fähigkeiten in Chemie geringer als die Interventionsgruppe ein, womit ihr Kompetenzerleben innerhalb der Reihe auch geringer sein dürfte. Zusätzlich zeigt die Vergleichsgruppe in zwei der vier Stunden einen signifikant höheren Intrinsic und Extraneous Load sowie einen geringeren Germane Load. Der Extraneous und Germane Load zeigen starke negative Korrelationen zu den Variablen der intrinsischen Motivation. Der Mental Effort ist in der Interventionsgruppe allerdings leicht erhöht. Dabei muss jedoch beachtet werden, dass ein erhöhter Mental Effort auch durch eine höhere Auseinandersetzung mit dem Lernmaterial, in Folge eines erhöhten Germane Load, verursacht werden kann (Wouters et al., 2009). Insgesamt zeigen Schülerinnen einen höheren Intrinsic und Extraneous Load, wodurch die geringere Förderung des Interesses im Pre-/Post-Test erklärt werden könnte. Die Auswirkung auf das Interesse kann beispielsweise über die Autonomie mediiert werden (Patall et al., 2018), wobei der Effekt bei Schülerinnen größer ist. Dieser Zusammenhang wird später näher erläutert. Die Ergebnisse für das Sachinteresse an Naturwissenschaften zeigen in der Entwicklung über die Unterrichtsreihe keine signifikanten Unterschiede zwischen den Gruppen.

Für die intrinsische Motivation, gemessen am Interesse (SDT) beim Lernen mit Arbeitsblättern, zeigen sich ebenfalls signifikante Unterschiede mit einer hohen Effektstärke ( $\hat{\xi} = 0.62 - 0.81$ ) zwischen den Gruppen. Die Vergleichsgruppe zeigt ein deutlich niedrigeres Interesse beim Lernen mit Arbeitsblättern als die Interventionsgruppe. Der Interessenszuwachs in der konkreten Lernsituation (Krapp, 1998), der sich beim Lernen mit den digitalen Arbeitsblättern gegenüber der analogen Variante zeigt, nimmt gegebenenfalls einen Einfluss auf das situative Interesse. Das situative Interesse ist dabei besonders wichtig bei Schüler:innen mit geringem individuellem Interesse. Damit würden HyperDocs als catch-Komponente dienen, wie es für Computeranwendungen bereits dokumentiert ist (Mitchell, 1992; Mitchell, 1997). Das situative Interesse könnte anschließend in das individuelle Interesse übergehen. Anzeichen dafür zeigen sich in der Zunahme des Fachinteresses Chemie bei den Jungen. Die Förderung des Interesses beim Bearbeiten der Lerninhalte durch die digitale Intervention kann auf den Einsatz von Technologie zurückzuführen sein (Renninger et al., 2018). Die signifikante Interaktion „Gruppe\*Messzeitpunkt“ deutet darauf hin, dass die Gruppen sich über die Unterrichtsreihe unterschiedlich

entwickeln. Während bei den Tabletbenutzer:innen das Interesse (SDT) beim Bearbeiten der Arbeitsblätter insbesondere in den ersten beiden Stunden abnimmt und danach stagniert, fällt das Interesse bei der Vergleichsgruppe ab der zweiten Stunde kontinuierlich ab. In beiden Gruppen könnte der prinzipielle Abfall durch den Neuheitseffekt erklärt werden. In anderen Studien konnte gezeigt werden, dass mit zunehmender Einsatzdauer von digitalen Medien die Effektstärken der Interventionen in Bezug auf die Motivation abnehmen (Higgins et al., 2019). Neuheit kann außerdem zur Interessensentwicklung beitragen (Renninger et al., 2018). Es gibt jedoch auch Autor:innen, die in ihren Studien zu digitalen Medien keinen Einfluss von Neuheit auf die Motivation feststellten (Jeno et al., 2017). Die gleichen Forscher:innen konnten sogar einen gegenteiligen Effekt feststellen (Jeno et al., 2019a). Sie führen die motivationsfördernden Effekte auf die Unterstützung der Bedürfnisbefriedigung nach Autonomie, Kompetenzerleben und Eingebundenheit von mobilen Anwendungen zurück (Jeno et al., 2019a; Jeno et al., 2019b). Auch Renninger et al. (2018) konnten einen interessensfördernden Aspekt von Autonomie feststellen. Der Neuheitseffekt sollte außerdem differenzierter betrachtet werden. Neben dem eigentlichen Medium können auch weitere Dinge, wie die Unterrichtsform oder das Thema, zur Neuheit beitragen (Renninger & Bachrach, 2015). Andere Autor:innen sehen den Neuheitseffekt als ein weiteres Grundbedürfnis in Bezug auf die Selbstbestimmungstheorie an (González-Cutre et al., 2016; González-Cutre & Scilia, 2019; González-Cutre et al., 2020) und zeigen Zusammenhänge zwischen den einzelnen Konstrukten auf. Letztlich kann für die vorliegenden Daten keine klare Aussage getroffen werden, da im Laufe der Erhebung auch die wahrgenommene Wahlfreiheit (Autonomie) und das Kompetenzerleben abnehmen, welche in starkem Zusammenhang mit dem Interesse (SDT) beim Lernen mit den HyperDocs stehen (Patall et al., 2008), sodass das sinkende Interesse auch durch die Abnahme der Autonomie und des Kompetenzerlebens verursacht wird. Zudem steigen der Intrinsic und Extrinsic Load an und der Germane Load nimmt ab, wodurch auch das Interesse abnehmen könnte. Demnach wäre dessen Abnahme durch den steigenden Anforderungsbereich zu erklären. Die Ergebnisse in Bezug auf das Interesse beim Lernen mit HyperDocs liegen allerdings in Übereinstimmung mit anderen Studien, in denen digitale Medien eingesetzt wurden (Bruckermann et al., 2016). Hierbei ist jedoch zu beachten, dass das Interesse und die Motivation auch von dem Lerngegenstand und dem Einsatzzweck des digitalen Werkzeuges abhängig sind (Herbst et al., 2016; Hillmayr et al., 2020).



Auch bei der wahrgenommenen Wahlfreiheit (Autonomie) beim Lernen mit Arbeitsblättern liegt die Vergleichsgruppe in allen Stunden signifikant mit einer hohen Effektstärke ( $\xi = 0.65$ ) unter der Interventionsgruppe. Bei beiden Gruppen ist eine Abnahme über die Zeit zu verzeichnen, die am stärksten während der ersten beiden Stunden ausfällt. Die Abnahme der wahrgenommenen Wahlfreiheit als ein Indikator für das Autonomieempfinden kann in beiden Gruppen über den Zusammenhang zu dem Konstrukt der intrinsischen Motivation nach der Selbstbestimmungstheorie begründet werden. Eine weitere Möglichkeit wurde in den Stundeninhalten selbst und in den Arbeitsblättern gesehen, so dass diese womöglich die Wahlfreiheit und damit die Autonomie zunehmend einschränken. Bei erneuter Betrachtung der Materialien konnten allerdings keine charakteristischen Unterschiede zwischen den Arbeitsblättern oder Arbeitsaufträgen festgestellt werden, die die Abnahme der wahrgenommenen Wahlfreiheit erklären würden. Ein möglicher Grund könnte in der Vielzahl der dargebotenen Hilfen liegen, sodass die Selbstregulation mit der Einsatzdauer abnimmt (Patall et al., 2008). Außerdem erachten die Schüler:innen die Hilfen zunehmend als weniger nützlich, wodurch die Autonomie ebenfalls abnehmen könnte, da die Wahlmöglichkeiten als weniger persönlich relevant wahrgenommen werden (Patall et al., 2008). Dieser Erklärungsansatz begründet allerdings nicht die unterschiedlich starke Abnahme zwischen beiden Gruppen. Das erhöhte Erleben von Wahlfreiheit beim Lernen mit HyperDocs kann auf eine höhere Auswahl der Hilfen sowie die Art der Texteingabe (Tastatur / Stift) zurückgeführt werden. Damit besteht pro Aufgabe ein höheres Angebot an Wahlmöglichkeiten, was sich ebenfalls positiv auf die wahrgenommene Wahlfreiheit (Autonomie) auswirken kann (Patall et al., 2008). Bereits in anderen Studien konnte ein autonomiefördernder Effekt von Wahlmöglichkeiten, wie beispielsweise die Hintergrundmusik, gefunden werden (Schneider et al., 2018). Der autonomieunterstützende Effekt von mobilen Anwendungen konnte in anderen Studien gezeigt werden (Bruckermann et al., 2016; Jenö et al., 2019b). Schneider und Kolleg:innen (2018) konnten außerdem den Zusammenhang zwischen der wahrgenommenen Wahlfreiheit, dem Autonomieerleben, der intrinsischen Motivation und dem Intrinsic Cognitive Load aufzeigen, sodass Schüler:innen mit mehr Wahlmöglichkeiten ein höheres Autonomieerleben, eine geringere intrinsische Last und eine höhere intrinsische Motivation erfahren. In der zweiten und vierten Stunde geben die Lernenden der Vergleichsgruppe eine höhere intrinsische Last an. Aufgrund der omnipräsenten Nutzung von digitalen Endgeräten im Alltag fühlen sich die Lernenden bei der Nutzung von analogen Materialien in ihrer Wahlmöglichkeit und damit in ihrer Autonomie gegebenenfalls eingeschränkt.

Damit würden digitale Medien nicht zwangsläufig die Autonomie erhöhen (Interventionsgruppe), sondern sie reduzieren diese lediglich, im Gegensatz zu analogen Arbeitsformen (Vergleichsgruppe), nicht. Für diese Vermutung können zum aktuellen Zeitpunkt jedoch keine unterstützenden Belege gefunden werden.

Beim Kompetenzerleben beim Lernen mit Arbeitsblättern liegen ebenfalls signifikante Unterschiede zwischen den Gruppen vor. In allen Stunden liegt das Kompetenzerleben der Vergleichsgruppe unterhalb dem der Interventionsgruppe. Die Interaktion „Gruppe\*Messzeitpunkt“, die eine stärkere Abnahme des Kompetenzerlebens der analogen Nutzer:innen gegenüber derjenigen der digitalen widerspiegelt, erreicht ein signifikantes Niveau. Damit zeigt sich insbesondere die dämpfende Wirkung von HyperDocs beim Abfall des Kompetenzerlebens, wodurch Schüler:innen im Verlauf einer Unterrichtsreihe, in der die Progression und die Anforderungsbereiche normalerweise zunehmen, länger ihr Kompetenzerleben bei der Bearbeitung von Arbeitsblättern aufrechterhalten können. Die glättende Wirkung der HyperDocs zeigt sich auch in der verringerten Standardabweichung in der Interventionsgruppe, wodurch weniger extreme Bewertungen auftreten. Die Gründe für das geringe Kompetenzerleben in der Vergleichsgruppe wurden bereits genannt: Die Schüler:innen dieser Gruppe zeigen teilweise einen höheren Cognitive Load und betrachten die Hilfen als weniger nützlich. Ein verringertes Kompetenzerleben tritt außerdem bei Überforderung auf (Ryan & Deci, 2017, S. 11), wie es die Daten beim Cognitive Load nahelegen. Zudem ist ein Zusammenhang zwischen Autonomie und Kompetenzerleben bekannt (Patall et al., 2008), sodass die verminderte Autonomie auch zu einem verminderten Kompetenzerleben beitragen kann. In der Interventionsgruppe liegt erneut eine stärkere Abnahme zwischen der zweiten und dritten Stunde vor, was wahrscheinlich auf den höheren Anforderungsbereich der dritten Stunde zurückzuführen ist.

Die Variable „Druck“ wurde lediglich deskriptiv analysiert. Dort zeigen sich bei den Jungen leichte Unterschiede zugunsten der Interventionsgruppe. Beide Gruppen erfahren eine leichte Zunahme des Drucks, was ebenfalls auf den erhöhten Cognitive Load in den letzten Stunden zurückgeführt werden könnte, da eine mittelstarke Korrelation zu diesem Konstrukt besteht. Gleichzeitig nimmt jedoch auch die wahrgenommene Wahlfreiheit (Autonomie) ab, was eine Erhöhung des Drucks zur Folge haben könnte, da sich die Schüler:innen in ihrer Bearbeitung weniger selbstständig fühlen. Der Druck liegt

insgesamt auf einem niedrigen Niveau, was darauf hindeutet, dass die Lernenden über die gesamte Erhebung kaum Druck beim Bearbeiten der Arbeitsblätter verspüren.

Interessanterweise lässt sich bei allen eben diskutierten Variablen der intrinsischen Motivation eine starke Abnahme zwischen der zweiten und dritten Stunde feststellen. Ein Erklärungsansatz könnte im Stundeninhalt begründet sein. In der dritten Stunde erfahren Schülerinnen einen starken Anstieg des Extraneous Load. Dieser könnte auf eine Aufgabe zurückzuführen sein, in der Lernende die Messreihe der photometrischen Messung in ein Koordinatensystem eintragen sollten, um die Kalibrierungsgerade zu erstellen. Der Intrinsic Load hingegen zeigt keine auffällige Zunahme und die Hilfenutzung nimmt in dieser Stunde ab. Zum Intrinsic Load bestehen jedoch, mit Ausnahme des Drucks und des Kompetenzerlebens, auch die schwächsten Zusammenhänge. Die Schüler:innen hätten sich zudem mehr Zeit zum Bearbeiten der Aufgaben in dieser Stunde gewünscht, was im Zusammenhang mit dem Zeichnen des Graphen auf ein Zeitproblem hindeutet. Dadurch wurden auch weniger Hilfen genutzt. In der gleichen Stunde erfährt die Bewertung der Nützlichkeit der Hilfen eine Abnahme und die Jugendlichen nutzten eher Hilfen aus dem Hilfebedarf heraus. Der Germane Load fällt bei den Schülerinnen ebenfalls wahrnehmbar in der dritten Stunde ab. Der Mental Effort steigt bei den Schülern. Die eben genannten Phänomene zeigen sich fast ausschließlich bei der Interventionsgruppe, was möglicherweise auf ein Problem bei der Eingabe der Kalibrierungsgeraden schließen lässt. Eventuell ist die Eingabe mit dem Apple Pencil für die Schüler:innen noch zu ungewohnt, um solche präzisen Eintragungen vorzunehmen. Zudem konnten die Lernende kein „digitales Lineal“ nutzen, um eine Kalibriergerade zu erstellen, sondern mussten sich einem analogen Lineal bedienen. Eine Kombination aus technischer sowie inhaltsbezogener Überforderung erscheint daher nachvollziehbar, da sich die Schüler:innen einerseits mit dem Lerngegenstand und andererseits mit der Technik auseinandersetzen mussten. Diese Beobachtung würde unter anderem den Abfall der Usability in Zusammenhang mit dem Recency-Effekt im Post-Test erklären (vgl. Kapitel 6.1.2).

Beim Vergleich der Schulformen zeigt sich in der Mittelstufe lediglich bei der wahrgenommenen Wahlfreiheit ein signifikanter Unterschied. Dort fühlen sich die Schüler:innen der Gesamtschule weniger autonom beim Lernen mit HyperDocs. An der Gesamtschule werden häufiger Formen der Binnendifferenzierung eingesetzt, bei denen die Schüler:innen autonomer lernen können (Letzel & Otto, 2019). Beim Interesse (SDT) und Kompetenzerleben beim Lernen mit HyperDocs liegen die Werte der Schülerinnen der

Gesamtschule leicht unterhalb von denen des Gymnasiums. Beim Druck liegen die Werte der Schüler etwas darunter. Lernende des Gymnasiums zeigen ein höheres Fachinteresse an Chemie, wobei dieses bei den Schülerinnen über die Reihe abnimmt. Es liegt jedoch am Ende immer noch über dem der Mädchen der Gesamtschule. Beim Fachinteresse Naturwissenschaften liegen die Werte der Schüler:innen der Gesamtschule unterhalb von denen des Gymnasiums. Unter Berücksichtigung der anderen Variablen fällt eine erhöhte Hilfenutzung der Schüler:innen der Gesamtschule in der zweiten und dritten Stunde auf. Zusätzlich bewerten insbesondere die Schülerinnen der Gesamtschule die Usability im Post-Test schlechter. Der Zusammenhang zwischen Usability und Interesse ist bereits bekannt (Karapanos et al., 2018). Zudem zeigen die Lernenden in zwei der vier Stunden einen höheren Intrinsic Load, in allen Stunden einen höheren Extraneous Load und die Schülerinnen der Gesamtschule einen niedrigeren Germane Load. Die Korrelationen zwischen der intrinsischen Motivation und dem Cognitive Load zeigen sich bei fast allen Variablen, insbesondere beim Extraneous und Germane Load, als signifikant. Die Ergebnisse deuten darauf hin, dass die Unterrichtsreihe trotz Lernhilfen für die Gesamtschüler:innen anspruchsvoller war als für die Gymnasiast:innen.

#### *Intrinsische Motivation und Interesse F\_MI: Oberstufe*

In der Oberstufe zeigen sich in Bezug auf die Zusammenhangsanalysen und die Abnahme des Interesses in der Interventionsgruppe die gleichen Ergebnisse wie in der Mittelstufe. Die Korrelationen zwischen dem Extraneous und Germane Load sind spätestens in der vierten Stunde zu allen Variablen der intrinsischen Motivation signifikant. Der Intrinsic Load zeigt zum Kompetenzerleben und zum Druck signifikante Zusammenhänge. Das Interesse (SDT) beim Lernen mit HyperDocs nimmt in der Interventionsgruppe über die Unterrichtsreihe bis zur dritten Stunde ab und bildet dort einen Tiefpunkt. Bei der Vergleichsgruppe sind ähnliche Ergebnisse bei den Schülerinnen beobachtbar, jedoch steigt bei den Jungen tendenziell das Interesse beim Lernen mit den differenzierten Arbeitsblättern an und erreicht in der vierten Stunde einen Höhepunkt. Beim Fachinteresse Chemie lässt sich eine signifikante Dreifache-Interaktion beobachten, wonach sich die Gruppen nach Geschlecht und Messzeitpunkt voneinander unterscheiden. Bei Schülern der Vergleichsgruppe nimmt das Interesse über die Unterrichtsreihe ab, bei Schülerinnen zu. Bei der Interventionsgruppe lässt sich nach Geschlecht genau der gegenteilige Effekt beobachten. Das Sachinteresse an Naturwissenschaften zeigt für die Schüler die gleichen

Trends. Bei den Schülerinnen nimmt in beiden Gruppen das Sachinteresse ab. Es liegen jedoch keine signifikanten Ergebnisse vor.

Die Abnahme des Interesses (SDT) beim Lernen mit Arbeitsblättern kann mithilfe des Cognitive Load teilweise erklärt werden. Sowohl der Intrinsic als auch der Extraneous Load zeigen einen Hochpunkt in der dritten Stunde, genauso wie der Mental Effort, wo das Interesse (SDT) einen Tiefpunkt erreicht. Bei beiden Loadtypen liegen die Werte der Schüler der analogen Nutzung in der zweiten und dritten Stunde über denen der digitalen. Der Germane Load zeigt jedoch keine deutliche Änderung in der dritten Stunde. Hierin liegt jedoch ebenfalls die Vergleichsgruppe unterhalb der Interventionsgruppe. Zusätzlich ist der Mental Effort ab der zweiten Stunde bei den Schülern der Vergleichsgruppe gegenüber der Intervention erhöht. Mit dem erhöhten Cognitive Load steigt auch die Nutzung der Lernhilfen in der dritten Stunde deutlich an. Die Schüler:innen geben zudem an, häufiger eine Hilfe aus dem Hilfebedarf heraus genutzt zu haben. In diesem Zusammenhang bewerten sie die Hilfen auch als nützlicher. Gründe für den erhöhten Load und die damit wahrscheinlich verbundene Hilfenutzung könnten in dem Lerngegenstand der dritten Stunde liegen. In der Unterrichtsstunde wurden der Aufbau und die Funktionsweise eines Photometers erarbeitet. Das Thema erweist sich in der Unterrichtspraxis als durchaus komplex, da es unter anderem auf mathematische Zusammenhänge, wie zum Beispiel das Lambert-Beersche-Gesetz zurückgreift. Weiterhin spielen physikalische Zusammenhänge eine große Rolle, wenn es um die Betrachtung der Wellenlänge und Absorptionsspektren geht. Auch die Abnahme der wahrgenommenen Wahlfreiheit (Autonomie) und des Kompetenzerlebens beeinflusst das Interesse negativ.

Bei der wahrgenommenen Wahlfreiheit beim Lernen mit Arbeitsblättern zeigt sich ein Tiefpunkt in der dritten Stunde bei den Schülerinnen der Vergleichs- und Interventionsgruppe. Sie starten auf dem gleichen Niveau bezüglich der wahrgenommenen Wahlfreiheit, wobei jedoch die Vergleichsgruppe diesbezüglich im Verlauf stärker abnimmt. Bei den Schülern ist dieser Trend nicht eindeutig zu erkennen. Zudem liegt die Vergleichsgruppe darin deutlich unterhalb der Interventionsgruppe. Zur vierten Stunde kommt es allerdings zu einer Annäherung der beiden Gruppen. Es können keine anderen offensichtlichen Gründe für die Abnahme der wahrgenommenen Wahlfreiheit in den beiden Gruppen gefunden werden. Möglicherweise liegen stundenspezifische Faktoren vor, wobei auf didaktischer und methodischer Ebene keine Gründe für den stetigen Abfall in der Interventionsgruppe gefunden werden konnten. Eventuell wirkt der Neuheitseffekt über die

wahrgenommene Wahlfreiheit, wodurch ein Abfall der Autonomie in Abhängigkeit von der Zeit zu einer Abnahme der intrinsischen Motivation führt. Für diese Hypothese spricht die stetige Abnahme der wahrgenommenen Wahlfreiheit (Autonomie) über die Unterrichtsstunden. Gegen diese Vermutung spricht die Abnahme der wahrgenommenen Wahlfreiheit in der Vergleichsgruppe. Dieser Erklärungsansatz könnte auch für die Mittelstufe herangezogen werden. Die generellen Unterschiede zwischen den Nutzer:innen der digitalen und analogen Variante können durch die höheren Auswahlmöglichkeiten bezüglich der digitalen Texteingabe und den Hilfen begründet werden. Dieser Sachverhalt wurde bereits für die Mittelstufe diskutiert.

Auch das Kompetenzerleben beim Lernen mit Arbeitsblättern erfährt eine Abnahme und in der dritten Stunde einen Tiefpunkt, der sich bei den Mädchen der Vergleichsgruppe sehr deutlich zeigt. Zwischen den Werten der Gruppen liegt eine Einheit des Mittelwertes. Dadurch begründet wird auch die signifikante Dreizehner-Interaktion, die bei der Analyse der einfachen, einfachen Haupteffekte eine signifikante Zweizehner-Interaktion „Gruppe\*Geschlecht“ zeigt. Die Werte der Jungen der analogen Variante liegen um circa 0.3 bis 0.4 Einheiten unter denen der digitalen. Eine mögliche Ursache ist erneut der Zusammenhang zur wahrgenommenen Wahlfreiheit (Autonomie) und dem Cognitive Load, der bereits weiter oben, bei der Abnahme des Interesses, näher analysiert wurde. Dabei kann ebenfalls die Schwierigkeit des Lerngegenstands als möglicher Erklärungsansatz herangezogen werden, wonach sich die Lernenden überfordert fühlen. Dennoch erfahren die digital lernenden Schüler:innen einen geringeren Einbruch des Kompetenzerlebens. Dadurch zeigt sich erneut, wie bereits bei der Mittelstufe diskutiert, das dämpfende Potential von HyperDocs bei Abnahme des Kompetenzerlebens, wenn Schüler:innen neue Inhalte erschließen müssen und ein höherer Anforderungsbereich vorliegt. Auch in der Oberstufe liegt in der Interventionsgruppe überwiegend eine geringere Standardabweichung vor, was dieses Ergebnis stützt.

Der Druck beim Lernen mit Arbeitsblättern steigt bei der Interventionsgruppe zur dritten Stunde an. Zwischen den Gruppen besteht aber insgesamt ein geringer Unterschied, der insbesondere in der zweiten und dritten Stunde zugunsten der Interventionsgruppe erkennbar wird. Die möglichen Gründe wurden bereits diskutiert. In der Literatur wird zudem für die älteren Klassenstufen zunehmend von negativen Emotionen gegenüber den Fächern Chemie und Physik berichtet, die für beide Geschlechter gelten (Dávila-Acedo et al., 2021).

Zusammenfassend zeigen sich für die Mittel- und Oberstufe makroskopisch die gleichen Tendenzen. So nimmt die intrinsische Motivation beim Lernen mit Arbeitsblättern über die Unterrichtsreihe ab, wobei sich stundenspezifische Charakteristika identifizieren lassen, die für eine stärkere Abnahme sorgen. Diese Abnahme des Interesses (SDT), der wahrgenommenen Wahlfreiheit (Autonomie) und des Kompetenzerlebens beim Lernen mit Arbeitsblättern lässt sich vermutlich mit der steigenden kognitiven Last begründen, die sich auch in einer häufigeren Nutzung der Lernhilfen widerspiegelt. Die Interventionsgruppe zeigt in den überwiegenden Stunden in den einzelnen Variablen der intrinsischen Motivation eine höhere Bewertung, wobei insbesondere in der Oberstufe eine differenzierte Betrachtung, je nach Geschlecht, erforderlich ist. Hier zeigen sich die Unterschiede der Interventionsgruppe insbesondere für die weiblichen Lernenden. Diese Unterschiede sind ebenfalls beim Fachinteresse Chemie zu erkennen. Während in der Mittel- und Oberstufe insbesondere die Schüler der Interventionsgruppe vom Einsatz der Tablets profitieren, führt bei den Schülerinnen dieser Einsatz zumindest in der Oberstufe zu einer Abnahme des Interesses. Insgesamt liegt ein interessensförderndes Potential vor, das zum situativen Interesse als catch-Komponente beitragen könnte. Damit würde, zumindest für die Schüler, die Steigerung des Fachinteresses Chemie erklärt werden können. Wichtige Mediatoren bilden dabei die wahrgenommene Wahlfreiheit und das Kompetenzerleben. Hierbei zeigt sich auch eine dämpfende Wirkung der HyperDocs auf die Minderung des Kompetenzerlebens, wenn neue Inhalt erschlossen werden und natürlicherweise das Erleben von Kompetenz aufgrund des steigenden Anforderungsbereichs sinkt.

Die geschlechtsspezifischen Unterschiede wurden oben bereits teilweise diskutiert und werden nun bei der nächsten Forschungsfrage detaillierter analysiert:

*F\_M2: : Inwieweit unterscheiden sich das Fachinteresse Chemie und die intrinsische Motivation zwischen den Geschlechtern beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

*Intrinsische Motivation und Interesse F\_M2: Mittelstufe*

In Teilen wurde auf die Differenzen zwischen den Geschlechtern bereits in der vorherigen Forschungsfrage eingegangen, wenn diese in Abhängigkeit von der Gruppenzugehörigkeit auftraten. Die Unterschiede werden nun nochmals differenzierter und detaillierter dargelegt. Beim Fachinteresse Chemie liegen die weiblichen Lernenden leicht unter den Jungen, wobei die Mädchen der Vergleichsgruppe das geringste Interesse aufweisen. Hierfür finden sich wieder die bekannten Belege aus der Literatur, die bereits bei

Beantwortung der ersten Forschungsfrage genannt wurden (Osborne et al., 2003, Trumper, 2006b). Zudem zeigen sie, nach eigenen Angaben, geringere Fähigkeiten in Chemie, was einen Einfluss auf die Interessensentwicklung, zum Beispiel über das Kompetenzerleben, nehmen könnte. Bei allen Teilgruppen steigt das Fachinteresse kaum wahrnehmbar an, bei den Jungen der Tabletgruppe jedoch am meisten. Diese Ergebnisse zeigen sich auch in anderen Studien, bei denen insbesondere die Schüler von der Nutzung digitaler Endgeräte profitieren, auch im Fach Biologie, in dem keine geschlechtsspezifischen Unterschiede gefunden werden (Bruckermann et al., 2016). Begründet werden könnte der stärkere Effekt bei den Jungen durch die höhere Affinität zum Lernen mit digitalen Werkzeugen (Fabian & Topping, 2019; Tossavainen & Hirsto, 2017). Ein weiterer denkbarer Grund für die Unterschiede zwischen den Geschlechtern könnten geringere tabletbezogene Kompetenzen der Schülerinnen gegenüber den Schülern sein. Durch die geringeren digitalen Kompetenzen erleben die Schülerinnen gegebenenfalls weniger Fachkompetenz, was zu einer Verringerung des Fachinteresses führt. Die Unterschiede bei den tabletbezogenen Kompetenzen konnten jedoch nicht oder nur teilweise, mit einer sehr geringen Differenz, beobachtet werden, sodass diese die Unterschiede ebenfalls nicht erklären können. Beim Sachinteresse an Naturwissenschaften steigt das Interesse der Jungen sowohl in der Interventionsgruppe als auch in der Vergleichsgruppe. Dies mündet in einen signifikanten Haupteffekt des Messzeitpunkts und Geschlechts. Die Unterrichtsreihe zeigt sich für Schüler damit insgesamt als interessanter. Bei den Schülerinnen bleibt das Interesse konstant. Hierbei können die gleichen Wirkmechanismen (geringeres Kompetenzerleben, weniger Interesse an Naturwissenschaften) angenommen werden, wobei der Effekt nicht allein auf die HyperDocs zurückzuführen ist.

In Bezug auf das Interesse (SDT) beim Lernen mit Arbeitsblättern liegt in der Mittelstufe eine signifikante Zweiwege-Interaktion „Geschlecht\*Gruppe“ vor. Die Schülerinnen der Interventionsgruppe starten in der ersten Stunde mit einem höheren Interesse, das jedoch zur dritten Stunde abfällt und auf dem gleichen Niveau wie bei den Jungen liegt. Dieses Ergebnis steht damit zum Teil in Widerspruch zu anderen Studien, die für Mädchen ein geringeres Interesse in den Naturwissenschaften, mit Ausnahme von Biologie, zeigten (Osborne et al., 2003, Trumper, 2006b). Als Grund dafür wird das unterschiedliche Selbstkonzept der Schüler:innen in den jeweiligen Fächern herangezogen (Jansen et al., 2013, Osborne et al., 2003, Sax et al., 2015). Es ist daher denkbar, dass der Einsatz der HyperDocs das Selbstkonzept der Nutzerinnen beim Bearbeiten der Arbeitsblätter stärkt. Dabei können jedoch nicht allein die Lernhilfen dafür verantwortlich gemacht werden,



denn bei Betrachtung der weiblichen Lernenden fällt insbesondere der hohe Unterschied zwischen den beiden Gruppen (Interventions- und Vergleichsgruppe) in der ersten Stunde auf. Die Differenz zwischen der Interventions- und der Vergleichsgruppe beträgt 1.5 Einheiten des Mittelwertes zugunsten der Intervention. In der Vergleichsgruppe liegen die Werte der Schülerinnen damit unter denen der Schüler. Folglich müssten die digitalen Lernhilfen gegenüber der analogen Unterstützung zu einer höheren Selbstwirksamkeit führen. Gegebenenfalls nutzen anteilig mehr Schülerinnen in einer digitalen Lernumgebung die Lernhilfen. Diese Hypothese kann allerdings mit den vorliegenden Daten nicht genau überprüft werden. Das Sachinteresse an Naturwissenschaften konnte bei den Schülern beider Gruppen gleichermaßen gefördert werden, wobei die Ergebnisse nicht signifikant sind. Die Schülerinnen erfahren keine nennenswerte Zunahme. Warum letztlich das gesteigerte Interesse beim Bearbeiten der Arbeitsblätter bei den Schülerinnen nicht in das Fachinteresse und Sachinteresse mündet, kann nicht genau gesagt werden. Es ist davon auszugehen, dass es sich bei dem Interesse um ein stabiles Persönlichkeitsmerkmal handelt, das bei den Mädchen womöglich einer längeren Intervention bedarf, um beispielsweise das Selbstkonzept im Fach Chemie zu stärken. Dafür müsste eine weitere hold-Komponente für eine Stabilisierung des situationalen Interesses vorhanden sein (Krapp, 2007).

Bei der wahrgenommenen Wahlfreiheit (Autonomie) beim Lernen mit Arbeitsblättern liegen die Geschlechter auf einem ähnlichen Niveau. Im Verlauf der Unterrichtsreihe nimmt bei den Schülerinnen, die mit HyperDocs arbeiten, tendenziell die wahrgenommene Wahlfreiheit ab der dritten Stunde stärker ab. Die Ergebnisse stehen in Einklang mit anderen Studien, die ebenfalls bei Schülerinnen ein geringes Autonomieerleben feststellen konnten (Patall et al., 2018). Die Gründe für die Unterschiede ab der dritten Stunde können nicht eindeutig erklärt werden. Sinnvoll erscheint der Bezug zum Intrinsic und Extraneous Cognitive Load. Beide Loadtypen erfahren bei den Schülerinnen in der dritten Stunde ein Maximum, das über dem der Schüler liegt. Gerade der Extraneous Load zeigt mittelstarke Korrelationen zur wahrgenommenen Wahlfreiheit.

Beim Kompetenzerleben beim Lernen mit Arbeitsblättern liegen wieder größere Differenzen zwischen den Geschlechtern vor. Die Analysen zeigen eine knapp nicht-signifikante Zweiwege-Interaktion „Geschlecht\*Gruppe“ und eine signifikante Interaktion „Geschlecht\*Messzeitpunkt“. Während sich in den ersten beiden Stunden die Jungen und Mädchen der Interventionsgruppe auf dem gleichen Niveau befinden, nimmt das

Kompetenzerleben bei den Schülerinnen in der dritten Stunde stärker ab. Begründet werden könnte die Abnahme auch mit dem gleichzeitigen Abfall des Kompetenzerlebens, da ein theoretischer Zusammenhang zwischen den beiden Variablen besteht (Patall et al., 2008). Die Ergebnisse lassen darauf schließen, dass die Schülerinnen mit dem Zeichnen der Kalibriergerade und der Auswertung der Messreihe stärker gefordert waren als Schüler. Bei der Vergleichsgruppe liegen die Werte der Schüler über denen der weiblichen Lernenden. Diese Ergebnisse zeigten sich bereits in anderen Studien mit anderen digitalen Tools (Bruckermann et al., 2016; Conradt & Bogner, 2016). In der vierten Stunde unterscheiden sich die Schülerinnen der beiden Erhebungsgruppen um eine ganze Einheit des Mittelwertes, wodurch HyperDocs die Schülerinnen in ihrem Kompetenzerleben unterstützen und sich die bereits diskutierte dämpfende Wirkung bei der Minderung des Kompetenzerlebens zeigt.

Unterschiede im Druck sind beim Lernen mit Arbeitsblättern zwischen den Geschlechtern kaum vorhanden. Die Schüler erfahren teilweise mehr Druck. In der dritten Stunde ist erneut der charakteristische Abfall beziehungsweise hier eine Zunahme des Drucks bei den Mädchen zu beobachten.

Die eben genannten Unterschiede lassen sich größtenteils auf die Korrelationen der Variablen der intrinsischen Motivation untereinander und den Cognitive Load zurückführen. Der gleiche Erklärungsansatz wurde bereits bei der ersten Forschungsfrage herangezogen und wird nun in Bezug auf das Geschlecht dargelegt. Unter Betrachtung des Intrinsic Load fällt ein stärkerer Anstieg bei den Werten der Schülerinnen in beiden Erhebungsgruppen ab der zweiten Stunde auf. Die Interaktion „Geschlecht\*Messzeitpunkt“ ist nur knapp nicht signifikant, was diese These unterstützt. Beim Extraneous Load liegen beide Geschlechter in der Vergleichsgruppe über der Interventionsgruppe. Ab der zweiten Stunde kommt es hier auch zu einer stärkeren Zunahme. In der dritten Stunde zeigt sich eine deutliche Zunahme des Load bei den digitallernenden Schülerinnen. Demnach erfahren die Schülerinnen durch den Einsatz von HyperDocs eine höhere Last als Schüler. Zudem liegen die Fähigkeiten in Chemie der Schülerinnen nach eigenen Angaben im Mittel unter denen der Schüler. Dadurch erfahren sie womöglich einen höheren Cognitive Load und ein geringeres Kompetenzerleben. Es ist außerdem bekannt, dass ein erhöhtes multimediales Scaffolding die extrinsische Last erhöhen kann (Bruckermann et al., 2017). Der t-Test ist signifikant in Bezug auf das Geschlecht. Dadurch erreicht die Interaktion „Geschlecht\*Messzeitpunkt“ ein signifikantes Niveau. Beim Germane Load liegen die

Werte der Schülerinnen jedoch über denen der Schüler, was sich nicht eindeutig mit dem theoretischen Konstrukt des Cognitive Load vereinbaren lässt. Beim Mental Effort hingegen erfahren die Mädchen beider Gruppen erneut ab der zweiten Stunde eine höhere Zunahme. Die Unterschiede im Cognitive Load führen vermutlich auch zu den Unterschieden in der Hilfenutzung. Der Haupteffekt des Geschlechts ist zwar knapp nicht signifikant, jedoch nutzen die Schülerinnen in der zweiten und vierten Stunde häufiger eine Hilfe als die Jungen. Bei der Usability hingegen treten keine nennenswerten Unterschiede auf, die die Differenzen in der Interventionsgruppe erklären könnten. Genauso liegt die tabletbezogene Selbstwirksamkeit der Schülerinnen zwar unter der der Schüler, sie befindet sich insgesamt jedoch auf einem hohen Niveau. In der Gesamtschule zeigen sich die gleichen Trends wie am Gymnasium, wonach bei den meisten Variablen die Werte der Schülerinnen unterhalb der Schüler liegen.

#### *Intrinsische Motivation und Interesse F\_M2: Oberstufe*

Auch in der Oberstufe liegen deutliche Unterschiede zwischen den Geschlechtern vor. So zeigen Schülerinnen beider Untersuchungsgruppen zur ersten Stunde ein höheres Interesse (SDT) beim Lernen mit Arbeitsblättern. Die Unterschiede sind signifikant. Danach nähern sich die Schüler der Vergleichsgruppe denen der digitalen Intervention an, während bei den Schülerinnen die beiden Gruppen parallel verlaufen und den bereits diskutierten Tiefpunkt in der dritten Stunde erreichen. Danach zeigen die Mädchen, wie auch die Jungen, der Vergleichsgruppe eine deutliche Zunahme und nähern sich der Interventionsgruppe an. Der eben beschriebene Verlauf führt zu einer signifikanten Zweibege-Interaktion. Bei Betrachtung der Werte des Fachinteresses Chemie zeigen die Schülerinnen, die mit HyperDocs lernten, eine Abnahme, während die weiblichen Lernenden der analogen Gruppe eine Zunahme verzeichnen. Die gegenteiligen Bewegungen sind bei den Jungen zu sehen, wodurch die Dreibege-Interaktion ein signifikantes Niveau erreicht. Die Unterschiede zwischen den Geschlechtern sind womöglich mit der unterschiedlichen Affinität zu iPads und dem damit verbundenen Nutzen zum Lernen zu begründen (Fabian & Topping, 2019; Tossavainen & Hirsto, 2017), wobei Schülerinnen lediglich eine geringere Selbstwirksamkeit in Bezug auf Tablets aufweisen. Beim Lernen sehen sie, laut eigenen Angaben, Tablets als genauso nützlich zum Lernen an. Beim Sachinteresse an Naturwissenschaften zeigen sich bei den Jungen die gleichen Tendenzen wie beim Fachinteresse Chemie. Bei den Mädchen führt die Intervention bei beiden Gruppen zu keiner großen Veränderung über den Zeitraum der Reihe.

Bei der wahrgenommenen Wahlfreiheit (Autonomie) beim Lernen mit Arbeitsblättern liegen die Geschlechter der Intervention gleich auf, während in der Vergleichsgruppe die Schülerinnen höher liegen. Sie erfahren allerdings eine stärkere Abnahme der Werte als die Interventionsgruppe. Auch bei den männlichen Lernenden nimmt die wahrgenommene Wahlfreiheit stärker ab. Sie nähert sich allerdings in der vierten Stunde wieder derjenigen der Interventionsgruppe an. Der beschriebene Verlauf mündet jedoch in keine signifikanten Effekte, die das Geschlecht betreffen.

Auch beim Kompetenzerleben beim Lernen mit Arbeitsblättern sind die Geschlechter der digitalen Gruppe gleich auf. In der Vergleichsgruppe starten die Schülerinnen über den Schülern, wobei ihre Werte in der dritten Stunde deutlich abfallen und damit unter denen der Jungen liegen. Hierbei findet jedoch eine Erholung in der vierten Stunde statt. Beim Druck zeigen sich die Geschlechter aller Gruppen in etwa gleich.

Als Erklärungsansatz für die Unterschiede zwischen den Geschlechtern kann in Teilen der Cognitive Load herangezogen werden, wobei hier gemischte Ergebnisse bestehen. Die Schülerinnen der Intervention zeigen einen höheren Intrinsic Load, die Lernenden der Vergleichsgruppe liegen leicht darunter. Die Effekte sind allerdings nicht signifikant. Auch beim Extraneous Load liegen die Schülerinnen beider Gruppen über den Schülern. Die höhere extrinsische Last bei den weiblichen Nutzerinnen der Tablets kann über deren geringere Affinität gegenüber iPads erklärt werden (Tossavainen & Hirsto, 2017). Da die Vergleichsgruppe unterhalb der Interventionsgruppe liegt, ist anzunehmen, dass HyperDocs in der Oberstufe eine leichte Erhöhung der extrinsischen Last verursachen. Dies ist womöglich auf die verschiedenen multimedialen Hilfen zurückzuführen (Bruckermann et al., 2017). In der zweiten und dritten Stunde steigt der Load jedoch bei den Schülern der Vergleichsgruppe deutlich an und liegt damit über dem der digitalen Gruppe. Die Ergebnisse sind allerdings erneut nicht signifikant. Der Germane Load ist hingegen bei den Schülerinnen der Intervention größer als bei den Schülern. Die Vergleichsgruppe liegt leicht darunter. Der Mental Effort ist bei Schülerinnen der Intervention wieder erhöht. Die Schüler der Vergleichsgruppe bewerten den Mental Effort ab der zweiten Stunde höher als die der Interventionsgruppe. Bei der Usability nimmt die Nützlichkeit bei den Schülerinnen stärker ab. Die Abnahme erreicht allerdings kein signifikantes Niveau. Die Zufriedenheit ist bei den Schülerinnen wiederum höher. Die anderen beiden Variablen der Usability sind in etwa gleich hoch bewertet. In der dritten Stunde nutzen die Schülerinnen zudem etwas häufiger eine Hilfe. Die Differenz ist jedoch gering.

Bei der Vergleichsgruppe nutzen die Schüler in der ersten und dritten Stunde mehr Hilfen. Es handelt sich dabei um eine Selbstangabe, wodurch die Ergebnisse verzerrt sein könnten. Die Usability und Hilfenutzung können daher nicht als unmittelbarer Erklärungsansatz herangezogen werden. Auch die tabletbezogene Selbstwirksamkeit liegt bei den Schülerinnen auf einem hohen Niveau. Die Ergebnisse sind in der Oberstufe nicht eindeutig interpretierbar, sodass die Unterschiede zwischen den Geschlechtern nicht vollständig erklärt werden können.

Zusammengefasst zeigen sich bezüglich dieser Forschungsfrage in der Mittelstufe Unterschiede zwischen den Geschlechtern. Teilweise profitieren die Schülerinnen in den ersten Stunden stärker vom Einsatz der HyperDocs. In der Vergleichsgruppe hingegen zeigen sich die üblichen geschlechtsspezifischen Unterschiede zugunsten der Jungen. In der Oberstufe zeigen sich zwischen den Geschlechtern weniger Unterschiede. Teilweise liegen die Schülerinnen bei dem Interesse (SDT) und der wahrgenommenen Wahlfreiheit beim Lernen mit Arbeitsblättern über den Schülern, wobei diese Effekte lediglich bei der Interventionsgruppe gelten. In der Vergleichsgruppe hingegen fallen die Werte der Schülerinnen im Verlauf der Unterrichtsreihe stark ab, womit ein positiver Einfluss von HyperDocs auf die weiblichen Lernenden feststellbar ist. Dieser positive Effekt mündet jedoch nicht in ein höheres Fachinteresse in Chemie. Als möglicher Grund für die Unterschiede kommt der Cognitive Load in Frage. Letztlich können die Unterschiede jedoch nicht allein damit begründet werden.

Die letzte Forschungsfrage beschäftigt sich mit dem bereits von Karapanos et al. (2018) aufgedeckten Zusammenhang zwischen dem Interesse und der Usability:

*F\_M3: Welcher Zusammenhang besteht zwischen der Usability und der intrinsischen Motivation beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

*Intrinsische Motivation und Interesse F\_M3: Mittelstufe*

In der Mittelstufe zeigen alle Variablen der intrinsischen Motivation zu allen Variablen der Usability signifikante positive Zusammenhänge. Die Stärke der Korrelationen nimmt in der vierten Stunde nochmals zu. Sie liegt auf einem mittleren bis hohen Niveau. Insbesondere das Interesse (SDT) beim Lernen mit HyperDocs korreliert stark mit der Usability. Damit können die Ergebnisse von Karapanos et al. (2018) auch für HyperDocs und den Chemieunterricht verifiziert werden. Über die Wirkungsrichtung kann keine Aussage getroffen werden. Hierbei sind beide Wirkmechanismen denkbar. Durch ein hohes

Interesse, unter anderem verursacht durch die digitale Intervention, bewerten die Schüler:innen die Usability ebenfalls als hoch. Sie setzten sich nicht nur mit dem Stundeninhalt besser auseinander, sondern auch mit dem System. Der Zusammenhang zwischen der Einfachheit der Nutzung und dem Spaß bei der Nutzung eines technischen Systems ist bereits bekannt (Venkatesh, 1999).

Auf der anderen Seite kann die hohe Usability der HyperDocs zu einer positiven Lernerfahrung führen, sodass der interessensfördernde Effekt des digitalen Mediums weiter verstärkt wird. Unter diesem Erklärungsansatz wäre die Usability als notwendige Voraussetzung für eine positive Erfahrung zu verstehen, wie es bereits einige Autor:innen formuliert haben (Hassenzahl et al., 2010). Weshalb die Stärke der Zusammenhänge in der vierten Stunde zunimmt, kann nicht genau gesagt werden. Gegebenenfalls wirkt das Vorhandensein des Neuheitseffekts in der ersten Stunde stärker und überdeckt damit in Teilen den Zusammenhang zur Benutzerfreundlichkeit. Auch das Kompetenzerleben, als ein Mediator zum Interesse (SDT), könnte positiv durch die Usability beeinflusst werden. Diejenigen Schüler:innen, die die HyperDocs als einfach zu nutzen und nützlich finden, werden sich aufgrund der geringeren technischen Hürden als kompetenter und selbstwirksamer fühlen als jene, die die Nutzung der digitalen Arbeitsblätter vor Probleme stellt. Die gleiche Begründung kann für den Druck herangezogen werden, wonach eine hohe Usability die Schüler:innen weniger unter Druck setzt, da sie sich nicht überproportional mit der technischen Bedienung auseinandersetzen müssen. Bei der wahrgenommenen Wahlfreiheit fällt die Begründung des Zusammenhangs schwerer. Hier kann vermutet werden, dass sich Lernende, die eine geringe Benutzerfreundlichkeit erfahren, von HyperDocs eingeschränkt fühlen, da sie auf mehr Probleme stoßen und damit den Lerninhalt nicht wie gewohnt erarbeiten können. Die Usability wäre demnach bei den eben beschriebenen Wirkmechanismen ebenfalls eine nötige Voraussetzung, damit digitale Lernumgebungen ihr motivationsförderndes Potential entfalten können.

#### *Intrinsische Motivation und Interesse F\_M3: Oberstufe*

In der Oberstufe fallen die Ergebnisse nicht eindeutig aus. Manche Zusammenhänge erreichen nicht das Signifikanzniveau, wie beispielsweise der Zusammenhang zwischen der wahrgenommenen Wahlfreiheit (Autonomie) und der Nützlichkeit. Tendenziell ist in der vierten Stunde eine Zunahme der Stärke der Korrelationen erkennbar, insbesondere bei den Zusammenhängen mit dem Interesse (SDT) beim Lernen mit HyperDocs. Weshalb die Zusammenhänge zwischen den Variablen in der Oberstufe schwächer sind oder sie

das Signifikanzniveau nicht immer erreichen, kann nicht eindeutig gesagt werden. Prinzipiell sollten die bereits beschriebenen Wirkungen zwischen den Variablen der Motivation und der Benutzerfreundlichkeit auch in der Oberstufe wirken. Möglicherweise zeigen die Schüler:innen auch unabhängig von der Usability eine hohe Motivation, da die Stichprobe Leistungskurse beinhaltet. Dies zeigt sich auch bei Betrachtung des Fachinteresses Chemie gegenüber der Mittelstufe. Die Schüler:innen weisen daher eine höhere Motivation am Fachinhalt auf und werden durch die Usability weniger stark beeinflusst. Gegebenenfalls setzen sie auch eine hohe Usability voraus, da sie häufiger mit digitalen Endgeräten lernen.

Zusammenfassend lässt sich für diese Forschungsfrage festhalten, dass die bisherigen Ergebnisse der Literatur auch für das Fach Chemie und die Mittelstufe verifiziert werden konnten. Es zeigen sich positive Zusammenhänge zwischen den Variablen der intrinsischen Motivation und denen der Usability. In der Mittelstufe sind die Korrelationen stärker. Als möglicher Wirkungsmechanismus wird die Richtung Usability zu Motivation bevorzugt.

### 6.1.2 Usability

Ziel dieser Arbeit ist es, verschiedene offene Fragen in Bezug auf die Usability in einem Lehr-Lern-Szenario zu beantworten. Im Folgenden werden die Forschungs- und Detailfragen beantwortet und die Ergebnisse diskutiert. Als erstes wurde folgende Forschungsfrage formuliert:

*F\_U1: Wie bewerten Schüler:innen die Usability von HyperDocs während des Einsatzes im Regelunterricht über den Zeitraum einer vierstündigen Unterrichtsreihe?*

In Bezug auf die diese Frage lässt sich festhalten, dass sowohl die Schüler:innen der Mittelstufe als auch der Oberstufe die Usability von HyperDocSystems als hoch einstufen. Bezogen auf die fünfstufige Likert-Skala der einzelnen Variablen der Benutzerfreundlichkeit (Nützlichkeit, Zufriedenheit, Einfachheit der Nutzung, Einfachheit des Lernens der Nutzung) liegt der Mittelwert der Variable Nützlichkeit im Bereich von 3.50 (Oberstufe) und 4.00 (Mittelstufe), während die anderen Variablen eine höhere Bewertung erfahren. Gerade die Einfachheit der Nutzung und die Einfachheit des Lernens werden nahezu mit dem maximalen Skalenwert beurteilt. Für diese beiden Aspekte der Usability liegt das Niveau in der Mittel- und Oberstufe gleich hoch. Die Nützlichkeit und

Zufriedenheit bewerten Lernende der Oberstufe geringer als die der Mittelstufe. Ein Grund hierfür könnte in der höheren Erwartungshaltung der Lernenden der Oberstufe liegen, da mehr Jugendliche mit einem Tablet lernen und gegebenenfalls größere Ansprüche an ein Tool wie HyperDocSystems stellen. Gegen diese Hypothese spricht jedoch der nicht-signifikante Zusammenhang zwischen der Nutzungshäufigkeit von Tablets im privaten und schulischen Bereich und der Usability. Im Vergleich zu einfachen Notiz-Apps, wie beispielsweise Goodnotes oder Notability, bietet HyperDocSystems jedoch weniger Funktionalitäten und Individualisierungen im Bereich der Freihandeingabe. Insbesondere in der Oberstufe müssen häufiger umfangreichere Texte beziehungsweise Antworten auf Aufgaben geschrieben werden, wodurch die Schreibfunktionalität einen großen Stellenwert einnimmt. Auch diese Hypothese lässt sich jedoch nicht mit den zur Verfügung stehenden Daten eindeutig belegen. Sowohl die Mittel- als auch die Oberstufe bewerten die Texteingabe mittels Tastatur oder Hand ebenfalls mit einem Mittelwert von über 4.00. In der Oberstufe lässt sich zwar eine höhere Abnahme der Usability über den Messzeitraum beobachten, allerdings bleibt der Wert über 4.00. Die Darstellung der Arbeitsblätter und Hilfen wird ebenfalls bei beiden Gruppen als sehr hoch eingestuft. Sie liegen in etwa auf dem gleichen Niveau. Ein weiterer Erklärungsansatz für die unterschiedliche Bewertung der Usability zwischen Mittel- und Oberstufe könnte in dem Interesse liegen. Karapanos et al. (2018) konnte bereits einen Zusammenhang zwischen der Usability und dem Interesse aufzeigen. Dieser Zusammenhang konnte in dieser Studie bestätigt werden. Demnach liegt das Gesamtniveau des aktuellen Interesses bei der Oberstufe unter dem der Mittelstufe. Ein weiterer Zusammenhang besteht zum Cognitive Load, der bereits aus der Literatur bekannt ist (Ibili & Billingham, 2019; Koć-Januchta et al., 2022). Auch bei den Zusammenhangsanalysen dieser Variablen konnten viele signifikante negative Korrelationen gefunden werden. Da allerdings beide Gruppen in der ersten Stunde, auf die sich die erste Usabilitymessung bezieht, in etwa den gleichen Load erfahren haben, kann dieser Erklärungsansatz ausgeschlossen werden. Als letzter Erklärungsansatz wird der Entwicklungsstand der Schüler:innen herangezogen. Demnach ist zu vermuten, dass die Oberstufelernenden die Usability differenzierter bewerten als die der Mittelstufe. Diese Vermutung basiert auf der gleichen Argumentation wie die der Nutzungshäufigkeit, die weiter oben beschrieben wurde. Die älteren Schüler:innen haben demnach mehr Erfahrung mit Tablets und einen höheren Kompetenzstand, wodurch sie die Benutzerfreundlichkeit differenzierter und kritischer bewerten.



Insgesamt lässt sich jedoch für beide Gruppen eine hohe Usability feststellen, die neben der Gesamtusability des Systems, auch die einzelnen Komponenten, wie die Textfelder oder die Darstellung der Arbeitsblätter und Hilfen, betrifft. Besonders hervorzuheben ist hierbei die Einfachheit der Nutzung und die Einfachheit des Lernens der Nutzung, wodurch geringe Einstiegshürden in das System bestehen. Vermutlich können jüngere Schüler:innen ebenfalls schnell in HyperDocSystems eingeführt werden, wodurch sich die Anwendung für niedrigere Klassenstufen eignet.

Die nächste Forschungsfrage, die formuliert wurde, lautet wie folgt:

*F\_U2: Inwieweit bestehen Unterschiede zwischen den Geschlechtern bei der Bewertung der Usability von HyperDocs durch Schüler:innen über den Zeitraum einer vierstündigen Unterrichtsreihe?*

*Usability F\_U2: Mittelstufe*

Die Forschungsfrage wurde im Laufe der Auswertung um die Komponente „Schulform“ in der Mittelstufe ergänzt, da sich Unterschiede zwischen der Gesamtschule und dem Gymnasium bei der explorativen Datenanalyse zeigten. Diese Unterschiede erreichen bei den Variablen „Nützlichkeit“, „Zufriedenheit“ und „Einfachheit des Lernens der Nutzung“ kein signifikantes Niveau. Die Einfachheit der Nutzung zeigt eine signifikante Dreiwege-Interaktion, wonach sich die Geschlechter in Abhängigkeit von der Zeit und der Schulform voneinander unterscheiden. Die Interaktion wird dadurch hervorgerufen, dass bei den Schülerinnen der Gesamtschule die Usability zwischen den Messzeitpunkten stärker abnimmt als bei denen des Gymnasiums. Gleichzeitig fällt die Usability bei den Schülern des Gymnasiums etwas stärker ab als bei denen der Gesamtschule. Die anschließende Untersuchung des einfachen, einfachen Haupteffekts zeigte keine signifikanten Ergebnisse, was auf eine geringe Effektstärke dieser Interaktion schließen lässt. Ansonsten zeigen sich unter Betrachtung beider Messzeitpunkte keine signifikanten Unterschiede zwischen den Geschlechtern. Zum ersten Messzeitpunkt bewerten sowohl Mädchen als auch Jungen die Usability von HyperDocSystems unabhängig von der Schulform gleich hoch. Bei der Variable „Zufriedenheit“ liegen die Werte der Schülerinnen sogar über denen der Schüler. Dies gilt allerdings nur für den ersten Messzeitpunkt. Bei Betrachtung des zweiten Messzeitpunktes nimmt der Skalenwert der Variablen „Nützlichkeit“, „Zufriedenheit“ und „Einfachheit der Nutzung“ bei den Schülerinnen der Gesamtschule stärker ab als bei denen des Gymnasiums und bei den Jungen. Wie bereits bei Beantwortung der ersten Forschungsfrage zur Usability, kann hier die Vermutung geäußert werden, dass

das aktuelle Interesse die Bewertung der Usability beeinflusst hat. Demnach zeigen die weiblichen Lernenden der Gesamtschule ein geringeres Interesse als die des Gymnasiums. Zudem fällt das Interesse zur vierten Stunde stärker ab, wodurch auch der Abfall in der Usability bei dieser Teilgruppe begründet werden könnte. Einen weiteren Erklärungsansatz bieten der Cognitive Load, der in der Gesamtschule höher ist, und die tabletbezogenen Kompetenzen, die wiederum bei den Schüler:innen dieser Schulform geringer sind.

#### *Usability F\_U2: Oberstufe*

In der Oberstufe fallen die Ergebnisse in Bezug auf das Geschlecht weniger einheitlich aus, wobei bei dieser Gruppe aufgrund der geringen Stichprobengröße der Gesamtschule keine Analyse der Schulform erfolgte. In der Oberstufe liegen die Werte der Schülerinnen bei der Nützlichkeit über denen der Schüler. Allerdings fällt dieser Wert im Post-Test stärker ab, wodurch die Jungen diesbezüglich zum zweiten Messzeitpunkt über den Mädchen liegen. Bei der Zufriedenheit liegen die Werte der Schülerinnen sowohl zum ersten als auch zum zweiten Messzeitpunkt über denen der Schüler, und bei der Einfachheit den Variablen „Einfachheit der Nutzung“ und „Einfachheit des Lernens der Nutzung“ liegen keine Unterschiede vor. Dabei ist anzumerken, dass die Unterschiede bei allen Variablen nicht signifikant sind. Die geringen Unterschiede zwischen den Teilgruppen bei den zuletzt genannten Variablen haben sich bereits beim Vergleich zwischen der Mittel- und Oberstufe gezeigt. Ein möglicher Grund dafür ist die schwächere Korrelation des Interesses zur Einfachheit der Nutzung und des Lernens im Vergleich zur Nützlichkeit oder Zufriedenheit. Diese Beobachtung würde auch die bisherigen Hypothesen unter Einbezug des Interesses bekräftigen, wobei über die Kausalität letztlich keine Aussage getroffen werden kann.

In Bezug auf die zweite Forschungsfrage lässt sich zusammenfassen, dass zwischen den Geschlechtern und Schulformen keine signifikanten Unterschiede in der Mittelstufe bestehen. Tendenziell bewerten die Schülerinnen der Gesamtschule die Usability von HyperDocSystems im Verlauf der Nutzung als schlechter. Eine mögliche Ursache liegt in dem geringeren aktuellen Interesse dieser Schülerinnen. In der Oberstufe sind keine großen Unterschiede zwischen den Geschlechtern feststellbar. Die Ergebnisse sind als positiv einzustufen, da in anderen Studien Unterschiede zwischen den Geschlechtern zugunsten des männlichen Geschlechts gefunden wurden (Schmidt et al., 2013; Tolle et al., 2020; Ong & Lai, 2006). HyperDocs sind folglich für beide Geschlechter gleich gut nutzbar.

Die dritte Forschungsfrage wurde bereits in Teilen in die obere Diskussion miteinbezogen. Sie soll nun ausführlicher diskutiert werden:

*F\_U3: Inwieweit tritt eine Veränderung auf in der Bewertung der Usability durch Schüler:innen in Bezug auf HyperDocs über eine Einsatzdauer von vier Unterrichtsstunden?*

Bei allen vier Variablen zeigt sich sowohl in der Mittel- als auch in der Oberstufe eine signifikante Abnahme. Aufgrund der Anwendung einer robusten ANOVA konnten keine Effektstärken gewonnen werden. Unter Berücksichtigung der deskriptiven Daten zeigt sich jedoch eine stärkere Abnahme über die beiden Messzeitpunkte bei den Variablen „Nützlichkeit“ und „Zufriedenheit“. Die Ergebnisse stehen teilweise in Widerspruch mit der vorhandenen Literatur (Lee & Ha, 2019), wonach Anwender:innen mit zunehmender Nutzungsdauer ein Softwaresystem besser kennenlernen und dadurch die subjektive Usability des Systems steigt. Auf der anderen Seite wird in der Literatur der Recency-Effekt beschrieben, der eine mögliche Erklärungsgrundlage bietet (Hassenzahl & Sandweg, 2004). Manche Autor:innen konnten zeigen, dass die Bewertung der Usability auf die letzte Erfahrung vor der Messung zurückzuführen ist. Verstärkt wird dieser Effekt durch die sogenannte „peak-end-rule“ (Bruun & Ahm, 2015, S. 241), wonach das Erreichen eines Meilensteines besser erinnert wird (Wang, 2021). Unter diesem Erklärungsansatz lässt sich eventuell die stärkere Abnahme der Nützlichkeit und Zufriedenheit begründen, da die Schüler:innen Probleme beim Zeichnen der Kalibrierungsgerade aufwiesen. Bereits bei den Hilfen, die als Teil von HyperDocs angesehen werden können, zeigt sich beispielsweise eine Abnahme der Bewertung der Nützlichkeit. Gleichzeitig steigt sowohl in der Mittel- als auch in der Oberstufe der Cognitive Load über die Unterrichtsstunden an. Die negative Korrelation zwischen Usability und Cognitive Load konnte aufgezeigt werden und sie zeigte sich auch in anderen Studien (Koć-Januchta et al., 2022). Dadurch können vermutlich mehr Schüler:innen, aufgrund des erhöhten Cognitive Load, Aufgaben nicht vollständig bearbeiten, wodurch eine Abnahme der Zufriedenheit einhergeht und das System für die Bearbeitung von Aufgaben als weniger nützlich betrachtet wird. Zusätzlich sinkt mit der höheren kognitiven Last wahrscheinlich die subjektive Qualität des Lerninhaltes, wodurch insbesondere die Nützlichkeit sinken könnte (Khedr et al., 2011). Unterstützt wird diese Hypothese durch die bereits diskutierte Abnahme der intrinsischen Motivation und dem vorliegenden Zusammenhang der intrinsischen Motivation zu allen Variablen der Usability. Der Einfluss verschiedener Variablen

auf die Usability zeigt sich ebenfalls unter Berücksichtigung der signifikanten Abnahme der Variablen „Einfachheit der Nutzung“ und „Einfachheit des Lernens der Nutzung“. Bereits Hassenzahl et al. (2010) konnte den Zusammenhang zwischen einem positiven Erlebnis (Bedürfnisbefriedigung, unter anderem im Sinne der Selbstbestimmungstheorie) und der Usability zeigen. Wie bereits oben beschrieben, sollten die Nutzer:innen mit zunehmender Einsatzdauer mehr Kompetenzen im Umgang mit einer Anwendung erlangen. Dadurch sollte die Anwendung zunehmend einfacher zu nutzen sein und die verschiedenen Funktionen sollten ebenso schneller über die Zeit erlernt werden. Die beiden Variablen sollten daher mit zunehmender Einsatzdauer steigen. Da jedoch eine Abnahme zu beobachten ist, liegt die Vermutung nahe, dass die Bewertung der Usability von anderen Variablen beeinflusst wird. Dabei ist zusätzlich anzumerken, dass der gesamte Funktionsumfang von HyperDocSystems ab der ersten Stunde Anwendung fand und keine weiteren Funktionen in den folgenden Stunden ergänzt wurden. Gleichzeitig lässt sich eine Zunahme der tabletbezogenen Kompetenzen, zumindest bei den Gymnasiast:innen, feststellen. Damit kann eine Abnahme der Selbstwirksamkeit nicht der Grund für eine geringere Bewertung der Usability sein. Generell steigt jedoch nach Angaben der Schüler:innen das Ablenkungspotential der Tablets, was auch zu einer Verminderung der Nützlichkeit und Zufriedenheit führen könnte.

Zusammenfassend zeigt sich eine Abnahme aller Variablen der Usability über den Einsatz von vier Unterrichtsstunden. Die Abnahme der Benutzerfreundlichkeit lässt sich in der Mittel- und Oberstufe beobachten. Eine mögliche Erklärung liegt im Recency-Effekt in Zusammenhang mit der Interaktion anderer Variablen, wie beispielsweise dem Cognitive Load, der Motivation oder auch den Lernhilfen.

Ausgehend von den Forschungsfragen und dem aktuellen Stand der Literatur haben sich weitere Detailfragen ergeben, die nun auch beantwortet werden sollen. Bei allen drei Fragen handelt es sich um Fragen, die einen Zusammenhang zwischen der Usability und weiteren damit verbundenen Variablen untersuchen sollen.

Die erste Frage lautet:

*D\_U1: Welcher Zusammenhang besteht zwischen der Usability und der kognitiven Last beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe*

Sowohl in der Mittel- als auch in der Oberstufe zeigt sich ein signifikanter Zusammenhang zwischen der Usability und dem Cognitive Load. In der Mittelstufe besteht dabei

eine negative Korrelation zwischen dem Intrinsic und Extraneous Load und den Variablen der Usability, wobei der Zusammenhang zu der Einfachheit der Nutzung und der Einfachheit des Lernens stärker ist als zu den beiden anderen Variablen. Zum Germane Load besteht erwartungsgemäß ein positiver Zusammenhang, wobei auch hier zur Einfachheit der Nutzung und der Einfachheit des Lernens eine stärkere Korrelation vorliegt. Zusammenhänge zum Mental Effort sind, mit einer Ausnahme, nicht feststellbar. In der vierten Stunde nimmt die Stärke der Zusammenhänge beim Intrinsic und Extraneous Load tendenziell ab und das Signifikanzniveau wird nicht mehr erreicht. In der Oberstufe bestehen die gleichen Tendenzen wie in der Mittelstufe, wobei vor allem in der ersten Stunde weniger signifikante Korrelationen zwischen den Variablen der Usability und des Cognitive Load auftreten.

Die gefundenen negativen Zusammenhänge zwischen der Usability und dem Cognitive Load sind erwartungsgemäß und wurden bereits in der Literatur beschrieben (Ibili & Billinghamurst, 2019; Koć-Januchta et al., 2022). Die vorliegende Arbeit kann damit die bisherige Forschung stützen und insbesondere auch für den Einsatz von digitalen Arbeitsblättern verifizieren. Die höheren negativen Korrelationen bei den Variablen „Einfachheit der Nutzung“ und „Einfachheit des Lernens“ könnten darin begründet sein, dass eine höhere Last durch die Darstellung der Inhalte (Extraneous Load) die Interaktion mit den HyperDocs erschwert und zu einer Überladung des Arbeitsgedächtnisses führt, wodurch die Nutzung als weniger leicht und erlernbar wahrgenommen wird. Gleiches gilt für den Intrinsic Load, wobei hier kein direkter Bezug zu der Darstellung der Inhalte besteht, was sich jedoch auch in den geringeren Korrelationen widerspiegelt, die insbesondere in der vierten Stunde bei zwei der vier Variablen der Usability kein signifikantes Niveau mehr erreichen. Gleiches gilt für den Mental Effort, der mit der Interaktion der HyperDocs in keinem Zusammenhang steht. Da keine Aussage über die Kausalität der Zusammenhänge getroffen werden kann, ist auch eine Begründung ausgehend von einer schlechteren Usability auszuführen. So nehmen womöglich Schüler:innen, die weniger geübt im Umgang mit Tablets sind, die Nutzung der HyperDocs als weniger einfach wahr und lernen den Umgang damit langsamer. Dadurch könnte insbesondere der Extraneous Load der Lernenden überproportional steigen, was sich über beide Betrachtungspunkte zeigt. Genauso führt gegebenenfalls eine bessere Beurteilung der Usability durch zum Beispiel erfahrenere Nutzer:innen zu einem höheren Germane Load, da mehr Kapazität des Arbeitsgedächtnisses besteht, um die Inhalte zu verarbeiten. Die Abnahme der Stärke der Zusammenhänge in der vierten Stunde lässt sich nicht eindeutig begründen. Es ist

denkbar, dass ein stärkerer Zusammenhang mit anderen Variablen besteht, wie zum Beispiel dem Interesse beim Bearbeiten der Arbeitsblätter. Bei Betrachtung der Korrelationen zwischen der Usability und dem Interesse zeigen sich in der vierten Stunde stärkere Zusammenhänge, wodurch womöglich das Interesse einen stärkeren Einfluss als der Cognitive Load nimmt. Eine weitere Begründung in der Abnahme der Zusammenhänge könnte in der Selbstregulation der Lernenden liegen. Zunehmend wird ein Bezug zwischen dem (Extraneous) Cognitive Load und der Selbstregulation hergestellt (De Bruin et al., 2020; Eitel et al., 2020). Nach der Theorie würde bei längerer Einsatzdauer der HyperDocs die Selbstregulation sinken und damit Faktoren, die den Extraneous Load erhöhen würden, mehr Einfluss gewinnen. Diese Tendenz zeigt sich auch in den Daten. Schüler:innen geben sowohl in der Mittel- als auch in der Oberstufe zunehmend an, durch die Nutzung der Tablets abgelenkt zu werden. Dabei besteht dieses Ablenkungspotential jedoch bereits ab der ersten Stunde. Aufgrund dieser Beobachtung könnte vermutet werden, dass die Lernenden an Selbstregulation verlieren und sich damit die Bewertung des Extraneous Load von der Usability abkoppelt. Dadurch könnte auch die Zunahme dieser Art der kognitiven Last über die Unterrichtsstunden begründet werden.

Die zweite Detailfrage, die sich aus dem Technology Acceptance Model ergibt, ist folgende:

*D\_U2: Welcher Zusammenhang besteht zwischen der Usability und den tabletbezogenen Kompetenzen beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

Aufgrund der schlechten Gütekriterien wurde die Angst vor Tablets nicht in die Auswertung miteinbezogen. Erwartungsgemäß zeigen sich in der Mittel- und Oberstufe signifikante Korrelationen zwischen der Usability und der tabletbezogenen Selbstwirksamkeit beziehungsweise der Nützlichkeit von Tablets beim Lernen. Demnach kann dieser Aspekt des Technology Acceptance Model (Davis, 1989; Venkatesh, 2000) auch für den schulischen Kontext verifiziert werden. In der Mittelstufe liegen zwischen allen genannten Variablen signifikante Zusammenhänge vor. Wie bereits bei den anderen Zusammenhangsanalysen, nimmt die Stärke der Zusammenhänge in der vierten Stunde zu. Ein möglicher Grund für diese Zunahme ist die gegenseitige Beeinflussung dieser Variablen über den Unterrichtszeitraum. Denkbar ist beispielsweise, dass Schüler:innen mit einer anfangs geringeren Selbstwirksamkeit durch den Umgang mit den Tablets eine höhere Selbstwirksamkeit erfahren und damit zunehmend besser mit HyperDocs umgehen

können, womit letztlich auch eine bessere Bewertung der Usability erfolgt. Diese Wirkungsrichtung, in der die Selbstwirksamkeit bei der Computer- beziehungsweise hier Tabletnutzung auf die Einfachheit der Nutzung und damit auf die Nützlichkeit Einfluss nimmt, ist bereits hinreichend belegt (Venkatesh, 2000). Diese Argumentation gilt allerdings nur für die Lernenden mit einer geringen Selbstwirksamkeit, da insgesamt die Bewertung der Usability abnimmt, wie die Diskussion zu Forschungsfrage FU\_3 bereits zeigte. Auffällig ist weiterhin die stärkere Korrelation der Subskala „Tablets beim Lernen“ zu den Variablen der Usability im Vergleich zu der Selbstwirksamkeit selbst. Der stärkere Zusammenhang lässt sich sinnvoll begründen. Lernende, die Tablets zum Lernen als nützlich erachten, bewerten digitale Arbeitsblätter wahrscheinlich ebenso als nützlich und sind zufriedener mit einer solchen Arbeitsform. Gleichzeitig werden solche Schüler:innen bereits Erfahrungen im Umgang mit Tablets gemacht haben und erachten die Nutzung und das Erlernen der Nutzung von HyperDocs als einfacher. Genauso ist denkbar, dass sich Schüler:innen durch die hohe Usability der HyperDocs mit zunehmender Einsatzdauer selbstwirksamer fühlen und auch Tablets als sinnvoller zum Lernen erachten, wodurch sich die Zunahme der Stärke der Korrelationen in der vierten Stunde begründen lässt.

In der Oberstufe nimmt die Stärke der Korrelationen lediglich bei der Subskala „Tablets beim Lernen“ eindeutig zu. Bei der tabletbezogenen Selbstwirksamkeit kommt es tendenziell zu einer leichten Abnahme. Für die Zunahme der Stärke der Zusammenhänge der eben beschriebenen Variable lässt sich der gleiche Erklärungsansatz wie in der Mittelstufe heranziehen, wobei über die Wirkungsrichtung der Korrelationen erneut keine eindeutige Aussage getroffen werden kann. Gegebenenfalls sehen die Schüler:innen aufgrund der hohen Usability der HyperDocs generell Tablets als sinnvoller an zum Lernen. Andererseits kann auch eine höhere Bewertung der Variablen zu einer verbesserten Usability führen, da die Schüler:innen die Nützlichkeit über den Erhebungszeitraum kennengelernt haben.

Die letzte Detailfrage, die wie folgt formuliert wurde, lässt sich eindeutig beantworten:

*D\_U3: Welcher Zusammenhang besteht zwischen der Usability und der Nutzungshäufigkeit von digitalen Endgeräten beim Lernen mit HyperDocs?*

Die Analysen zeigen keinen Zusammenhang zwischen der Usability und der Nutzungshäufigkeit. Es wäre zu erwarten, dass Lernende, die häufiger ein Tablet benutzen, HyperDocs einfacher nutzen können und die Nutzung einfacher erlernen, da sie eine höhere

Selbstwirksamkeit aufweisen. Das war sowohl in der Mittel- als auch in der Oberstufe nicht der Fall. In der Oberstufe besteht gar ein negativer Zusammenhang zwischen der schulischen Nutzungshäufigkeit von Tablets und der Einfachheit des Lernens der Nutzung. Das Ergebnis ist allerdings mit  $p = 0.070$  nicht signifikant, wobei die Stärke des Zusammenhangs mit  $\rho = -0.40$  hoch ist. Insgesamt ist die Stichprobe in der Oberstufe mit  $n = 21$  jedoch gering. Ein möglicher Erklärungsansatz für die fehlenden Korrelationen könnte in der Art der Anwendung liegen, da Schüler:innen eine Anwendung wie HyperDocSystems mit hoher Wahrscheinlichkeit zuvor noch nicht genutzt haben und dadurch selbst bei einer hochfrequenten Tabletnutzung die Nutzung von HyperDocSystems nicht leichter fällt.

### 6.1.3 Lernhilfen und Vertiefungsaufgaben

Ziel dieser Arbeit war es unter anderem, Forschungslücken im Themenkomplex der Binnendifferenzierung mittels Lernhilfen, insbesondere in einem digitalen Kontext, zu beantworten. Dabei ist festzuhalten, dass im Bereich der digitalen Differenzierung der Forschungsstand gering ist und daher viele der Forschungsfragen einen explorativen Charakter aufweisen. Im Folgenden werden nun die einzelnen Fragen beantwortet und weitere Variablen, die im Zusammenhang mit der Nutzung der Lernhilfen stehen, in die Diskussion miteinbezogen.

Folgende Forschungsfrage wurde für diese Arbeit formuliert:

*F\_L1: Inwieweit bestehen Unterschiede zwischen den Geschlechtern in der Nutzungshäufigkeit von Lernhilfen über den Zeitraum einer vierstündigen Unterrichtsreihe?*

In der Mittelstufe wurde aufgrund der deskriptiven Unterschiede zwischen der Gesamtschule und dem Gymnasium bei den Variablen der Usability und der Motivation auch die Nutzung der Lernhilfen und Vertiefungsaufgaben unter Berücksichtigung der Schulform ausgewertet. Hierbei zeigen sich signifikante Interaktionen zwischen dem Geschlecht und dem Messzeitpunkt sowie der Schulform und dem Messzeitpunkt. Dabei liegt ebenfalls ein signifikanter Unterschied zwischen den Geschlechtern in zwei der vier Unterrichtsstunden vor, wonach in der zweiten und vierten Stunde Schülerinnen häufiger eine Hilfe verwenden. Die Effektstärken sind dabei als mittelstark einzustufen ( $\xi = 0.20 - 0.24$ ). Bei Betrachtung des Interesses (SDT) zeigen die Geschlechter in der Interventionsgruppe keine signifikanten Unterschiede. Bei den Schüler:innen der Gesamtschule liegt das



Interesse unter dem der Gymnasiast:innen, wodurch möglicherweise diese Teilgruppe die Lernhilfen aus Langeweile genutzt hat. Bei Analyse des Kompetenzerlebens und des Drucks zeigen Schülerinnen insbesondere bei der dritten und vierten Stunde (Kompetenzerleben) und ab der zweiten Stunde (Druck) ein geringes Kompetenzerleben und einen erhöhten Druck, wodurch die Nutzung von Lernhilfen steigen könnte. Zwischen den Variablen konnten jedoch keine signifikanten Korrelationen gefunden werden (vgl. Forschungsfrage F\_L2). Wird die Usability als möglicher Erklärungsansatz herangezogen, zeigt sich insbesondere bei den Schülerinnen eine Abnahme der Nützlichkeit über die Unterrichtsreihe. Gegebenenfalls verwenden die Schülerinnen daher in der Summe mehr verschiedene Hilfen, da sie die einzelnen Hilfen als weniger nützlich erachten. Dagegen spricht allerdings, dass in der dritten Stunde von den weiblichen Lernenden tendenziell weniger Lernhilfen genutzt werden. Ein begründeter Erklärungsansatz liegt im Cognitive Load. Insbesondere der Intrinsic Load sollte zu einer höheren Nutzung von Lernhilfen führen. Während sich keine Korrelationen zwischen der Nutzung und dem Intrinsic Load zeigen, ist jedoch ein höherer Intrinsic Load bei den Schülerinnen in der vierten Stunde erkennbar. Auch der Mental Effort liegt bei den Schülerinnen ab der zweiten Stunde höher. Dabei zeigt sich in der vierten Stunde auch eine signifikante schwache Korrelation zwischen dem Mental Effort und der Nutzung der Lernhilfen. Gleichzeitig besteht in der ersten Stunde eine schwache negative signifikante Korrelation zum Germane Load, was sich mit dem theoretischen Modell des Cognitive Load gut vereinbaren lässt. Insgesamt ist jedoch anzumerken, dass die Nutzung der Lernhilfen in den ersten beiden Unterrichtsstunden bei beiden Geschlechtern am höchsten ist und danach signifikant abnimmt und schließlich in der dritten Stunde einen Tiefpunkt erreicht. Gleichzeitig ist der Cognitive Load in den ersten beiden Stunden am niedrigsten. Womöglich liegt in den ersten Stunden ein größerer Einfluss des Interesses und der Neuheit vor, wonach die Lernenden zunächst alle Hilfen erkunden. Das zeigt sich auch in den Daten der Gründe für die Nutzung von Lernhilfen (vgl. Detailfrage F\_L4). Neben den eben genannten Interaktionen, wurde bereits in der Literatur beschrieben, dass Schülerinnen eine höhere Intention haben, Hilfen zu nutzen, ein geringeres Vermeidungsverhalten in Bezug auf die Hilfenutzung zeigen und die Kosten dieser Nutzung als geringer wahrnehmen (Cheong et al., 2004; Kessel und Steinmayr, 2013; Ryan et al., 2009). Dadurch kann die verstärkte Hilfenutzung zusätzlich erklärt werden. Dabei darf jedoch nicht die Art der Hilfenutzung (digital / analog) außer Acht gelassen werden. Unterschiede im Darbietungsformat konnten bereits gezeigt

werden (Beal et al., 2008; Karabenick & Knapp, 1988). Final können daher die genauen Gründe für die erhöhte Hilfenutzung nicht vollständig erklärt werden.

Zwischen den Schulformen liegen bezüglich der Nutzung der Lernhilfen zu den einzelnen Messzeitpunkten keine signifikanten Unterschiede vor, wobei die Gesamtschüler:innen insbesondere in der zweiten und dritten Stunde mehr Hilfen verwenden. Das Kompetenzerleben und der Druck sind zwischen den beiden Schulformen gleich. Ein möglicher Erklärungsansatz liegt darin, dass an der Gesamtschule der Leistungsstand niedriger und die Heterogenität höher sind, wonach auch mehr Lernhilfen verwendet werden. Zudem liegen der Intrinsic und Extraneous Load in der Gesamtschule höher, während der Germane Load insbesondere bei den Schülerinnen geringer ist. Daraus lässt sich schließen, dass der Lerngegenstand für die Schüler:innen der Gesamtschule schwieriger war. Das zeigt auch das verringerte Kompetenzerleben.

In der Oberstufe zeigen sich Unterschiede zwischen den Geschlechtern in der dritten und vierten Stunde. In diesen Stunden nutzen die Schülerinnen mehr Hilfen, wobei die Unterschiede zu den Schülern nicht signifikant sind. Prinzipiell können dafür die gleichen Erklärungsansätze wie in der Mittelstufe herangezogen werden, wobei sich in den Daten weniger Unterschiede bei den anderen Variablen zeigen. In Bezug auf die Motivation liegen beispielsweise nur beim Druck in der vierten Stunde leichte Unterschiede vor, wonach Schülerinnen mehr Druck verspüren. Auch zeigen weibliche Lernende eine stärkere Abnahme bei der Bewertung der Nützlichkeit der Usability. Die Unterschiede sind insgesamt aber gering. Beim Cognitive Load liegen die Werte der Schülerinnen insbesondere beim Intrinsic Load und dem Mental Effort über denen der Schüler. Darin begründet könnte demnach die umfangreichere Nutzung der Lernhilfen sein, wobei sich die Unterschiede, wie auch in der Mittelstufe, eher am Ende der Unterrichtsreihe zeigen. Eine mögliche Ursache könnte darin liegen, dass der Cognitive Load erst am Ende der Unterrichtsreihe an Einfluss gewinnt. Dafür spricht, dass die Korrelationen zwischen der Nutzung der Lernhilfen und dem Intrinsic Load und Extraneous Load eine schwache bis mittlere Stärke in der vierten Stunde zeigen. In der ersten Stunde ist dieser Zusammenhang negativ.

An dieser Stelle soll kurz auf die Vergleichsgruppe eingegangen werden, die aufgrund der Methodik der Datenerhebung der Hilfenutzung nicht gemeinsam mit der Interventionsgruppe ausgewertet werden konnte. Insgesamt zeigt die Nutzung der Lernhilfen über die Zeit sowohl in der Mittel- als auch in der Oberstufe den gleichen Verlauf wie in der

Interventionsgruppe. Hierbei sind in der Mittelstufe in den ersten drei Stunden ebenfalls Unterschiede zwischen den Geschlechtern vorhanden, wonach Schülerinnen häufiger eine Hilfe verwenden als Schüler, was durch einen signifikanten Haupteffekt bestätigt wird. Bei Betrachtung der weiteren Variablen kann ein geringeres Interesse, ein verringertes Kompetenzerleben, in Teilen ein erhöhter Intrinsic Load, ein niedrigeren Germane Load und ein leicht erhöhter Mental Effort bei den weiblichen Lernenden festgestellt werden. Es ist jedoch anzumerken, dass die Unterschiede in der Nutzung der Hilfen in der vierten Stunde nicht mehr beobachtbar sind, obwohl weiterhin Unterschiede in den eben genannten Variablen auftreten. So erreicht beispielsweise das Kompetenzerleben bei den Schülerinnen in der vierten Stunde ein Minimum, während der Intrinsic Load ein Maximum erreicht. Vermutlich liegen daher doch Gründe vor, die die geschlechterspezifischen Rollen betreffen, die bereits durch verschiedene Autor:innen gezeigt werden konnten (Cheong et al., 2004; Kessel und Steinmayr, 2013; Ryan et al., 2009).

In der Oberstufe nutzen die Schülerinnen hingegen in der ersten und dritten Stunde weniger Hilfen als die Schüler. Unter Berücksichtigung der verschiedenen Variablen können die gleichen Wirkmechanismen, die bereits weiter oben diskutiert wurden, angenommen werden. Dabei muss allerdings berücksichtigt werden, dass die Unterschiede zwischen den Geschlechtern nicht immer gegeben sind. Beispielsweise erreicht das Kompetenzerleben bei den weiblichen Lernenden in der dritten Stunde ein deutliches Minimum. Es liegt damit unter dem der Jungen. Für die weiteren Variablen zeigen sich insbesondere in der ersten Stunde Unterschiede zugunsten der Schülerinnen. Sie zeigen ein höheres Interesse, ein höheres Kompetenzerleben, einen niedrigeren Intrinsic Load, einen erhöhten Germane Load und einen niedrigeren Mental Effort.

Zusammenfassend lassen sich für die erste Forschungsfrage gemischte Ergebnisse festhalten. Während sowohl in der Interventions- als auch in der Vergleichsgruppe der Mittelstufe die Schülerinnen häufiger eine Hilfe verwenden und die Hilfenutzung in beiden Gruppen einen ähnlichen Verlauf nimmt, nutzen in der Oberstufe die Schüler häufiger eine Hilfe. Damit kann keine Generalisierung formuliert werden, wonach Mädchen häufiger eine Lernhilfe verwenden als Jungen. Vielmehr finden sich in den Daten Tendenzen, wonach sich die unterschiedliche Nutzung auf das Interesse, das Kompetenzerleben, den Intrinsic und Germane Load und den Mental Effort zurückführen lassen. Die Bezüge zu diesen Variablen können theoretisch begründet werden, auch wenn sich nur wenige Korrelationen zwischen den Variablen und der Hilfenutzung zeigen. Alle diese Variablen

haben womöglich ihren Ursprung in einer geringeren Selbstwirksamkeit der Schüler:innen im Fach Chemie.

### *Vertiefungsaufgaben*

Zusätzlich wurde auch die Nutzung der Vertiefungsaufgaben analysiert, wobei hierfür keine eigene Forschungsfrage formuliert wurde. Die Nutzungshäufigkeit lässt sich nicht valide ohne die tatsächliche Überprüfung der Bearbeitung der Vertiefungsaufgaben auswerten. Dennoch wurde aus eigenem Interesse die Nutzung grafisch aufgetragen. Dabei zeigten sich stundenabhängige Unterschiede zwischen den Geschlechtern. Im Gymnasium rufen die Schüler in den ersten beiden Stunden häufiger eine Vertiefungsaufgabe auf, an der Gesamtschule sind es in der ersten Stunde die Schülerinnen. Danach ist die Nutzung zwischen den Geschlechtern ausgeglichen. Unter Vernachlässigung der Schulform zeigen sich keine nennenswerten Unterschiede. Insgesamt nimmt die Betrachtung der Vertiefungsaufgaben über die vier Unterrichtsstunden deutlich ab. Das könnte darin begründet sein, dass die Lernenden zunehmend weniger Zeit und weniger Interesse daran zeigen. Die Unterschiede zwischen den Geschlechtern sind nicht eindeutig zu erklären. Mögliche Zusammenhänge zum Interesse beim Lernen mit HyperDocs könnten die Unterschiede zumindest bei den Gymnasiast:innen begründen. Die anderen Variablen können die Unterschiede nicht eindeutig erklären.

Die nächste Forschungsfrage bezieht sich auf die vermuteten Zusammenhänge zum Konstrukt der intrinsischen Motivation:

*F\_L2: Welcher Zusammenhang besteht zwischen der Nutzungshäufigkeit der Lernhilfen durch Schüler:innen und der intrinsischen Motivation beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

Die Daten zeigen in Bezug auf diese Forschungsfrage bei Betrachtung der Lernhilfen in der Mittelstufe keine signifikanten Korrelationen. In der Oberstufe besteht eine signifikante Korrelation zwischen der Hilfenutzung und dem Druck, wonach Lernende mit einem geringeren Druck mehr Lernhilfen verwenden. Die fehlenden Korrelationen zwischen den Variablen können womöglich auf die Komplexität der Hilfenutzung zurückgeführt werden. Das zeigt sich beispielsweise in der Verschiebung der Gründe für die Nutzung der Lernhilfen (vgl. Detailfrage F\_L4). Auch wenn keine direkte Korrelation zwischen den Variablen festgestellt wurde, zeigen sich dennoch tendenzielle Zusammenhänge bei Betrachtung des Geschlechts, wobei der Cognitive Load und der Mental Effort die Unterschiede besser erklären als die Motivation (vgl. Detailfrage D\_L1). Die

signifikante Korrelation in der Oberstufe zeigt sich lediglich in der ersten Stunde. Erklärbar ist sie womöglich dadurch, dass Schüler:innen mit einem geringeren Druck mehr Hilfen aufrufen und sich mehr Zeit lassen, um die Aufgaben zu bearbeiten. Aus der Anzahl der aufgerufenen Hilfen lässt sich daher nicht unmittelbar auf die Motivation der Schüler:innen schließen. In einer anderen Studie konnte hingegen gezeigt werden, dass eine hohe Interaktion mit dem Lernmaterial auf ein höheres Interesse schließen lässt (Urhahne et al., 2004). Andere Studien zeigen ebenfalls einen positiven Zusammenhang zwischen der Nutzung von Lernhilfen und motivationalen Variablen, wie beispielsweise dem Kompetenzerleben (Algharaibeh, 2020; Newman, 1990; Yang & Cao, 2013). In der eigenen Studie kann die Hilfenutzung bei HyperDocs nicht allein auf die motivationale Komponente zurückgeführt werden. Die Hilfenutzung bildet ein komplexeres Konstrukt aus lern- und sozialpsychologischen Faktoren, welche im Rahmen dieser Arbeit nicht alle erfasst wurden. Einen Einfluss bildet sicherlich der Anforderungsbereich, da bei höheren Anforderungsbereichen vermutlich mehr Lernhilfen genutzt werden. Das zeigt sich insbesondere in der dritten Stunde bei den Oberstufenschüler:innen.

In Bezug auf die nächste Forschungsfrage zeigen sich keine signifikanten Ergebnisse:

*F\_L3: Welcher Zusammenhang besteht zwischen der Nutzungshäufigkeit der Lernhilfen und dem Leistungsstand der Schüler:innen beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

In der Mittelstufe sind die Zusammenhänge eindeutig nicht signifikant. Es wurde vermutet, dass leistungsschwächere Schüler:innen häufiger eine Hilfe nutzen (Ryan et al., 1997; Ryan & Shin, 2011). Der Leistungsstand wurde allerdings über eine Selbsteinschätzung erhoben, was womöglich das Ergebnis verzerrt. Die Neuheit des Mediums sorgt gegebenenfalls für eine zusätzliche Verzerrung der Hilfenutzung, da viele Lernende die Hilfen zu Beginn aus Interesse und Neugier aufrufen (vgl. Detailfrage F\_L4). In der Oberstufe sind ebenfalls keine signifikanten Korrelationen vorhanden. In der ersten Stunde gibt es einen schwachen positiven Zusammenhang zwischen der Nutzung der Lernhilfen und dem Leistungsstand in Chemie, der auf dem Niveau  $p < 0.1$  signifikant ist. Demnach würden leistungsstärkere Schüler:innen häufiger eine Hilfe verwenden. Dieses Ergebnis ist in der Literatur teilweise gesichert (Ryan et al., 1997; Ryan & Shin, 2011). Andere Studien zeigen den gegenteiligen Zusammenhang (Alevin et al., 2006; Newman & Schwager, 1995). In der vierten Stunde zeigt sich dieser Zusammenhang jedoch nicht mehr, was auf zusätzliche Einflussfaktoren, die zum Teil oben bereits

beschrieben wurden, schließen lässt und womöglich auch die widersprüchlichen Ergebnisse in der Literatur erklärt.

Die letzte Forschungsfrage untersucht die Nutzung der Lernhilfen und Vertiefungsaufgaben detaillierter:

*F\_L4: Aus welchen Gründen nutzen Schüler:innen Lernhilfen und Vertiefungsaufgaben beim Lernen mit HyperDocs während einer vierstündigen Unterrichtsreihe?*

Im Vergleich zwischen der Interventions- und Vergleichsgruppe der Mittelstufe nutzen beide Gruppen weitestgehend aus den gleichen Gründen die Hilfen, wobei der Mittelwert der Zustimmung teilweise variiert. Auffallend ist die vermehrte Nutzung der Lernhilfen aus Interesse und aus Neugier in der Interventionsgruppe gegenüber der Vergleichsgruppe. In beiden Gruppen nimmt zwar die Zustimmung über die Zeit ab, allerdings verbleibt sie bei den Tabletutzer:innen auf einem höheren Niveau. Durch die digitalen Hilfen könnten gegebenenfalls Lernende zu einer intensiveren Auseinandersetzung mit dem Lerngegenstand angeregt werden. Bei den weiteren Nutzungsgründen zeigen sich keine auffälligen Unterschiede. Es muss allerdings berücksichtigt werden, dass zwischen dem Interesse und der Hilfenutzung keine signifikante Korrelation gefunden werden konnte. Die Abnahme des Interesses und der Neugier an den Hilfen kann durch den Neuheitseffekt begründet werden, der sowohl bei der digitalen als auch bei der analogen Durchführung auftritt. Gleichzeitig nimmt in beiden Gruppen die Nutzung aus dem Hilfebedarf zu, wodurch die Lernenden ihr Interesse weniger ausbilden können und die Hilfen zunehmend zielführender zur Bearbeitung der Aufgaben nutzen. Gegebenenfalls können die Schüler:innen mit der Zeit ihre metakognitiven Strategien in Bezug auf die Hilfenutzung ausbauen und damit auch zielgerichteter verwenden. Zu diesem Ergebnis kamen auch Großmann und Woest (2014, 2015). Die hohe Standardabweichung und die teilweise niedrige Zustimmung in der Vergleichsgruppe zeigen, dass die Lernenden die Hilfen noch aus anderen Gründen verwenden.

In der Oberstufe finden sich die gleichen Verteilungsmuster wieder, wobei die Lernenden der Vergleichsgruppe die Hilfen gegenüber der Mittelstufe eher aus Neugier und Interesse aufrufen. Das Niveau liegt allerdings immer noch unter dem der Interventionsgruppe. Gleichzeitig ist bei den analog Lernenden die Abnahme über die vier Messzeitpunkte größer. Somit kann bei den digitalen Nutzer:innen das Interesse an den Hilfen länger aufrechterhalten werden. In der Oberstufe nutzen die Lernenden die Hilfen außerdem häufiger aus dem tatsächlichen Hilfebedürfnis heraus. Ein Grund hierfür könnte in einer

höheren Selbststeuerung liegen, sodass die Schüler:innen eher erkennen, wann sie eine Lernunterstützung verwenden sollten.

Die Vertiefungsaufgaben werden aus ähnlichen Gründen aufgerufen. In der Interventionsgruppe der Mittelstufe und Oberstufe nimmt die Nutzung aus Neugier und Interesse mit den Stunden zunehmend ab. Diese Abnahme kann über das generell abnehmende Interesse beim Lernen mit HyperDocs begründet werden. Gleichzeitig nimmt die Nutzung aus Langeweile zu. Die Schüler:innen stimmen diesem Grund jedoch am geringsten zu. Die Zunahme kann über den zuvor genannten Grund erklärt werden. Ebenso rufen die Lernenden die Aufgaben zunehmend weniger auf, um ihr Fachwissen zu erweitern oder Zeit zu überbrücken. Die Abnahme dieser Gründe kann durch die als geringer eingeschätzte Bearbeitungszeit erklärt werden, wonach die Lerngruppen keine Zeit mehr hatten, um die Aufgaben zu betrachten und zu bearbeiten. Unter Betrachtung der Stärke der Zustimmung rufen die Schüler:innen insgesamt die Vertiefungsaufgabe aus Interesse und als zeitlichen Puffer auf. In der Vergleichsgruppe zeigt sich ein ähnliches Bild. Die Nutzung aus Langeweile oder um das Fachwissen zu verbessern, bleibt über die Zeit stabil. Die Betrachtung als zeitlicher Puffer nimmt jedoch stark ab, was sich ebenfalls durch die geringere Bearbeitungszeit erklären lässt.

Im Folgenden werden nun die Detailfragen diskutiert. Die erste Detailfrage bezieht sich auf den Zusammenhang zwischen dem Cognitive Load und der Nutzung der Lernhilfen:

*D\_L1: Welcher Zusammenhang besteht zwischen der Nutzung von Lernhilfen und der kognitiven Last beim Lernen mit HyperDocs über den Zeitraum einer vierstündigen Unterrichtsreihe?*

Die Zusammenhangsanalysen fallen erneut gemischt aus. In der ersten Stunde besteht in der Mittelstufe lediglich eine signifikante schwach negative Korrelation zum Germane Load, die nach der Theorie des Cognitive Load nachvollziehbar erscheint. Eine mögliche Erklärung geht von der Wirkungsrichtung Germane Load zu Lernhilfen aus. Demnach nutzen Lernende mit einem niedrigeren Germane Load mehr Lernhilfen, da die übrige Kapazität des Arbeitsgedächtnisses durch den hohen Intrinsic und Extraneous Load stark limitiert ist. In den Daten findet sich jedoch keine signifikante Korrelation der Hilfenutzung zum Intrinsic oder Extraneous Load und in der vierten Stunde ist der Zusammenhang zum Germane Load ebenfalls nicht mehr signifikant, wodurch diese These nicht eindeutig belegt werden kann. In der vierten Stunde besteht eine schwache bis mittelstarke signifikante Korrelation zum Mental Effort. Auch hier erscheint die

Wirkungsrichtung Mental Effort zu Lernhilfen schlüssig, sodass Lernende, die sich kognitiv intensiver mit den Inhalten auseinandersetzen müssen, auch mehr Lernhilfen verwenden. Unter Berücksichtigung der Literatur nutzen Schüler:innen mit einer hohen kognitiven Last Hilfen, um einen direkten Lösungsweg zu erhalten (Dong et al., 2020). Letztlich muss auch in Erwägung gezogen werden, dass die Nutzung von Lernhilfen, aufgrund der parallelen Verarbeitung der Hilfen und des Lerninhaltes, zusätzlichen Cognitive Load verursacht (Alevén et al., 2003). Die ausbleibenden signifikanten Ergebnisse können allerdings auch positiv gesehen werden, wenn die Wirkungsrichtung Lernhilfen zu Cognitive Load betrachtet wird. Demnach erhöht der Einsatz der Lernhilfen den Cognitive Load nach den vorliegenden Daten nicht. In anderen Studien konnte eine Zunahme des Extraneous Cognitive Load beobachtet werden, wenn Lernhilfen mit multimedialer Unterstützung kombiniert wurden (Bruckermann et al., 2017).

Weshalb die genannten Korrelationen lediglich in einer der beiden Stunden (erste oder letzte) das Signifikanzniveau erreichen, beziehungsweise andere Variablen dieses nicht erreichen, kann nicht eindeutig gesagt werden. Möglicherweise liegen in den jeweiligen Stunden weitere Einflussfaktoren vor. Die bisherige Diskussion zu den anderen Forschungsfragen stützt diese Vermutung. Beispielsweise geben die Schüler:innen besonders am Anfang an, Lernhilfen aus Interesse und Neugier zu öffnen. Diese Hypothese wird bekräftigt, wenn die Daten der Oberstufe mitberücksichtigt werden. Hierbei liegt in der ersten Stunde eine signifikante negative Korrelation zum Intrinsic Load und Mental Effort vor. In der vierten Stunde hingegen werden diese beiden Korrelationen positiv, wobei die Effektstärke beim Mental Effort zusätzlich zunimmt. Über die Unterrichtsreihe hinweg kommt es somit zu einer völligen Umkehrung der Zusammenhänge, was für eine Veränderung der Nutzungsgründe spricht. Schüler:innen weisen teilweise geringere metakognitive Fähigkeiten auf, um Hilfen zu nutzen (Alevén & Koedinger, 2000; Vaessen et al., 2014). Es ist zu vermuten, dass die Lernenden im Verlauf der Unterrichtsreihe diese Strategien ausbauen und gezielter Hilfen nutzen konnten.

Die weitere Detailfrage hat einen explorativen Charakter und soll den Forschungsstand im Bereich der digitalen Differenzierung durch Lernhilfen erweitern:

*D\_L2: Inwieweit bevorzugen Schüler:innen eine bestimmte Darstellungsform der Hilfen bei der Nutzung von HyperDocs während einer vierstündigen Unterrichtsreihe?*



In der Mittel- und Oberstufe nutzen die Schüler:innen überwiegend Hilfen in Form eines Textes oder Bildes (zwischen 30 und 60 Prozent). In der ersten Stunde nutzen die Lernenden in der Mittelstufe (10 %) und Oberstufe (7 %) jedoch ebenfalls die Audiohilfe. Hierbei handelt es sich allerdings um die bereinigten Daten. Unter Berücksichtigung der Rohdaten, die keine Korrektur durch Unterschreitung der theoretischen Betrachtungsdauer erfahren, zeigt sich eine leichte Zunahme der Nutzung von Audio- und Videohilfen, wobei die Zunahme, insbesondere in der Oberstufe, gering ausfällt.

Diese Ergebnisse scheinen aus forschungsmethodischer und schulpraktischer Sichtweise plausibel. Durch die vorgenommene Auswertungsmethode werden Bildhilfen anteilig eher übergewichtet, da pro Bild eine theoretische Wahrnehmungszeit von einer Sekunde pro Bild angenommen wurde. Diese Wahrnehmung ist allerdings stark vom Inhalt abhängig, jedoch kann in der Literatur kein eindeutiger Wert identifiziert werden. In der Regel erfolgt die Objekterkennung innerhalb weniger Millisekunden. Auf der anderen Seite werden die Video- und Audiohilfen dadurch untergewichtet, da diese lediglich bei voller Spieldauer berücksichtigt wurden. Für die initiale Betrachtung liefert diese Methode aussagekräftige Ergebnisse. Bei erneuter Einsicht in die Hilfe wird wahrscheinlich nur ein Teil der Hilfe, der die wichtigen Informationen liefert, betrachtet. Aus schulpraktischer Sichtweise entsprechen die Ergebnisse den Erwartungen. Zwar nutzen Jugendliche in ihrer Freizeit Videos (Mpfs, 2022), allerdings ist diese Methode für den Unterricht, ebenso wie Sprachnachrichten, unpraktisch. Zunächst ist ein Kopfhörer erforderlich, um die akustischen Inhalte hören zu können. Dadurch sind die Schüler:innen von ihrer Umgebung stärker abgeschirmt, wodurch sie Anmerkungen der Mitschüler:innen oder Lehrkräfte unter Umständen nicht mehr wahrnehmen. Zusätzlich stehen der Nutzung organisatorische Gründe im Weg, wie beispielsweise das prinzipielle Vorhandensein von Kopfhörern. Für die Studie wurden zwar prinzipiell Kopfhörer bereitgestellt, allerdings mussten die Lernenden sich diese am Pult abholen. Ein weiterer praktischer Grund für die Bevorzugung der Bild- und Texthilfen liegt in der alltäglichen Nutzung des Schulbuches und Heftes als Lernmedium. Diese bestehen ausschließlich aus Texten und Bildern und stellen damit für Schüler:innen ein gewohntes Lernmedium dar. Aus lernpsychologischer Sicht spricht gegen die Verwendung von Videos und Audios eine mögliche Erhöhung des Extraneous Cognitive Load, da eine zusätzliche Interaktion, zum Beispiel Starten und Stoppen der Mediadatei, erforderlich ist. Weiterhin konnte in einer anderen Studie gezeigt werden, dass geschriebener Text in Hilfen in Bezug auf die Lernleistung gesprochenem Text überlegen ist (Sanchez & Garcia-Rodicio, 2008). Können Lernende, wie im Falle

dieser Arbeit, frei wählen, entscheiden sie sich womöglich intuitiv für die einfacher zu verarbeitende Hilfeform. Einen weiteren Einfluss nimmt gegebenenfalls der Stundeninhalt, wonach ein visueller Schwerpunkt in den Stunden auch zu einer höheren Nutzung der Bild- und Videohilfen führt. Für diese Vermutung finden sich Anhaltspunkte in den Daten. Schüler:innen nutzten in der Mittel- und Oberstufe vermehrt Bild- und Videohilfen in der Unterrichtsstunde, in die Messergebnisse der photometrischen Messung (Mittelstufe: Stunde 3, Oberstufe: Stunde 2) ausgewertet wurden. Diese Tendenz findet sich auch, unter Ausschluss der ersten Stunde, in der Vergleichsgruppe.

Bei weiterer Betrachtung der Vergleichsgruppe, bei der anhand einer Selbstangabe lediglich zwischen Text- und Bildhilfe (verbal und visuell) unterschieden wurde, zeigt sich eine leichte Tendenz zugunsten der visuellen Hilfe. Die Ergebnisse sind insbesondere für die erste Stunde eindeutig. Hierbei kann ein Zusammenhang zu dem Thema, der Behandlung von Gefahrensymbolen, vermutet werden. In der zweiten Stunde der Mittelstufe und der dritten Stunde der Oberstufe ist die Nutzung ausgeglichen. In diesen Stunden fand überwiegend Textarbeit statt, was die ausgeglichene Verwendung erklären könnte. Bei den anderen Stunden werden eher wieder visuelle Inhalte behandelt (Zeichnen eines Graphen und Teilchenmodells), was die Tendenz in Richtung visueller Hilfen erklären könnte. Die eben genannten Tendenzen zeigen sich in der Interventionsgruppe lediglich eindeutig in der ersten Stunde bei Betrachtung der bereinigten Hilfeaufrufe. In der zweiten und dritten Stunde der Mittel- und Oberstufe sind ähnliche Tendenzen wie in der Vergleichsgruppe zu erkennen. Die geringen Unterschiede in den beiden Untersuchungsgruppen sind womöglich auf die Methodik zur Erfassung der Verteilung zurückzuführen (Selbstangabe gegen maschinelle Auswertung).

## **6.2 Fachdidaktisches Fazit zur Unterrichtsreihe**

Die beiden Unterrichtsreihen für die Mittel- und Oberstufe erwiesen sich als durchführbar und sie entsprachen in vielen Teilen dem richtigen Anforderungsbereich, um sie in den jeweiligen Lerngruppen durchführen zu können. Die Entscheidung, auf eine erprobte Reihe zurückzugreifen und diese dem notwendigen Design der Studie anzupassen, erwies sich ebenfalls als zielführend. Die jeweiligen Unterrichtsstunden konnten in der vorgegebenen Zeit gehalten werden, ohne dass die Schüler:innen einen großen Zeitdruck verspürten. Die Schwerpunkte der jeweiligen Stunden bilden wichtige Grundlagen und Kompetenzen für den Chemieunterricht. Gleichzeitig konnte durch die weniger üblichen Themen

der Azofarbstoffe im gesundheitlichen Kontext und die Konzentrationsbestimmung mittels des Smartphotometers eine große Stichprobe gewonnen werden, die kein Vorwissen in Bezug auf den Anwendungsfall aufwies. Die Unterrichtsreihe erwies sich in der Vergleichsgruppe ebenfalls als motivierend. Dies zeigte sich unter anderem an der regen Unterrichtsbeteiligung der Schüler:innen. Die Motivation wurde teilweise durch die Progression in den Stunden etwas gemindert. Auch die Hilfen erfüllten in diesem Zusammenhang ihren didaktischen und methodischen Zweck, sodass die Lernenden selbstständig arbeiteten und Rückfragen selten auftraten. Teilweise wurden die Vertiefungsaufgaben von Schüler:innen bearbeitet und im Unterricht besprochen. Damit zeigt sich auch deren didaktisches Potential.

Aus methodischer Sicht eigneten sich die HyperDocs, unabhängig von der Studie, für den Einsatz im Unterricht. Die Schüler:innen konnten die digitalen Arbeitsblätter nahezu problemlos verwenden. In diesem Zusammenhang bewerteten die Lernenden die Einschränkungen durch die Verwendung der iPad, wie zum Beispiel die Ermüdung der Augen oder das Ablenkungspotential, als gering. Zudem zeigte sich das interessenfördernde Potenzial der HyperDocs.

In einzelnen Stunden zeigten sich jedoch auch didaktische Schwächen. Die erste Stunde besitzt in beiden Stufen einen zu niedrigen Anforderungsbereich. Die ursprüngliche Idee, die Stunde einfach zu gestalten, um eine problemlose Einführung in die Methodik der HyperDocs und Lernhilfen zu ermöglichen, konnte zwar umgesetzt werden, allerdings könnten fachliche Inhalte stärker in den Fokus gerückt werden. Beispielsweise könnte die Follow-Up-Stunde „Zucker, das weiße Gold“, die noch in den Pilotierungen gehalten wurde, auch als Einstiegsstunde dienen, indem zunächst die Zuckerthematik bei dem Sirup thematisiert wird. Diese wurde von einigen Lernenden auch in der zweiten Stunde bei Einführung des Sirups genannt. Auf der anderen Seite stellte die dritte Stunde in der Oberstufe die Lernenden vor eine größere Herausforderung, die sich insbesondere auch im hohen Cognitive Load und in dem geringen Interesse bemerkbar macht. Zwar konnte die Stunde im vorgegebenen Zeitrahmen gehalten werden, allerdings hätte für einen sicheren Umgang mit den Begrifflichkeiten und der Methodik der Photometrie den Lernenden wahrscheinlich mehr Zeit eingeräumt werden müssen.

In der Mittelstufe stellte das Zeichnen der Kalibriergeraden die digital Lernenden vor eine größere methodische Herausforderung. Hier zeigte sich, dass viele Schüler:innen noch keine umfangreichen Erfahrungen mit dem Apple Pencil aufwiesen. Das ist allerdings

eher eine Herausforderung der generellen Methode der HyperDocs als ein Problem in der fachdidaktischen Planung. Eventuell hätten die Schüler:innen durch kleinschrittigere Aufgaben langsamer an den Prozess herangeführt werden müssen. Auf der anderen Seite konnte der Arbeitsauftrag mit den Hilfen durch viele Lernende selbstständig erarbeitet werden. Die methodischen Schwächen der Studie werden im nachfolgenden Kapitel diskutiert.

### **6.3 Methodenkritik und Limitationen**

Im Laufe der Erhebung haben sich einige Schwächen des Untersuchungsdesigns und der eingesetzten Methoden gezeigt, die im Folgenden diskutiert werden sollen. Am Ende des Unterkapitels werden Änderungen beziehungsweise Empfehlungen für zukünftige Projekte gegeben, um noch zuverlässigere Daten zu erhalten. Zusätzlich wird auf die hier angenommenen Kompromisse, die in weiteren Forschungsprojekten vermieden werden sollten, und die daraus resultierenden Limitationen eingegangen. Ein allgemeiner Aspekt, der bei nahezu allen betrachteten Variablen auftritt, ist die hohe Standardabweichung, die zum Teil eine Einheit des Mittelwertes beträgt. Diese Beobachtung lässt vermuten, dass teilweise erhebliche individuelle Unterschiede zwischen den Lernenden vorliegen.

#### **6.3.1 Methodenkritik**

Zunächst lässt sich positiv hervorheben, dass die eingesetzten Methoden weitestgehend adäquat in Bezug auf die aufgestellten Forschungsfragen waren. Die Unterrichtsreihe konnte in allen Klassenstufen in der angedachten Zeit durchgeführt werden. Zudem zeigte sich HyperDocSystems als ein geeignetes und zuverlässiges Tool, um digitale Arbeitsblätter mit Differenzierungsangebot einzusetzen und die Nutzungsdaten zu erfassen. Auch die Erhebung über analoge Fragebögen erwies sich als sinnvoll, da die Datenerhebung flexibel an die Herausforderungen einer Erhebung im Regelunterricht angepasst werden konnte. Insgesamt zeigen sich jedoch auch einige methodische Schwächen der Studie, die im Folgenden dargelegt werden.

#### **Messinstrumente**

Aufgrund der Corona-Pandemie konnten die Pilotierungen teilweise nicht in vollem Umfang durchgeführt werden, wodurch die Testinstrumente ebenfalls nicht vollumfänglich mit Schüler:innen getestet werden konnten. Zudem musste der jeweilig untersuchte Begriff in den Instrumenten, zum Beispiel „activity“ beim Instrument für die intrinsische

Motivation, durch „digitale Arbeitsblätter“ ersetzt werden. Daher wiesen manche Messinstrumente, insbesondere der Fragebögen zur Erfassung des Cognitive Load und der intrinsischen Motivation, für einige Subskalen eine geringe Reliabilität und Konstruktvalidität auf, die durch entsprechendes Entfernen von Items oder Skalen jedoch verbessert werden konnten. An dieser Stelle kann auch gesagt werden, dass es sich bei dem Instrument zur Erfassung des Load um ein reliables Messinstrument handeln sollte, das ausreichend geprüft wurde. In einer weiteren Studie der Autor:innen zeigen sich erneut gute Reliabilitäten, jedoch weist das Faktorenmodell keinen exzellenten Fit auf (Klepsch & Seufert, 2020). Es ist anzumerken, dass die Validierung, beziehungsweise die durchgeführten Studien, mit Studierenden durchgeführt wurden und damit das Instrument womöglich nicht vollumfänglich ohne erneute Prüfung in der Schule eingesetzt werden kann.

Die Usability wurde in Form einer Selbsteinschätzung erhoben. Diese Methode wird teilweise als ungenügend erachtet (z. B. Holzinger, 2005), sodass weitere Verfahren eingesetzt werden sollten. Das war jedoch im Rahmen des vorliegenden Studiendesigns nicht möglich. In Bezug auf den USE bleibt die Frage offen, ob es sich bei der Einfachheit der Nutzung und der Einfachheit des Lernens der Nutzung um ein oder mehrere Konstrukte handelt. Lund (2001) findet Hinweise für beide Annahmen. In der vorliegenden Studie wurden ebenfalls zwei Konstrukte angenommen. Die Problematik der Konstruktvalidität zeigen auch Gao et al. (2018) auf. Durch eine Reduktion der Items konnte allerdings für die vorliegende Arbeit ein Modell mit einem guten Fit gefunden werden.

Bei vielen eingesetzten Skalen liegt keine Normalverteilung der Daten vor. Für die eingesetzten Verfahren stellt das in Anbetracht der Stichprobengröße kein Problem dar. Dennoch wurde beispielsweise bei der Faktorenanalyse ein spezieller Schätzer für nichtnormalverteilte Daten verwendet, wodurch die Ergebnisse noch einmal für normalverteilte Daten verifiziert werden müssten. Einerseits sollten nach der klassischen Testtheorie schiefe Daten vermieden werden, da viele statistische Tests die Normalverteilung der Daten voraussetzen. Andererseits erweist sich zum Beispiel die Usability aufgrund der Benutzeroberfläche von HyperDocSystems als hoch, womit nicht zwangsläufig ein falsch konstruiertes Messinstrument vorliegt. Zudem erweisen sich einige statistische Methoden als robust gegen die Verletzung der Normalverteilung (Rasch & Guiard, 2004) oder es bestehen alternative robuste Verfahren, die hier eingesetzt wurden.

### **Durchführung der Studie und Stichprobe**

In Bezug auf das Studiendesign ergeben sich weitere Aspekte, die bei der Bewertung der Datengüte zu berücksichtigen sind. Die Studie fand im Zeitraum der Corona-Pandemie statt. Das hatte keinen unmittelbaren Einfluss auf das Unterrichtsgeschehen, allerdings veränderte sich das Lernumfeld im Vorfeld der Erhebung dahingehend, dass die Schüler:innen im Zeitraum von März bis April im Homeschooling und von Mai bis Juli 2020 im Wechselunterricht unterrichtet wurden. In diesem Zeitraum haben sich die Lernenden, aber auch die Lehrkräfte, stärker mit dem digitalen Unterricht auseinandergesetzt, wodurch eine andere Beziehung und Einstellung gegenüber digitalen Geräten vermutet werden kann. Bedingt durch die Corona-Pandemie konnte zudem die Vergleichsgruppe erst zu einem späteren Zeitpunkt erhoben werden (Juni bis Juli 2021). Im Zeitraum Januar bis Mai 2021 wurde erneut im Homeschooling unterrichtet, wodurch natürlicherweise das Lernen in einem digitalen Rahmen stattfand. Es kann daher sein, dass in der unmittelbaren Zeit danach, in der die Schüler:innen wieder mit analogen Materialien unterrichtet wurden, eine veränderte Einstellung gegenüber diesem Unterricht bestand. In welche Richtung die eben beschriebene Veränderung wirkt, kann abschließend nicht gesagt werden. Im Unterricht konnten keine unmittelbaren negativen Folgen des Homeschoolings, zum Beispiel in Form von speziellen Aussagen, beobachtet werden. Ein Einfluss könnte jedoch darin bestehen, dass die Interventionsgruppe nach den Sommerferien und die Vergleichsgruppe vor den Sommerferien erhoben wurde. Dadurch könnte die Vergleichsgruppe eine geringere grundlegende Motivation als die Interventionsgruppe aufweisen. Allerdings kann auch argumentiert werden, dass die Schüler:innen der Vergleichsgruppe unmittelbar aus dem Homeschooling kamen und daher ein besonderes Interesse am regulären Unterricht zeigten. Es wurde in diesem Fall ebenfalls explizit darauf geachtet, alle Gruppen gleich zu behandeln und damit den Rosenthal- und Pygmalion-Effekt auszuschließen. Dafür wurde für beide Gruppen die gleiche Didaktik und Methodik, mit Ausnahme der HyperDocs, verwendet. Da die Untersuchungen von der gleichen Person durchgeführt wurden, erhöht das einerseits die Gefahr einer Ungleichbehandlung der jeweiligen Untersuchungsgruppen, andererseits kann so ein standardisierter ausreichend erprobter Unterricht durchgeführt werden, der unter anderem neben der groben Unterrichtsstruktur auf der Makroebene, auch auf der Mikroebene in Form der Unterrichtsgespräche und Impulse, gleiche Bedingungen schafft. Weiterhin waren im Vorfeld der Erhebung keine Informationen über den Leistungsstand oder andere Persönlichkeitsanlagen der Klassen bekannt. Hinsichtlich der Zusammensetzung der Vergleichsgruppe sei an dieser

Stelle auf die besondere Situation an einer rheinland-pfälzischen Schule hingewiesen, die nach dem G8-Abitur ausgebildet und damit im Vergleich zu anderen Schulen früher in den Oberstufenunterricht, beginnend ab der 10. Jahrgangsstufe, einführt. Aus einer solchen Schule stammten zwei achte und eine zehnte Klasse (formal Grundkurs), die zur Mittelstufe der Vergleichsgruppe gezählt wurden. Mit Rücksprache der Lehrkräfte und in Hinblick auf den Entwicklungsstand der Schüler:innen der achten Klasse, die sich am Ende der Jahrgangsstufe befanden, stellte diese Zuordnung kein Problem dar. Die 10. Jahrgangsstufe wurde ebenfalls von der Lehrkraft als eine durchschnittliche Klasse wahrgenommen, sodass sich gegen eine Zuordnung zur Oberstufe entschieden wurde. Bekräftigt wird diese Entscheidung dadurch, dass die Schüler:innen das Fach noch nicht abwählen konnten.

Zudem konnte leider keine umfangreiche Stichprobe der Gesamtschulen für die Oberstufe oder die Vergleichsgruppe gewonnen werden, wodurch Unterschiede zwischen den Schulformen lediglich für die Interventionsgruppe der Mittelstufe untersucht werden konnten. Insgesamt ist anzumerken, dass die Vergleichsgruppe circa um die Hälfte kleiner ausfällt als die Interventionsgruppe. Aufgrund dessen variieren zum Teil die Streuungsmaße zwischen den Gruppen, wodurch robuste Testverfahren eingesetzt werden mussten.

### **Teststatistik**

Auch in Hinblick auf die Auswertungsverfahren zeigen sich ein paar methodische Schwächen. Durch die gemischte Varianzanalyse mit zwei Zwischensubjekt- und einem Innersubjektfaktor entstehen kleine Teilstichproben, wodurch die statistische Power verringert wird. Durch die Schiefe der Daten, die fehlende Normalverteilung und das Vorhandensein von Ausreißern musste außerdem auf robuste Verfahren zur Datenauswertung zurückgegriffen werden. Dabei wird der Datensatz durch Winsorizing und Trimming verändert. Zusätzlich haben die verwendeten Funktionen unter anderem den Nachteil, nicht immer Effektstärken anzugeben. Auf der anderen Seite handelt es sich bei robusten Methoden um anerkannte Verfahren, die zudem zuverlässige Ergebnisse bei Verletzung der Voraussetzungen liefern können (Mair & Wilcox, 2020). Nichtsdestotrotz sind teilweise viele Ausreißer, insbesondere bei den Statistiken zu den Lernhilfen und Vertiefungsaufgaben, vorhanden, was zu einer Verzerrung der Daten führen könnte. Da es sich jedoch um plausible Messwerte handelt, wurden diese nicht aus dem Datensatz entfernt.

Einen weiteren Kritikpunkt stellt das multiple Testen dar, durch das eine Alpha-Fehler-Kumulierung stattfindet. Eine Korrektur der p-Werte wird insbesondere bei konkreten Hypothesentests empfohlen, die jedoch hier keine Anwendung fanden (Rubin, 2021). Andere Forscher:innen stellen zudem das Studiendesign in den Vordergrund und betonen die Relevanz der Ergebnisse. So können beispielsweise signifikante Ergebnisse für die Praxis irrelevant sein und knapp nicht-signifikante Ergebnisse einen hohen Praxisbezug aufweisen (Feise, 2002). Wiederum andere Autor:innen fordern keine Adjustierung oder eine Adjustierung aller jemals veröffentlichten Studien über landesweite Datenbanken (An et al., 2013). Bei den vorliegenden Ergebnissen stellt sich zudem die Frage, auf welcher Ebene die Alpha-Fehler-Kumulierung stattfinden soll. Beispielsweise können alle in dieser Studie gerechneten, oder nur die zu den einzelnen Variablen durchgeführten t-Tests, angepasst werden. Das Gleiche gilt für die robusten gemischten Varianzanalysen, die wiederum mehrere statistische Ergebnisse mit p-Werten erbringen. Aus den dargestellten Gründen wurde auf eine Korrektur der Alpha-Fehler-Kumulierung verzichtet.

### **Lernhilfen**

Die Erfassung der Nutzung der Lernhilfen und Vertiefungsaufgaben in HyperDocSystems erfolgte weitestgehend problemlos. Vereinzelt wurde der Schließzeitpunkt der Hilfen nicht erfasst, da die Schüler:innen direkt den Browser geschlossen oder das Tablet ausgeschaltet haben. Das betrifft jedoch fast ausschließlich diejenigen Hilfen, die als letztes, am Ende der Stunde, geöffnet wurden. An dieser Stelle ist zu erwähnen, dass die Nutzung des Differenzierungsangebots nicht vollständig kontrolliert werden konnte. Beispielsweise könnten benachbarte Schüler:innen gemeinsam eine Hilfe oder Vertiefungsaufgabe auf einem Tablet gelesen haben. Da allerdings jede:r Lernende ein eigenes Tablet verwendete, ist dieser Fall eher unwahrscheinlich. Wahrscheinlicher ist, dass interessierte Schüler:innen sich an der Nutzung ihrer Mitschüler:innen orientiert haben. Ein weiterer Faktor in Bezug auf den Vergleich zwischen Interventions- und Vergleichsgruppe stellt die Gestaltung der Lernhilfen dar. Die Vergleichsgruppe konnte natürlicherweise keine Hilfen in Form eines Videos oder einer Audiodatei nutzen. Insgesamt ist jedoch die Nutzung dieser Formate auch bei der Interventionsgruppe gering. Außerdem sind Hilfen in Form von Videos und Audios ein Teil des Digitalen und damit Bestandteil der Intervention. Bezüglich der Hilfen lassen sich weitere Kritikpunkte formulieren. Natürlicherweise variieren der Informationsgehalt und Nutzen verschiedener Modalitäten, beispielsweise zwischen Video und Text. Zwar wurde versucht, den Informationsgehalt möglichst gleich



zu halten, allerdings gelingt dies nicht immer, da beispielsweise Videos dynamische Vorgänge besser visualisieren können. Zudem konnten bei manchen Aufgaben keine Hilfen in Form eines Videos oder einer Bildfolge erstellt werden. Daher variiert zum Teil die Anzahl der zur Verfügung stehenden Hilfen pro Aufgabe und Arbeitsblatt. Aus diesem Grund wurde bei der Auswertung der Nutzungshäufigkeit die Anzahl der Hilfen auf einem Arbeitsblatt berücksichtigt. Weiterhin kann in diesem Zusammenhang auf die Methode der Auswertung der Hilfenutzung verwiesen werden. Insgesamt wird die Einschätzung einer Hilfe als „gelesen“ zu optimistisch sein, da für Bilder eine Betrachtungsdauer von einer Sekunde und für Texte die maximale Leserate einer 10.-Klässlerin oder eines 10.-Klässlers angenommen wurde. Auf der anderen Seite wird die Häufigkeit der „gelesenen“ Hilfen mit einem Video oder einer Audiodatei supprimiert, da das Medium vollständig betrachtet beziehungsweise gehört werden muss, damit es als „gelesen“ gilt. In der Realität wird eine Schülerin oder ein Schüler jedoch gelegentlich ein Video ab der Mitte schauen oder bei erneuter Betrachtung an die entsprechende Stelle im Video springen. Da jedoch insgesamt die Nutzung dieser beiden Hilfearten gering war, ist die Verzerrung gering.

Beim Nachvollziehen der Nutzungshäufigkeiten des Differenzierungsangebots in der Vergleichsgruppe wurde zudem auf eine Selbstangabe auf den Kärtchen und im Fragebogen zurückgegriffen. In der Praxis konnte beobachtet werden, dass die Schüler:innen die Nutzung der Lernhilfen nur unvollständig protokollierten. Für die Arbeit wurde auf die Selbstangaben auf dem Fragebogen zurückgegriffen. Die Daten sind gegebenenfalls verzerrt. Eine Videografiestudie könnte daher die Nutzung der Lernhilfen besser nachvollziehen. Zudem ist anzumerken, dass die Lernhilfen nicht unmittelbar an das individuelle Vorwissen der teilnehmenden Lerngruppen angepasst wurden und eher ein breites Spektrum an Hilfen eingesetzt wurde. Kleinert und Kolleg:innen (2021) weisen allerdings darauf hin, dass die Lernhilfen an die Lernvoraussetzungen der Lerngruppe angepasst werden sollten. Auf der anderen Seite wäre ein solcher Einsatz lediglich im Rahmen eines qualitativen Designs mit einem geringeren Stichprobenumfang zu realisieren.

### **HyperDocSystems**

In Verbindung mit HyperDocSystems kam es bei wenigen Schüler:innen durch Nutzungsfehler zu einem Verlust der bisherigen Erarbeitungen. Die Wiederherstellungsfunktion wurde danach von einigen Lernenden nicht immer verwendet, was sich manchmal in einer gewissen Frustration äußerte. Für zukünftige Erhebungen müsste diese Funktion

stärker in den Fokus gerückt werden, auch auf Ebene des Programms, indem beispielsweise der Hinweis über den gesamten Bildschirm platziert wird. Ein weiterer Aspekt bezüglich HyperDocSystems ist die Verwendung des Programms im Unterricht. In der vorliegenden Studie wurden nicht alle Potentiale, wie zum Beispiel das gemeinsame Betrachten von Schüler:innenlösungen, ausgeschöpft, da dieser Nutzung die fehlende technische Ausstattung an den Schulen entgegenstand. Durch die aktuelle Anlage der Studie ergibt sich dafür unter Betrachtung des Stundenablaufs eine hohe Vergleichbarkeit zwischen Interventions- und Vergleichsgruppe, aufgrund der ansonsten gleichen Unterrichtsdidaktik und -methodik der Stunden.

### **Unterrichtsreihe**

Zuletzt sei auf die generelle Gestaltung der Unterrichtsreihe hingewiesen. Innerhalb der Unterrichtsreihe variiert der Kompetenzschwerpunkt und damit die Gestaltung der Arbeitsblätter. Sowohl in der Mittel- als auch in der Oberstufe ist die eigentliche Unterrichtszeit in der ersten Stunde kürzer als in den anderen Stunden, da dort ein längeres Erhebungsinstrument Anwendung findet. In der Mittelstufe wird in der zweiten Stunde ein sehr textlastiges Arbeitsblatt verwendet, während in der dritten Stunde die mathematische graphische Auswertung im Vordergrund steht. Dadurch können die Hilfenutzung und das Interesse schwanken und möglicherweise den Neuheitseffekt korrumpieren. Ein Vergleich zwischen der Mittel- und Oberstufe wird ebenfalls erschwert, da in der Oberstufe in der dritten Stunde mehr Textarbeit erforderlich ist und in der zweiten Stunde die graphische Auswertung im Vordergrund steht. Dem ist allerdings hinzuzufügen, dass im Regelunterricht prinzipiell die Art der Schulstunden variieren wird und damit die Ergebnisse eher den im Feld vorliegenden Voraussetzungen entsprechen werden. Insbesondere ist bei der Durchführung einer Unterrichtsreihe eine gewisse Progression der Anforderungsbereiche zu erwarten. Gleichzeitig wurde bei der Untersuchung in der Oberstufe nicht zwischen Grund- und Leistungskurs differenziert und die gleiche Unterrichtsreihe in beiden Gruppen durchgeführt, um eine größere und breiter gefasste Stichprobe zu erhalten. Es bleibt offen, ob eine Unterscheidung in Grund- und Leistungskurs sinnvoll ist, da einerseits starke Grundkurse und andererseits schwache Leistungskurse existieren. Zudem sind die Leistungskurse oftmals sehr klein, was eine umfangreiche Erhebung deutlich erschwert.

### 6.3.2 Limitationen

Aus der Anlage der Studie und der Zusammensetzung der jeweiligen Untersuchungsgruppen ergeben sich Limitationen. Zunächst gelten die hier gewonnenen Erkenntnisse lediglich für die neunte und zehnte beziehungsweise die elfte und zwölfte Jahrgangsstufe im Fach Chemie. In den unteren Klassenstufen, die zum Teil einen geringen Kontakt mit digitalen Medien vorweisen, können die Ergebnisse in Bezug auf den Cognitive Load und die Motivation variieren. Zudem konnte für die Gesamtschule keine Vergleichsgruppe erhoben werden. Da jedoch bei den Gesamtschüler:innen der Interventionsgruppe ähnliche Tendenzen wie bei den Gymnasiast:innen gefunden wurden, sollten auch die Vergleiche zur Vergleichsgruppe ähnlich sein. Die Ergebnisse gelten jedoch nur für die Mittelstufe als gesichert.

Die Ergebnisse in Bezug auf das Interesse nach der Selbstbestimmungstheorie gelten für das Lernen mit Arbeitsblättern. Ein Einfluss auf das situative Interesse im Chemieunterricht ist zu vermuten. Damit könnte langfristig auch das individuelle Interesse gefördert werden. Zudem beschränken sich die aufgezeigten Ergebnisse zunächst auf den Chemieunterricht, da in anderen Fächern, insbesondere den Nicht-Naturwissenschaften, andere Interessen, auch zwischen den Geschlechtern, vorliegen. Weiterhin lassen sich aufgrund einer anderen Usability die Ergebnisse bezüglich HyperDocSystems nicht ohne weiteres auf jedes andere digitale Tool übertragen. Das betrifft insbesondere die Erkenntnisse bezüglich der Motivation. Hierbei wiesen bereits Hillmayr et al. (2020) auf Unterschiede zwischen den verschiedenen digitalen Tools hin. In diesem Zusammenhang ist zudem nicht auszuschließen, dass das motivationsfördernde Potential von HyperDocSystems nach einer längeren Einsatzdauer von mehr als vier Stunden weiter abnimmt. Die Nutzung der Lernhilfen kann nicht vollständig durch die in dieser Arbeit erhobenen Variablen erklärt werden. Ein möglicher Erklärungsansatz für die individuelle Nutzung der Lernhilfen könnte durch die Art des Help-Seeking gegeben werden, das jedoch ebenfalls nicht erhoben wurde. Aus den in diesen Studien beobachteten Vorteilen von HyperDocs gegenüber analogen Arbeitsblättern kann nicht zwangsläufig auf eine bessere Lernwirksamkeit geschlossen werden. Der Einfluss der Motivation auf die Lernwirksamkeit ist jedoch allgemein anerkannt (Lazowski & Hulleman, 2016). Auch der positive Einfluss von Lernhilfen und der Usability auf das Verständnis zeigt sich in Studien (z. B. Arnold et al., 2016; Meiselwitz & Sadera, 2008), weshalb insgesamt von einem lernwirksamen Effekt ausgegangen werden kann.

## **6.4 Empfehlungen und Implikationen**

Aufgrund der umfangreichen Erkenntnisse und Limitationen, die aus der vorliegenden Studie gewonnen werden konnten, sollen weiterführende Empfehlungen und Implikationen für das Lernen mit digitalen Arbeitsblättern und Lernhilfen gegeben werden. Dabei wird ein besonderer Schwerpunkt auf HyperDocSystems gesetzt. Die Empfehlungen beruhen einerseits auf den erhobenen Daten der Studie und andererseits auf den gewonnenen Erfahrungen aus den unterrichteten Stunden mit HyperDocSystems. Die Empfehlungen müssen daher in Teilen in weiteren Studien verifiziert werden.

### **6.4.1 Intrinsische Motivation und Interesse**

Bei Untersuchungen der Motivation und des Interesses im Zusammenhang mit digitalen Medien sollte die Usability mindestens einmal über den Untersuchungszeitraum erhoben werden, um gegebenenfalls ausbleibende Effekte der Intervention näher untersuchen zu können. Zusätzlich sollten Interventionen über einen längeren Zeitraum erfolgen, da beispielsweise das Interesse beim Lernen mit HyperDocs über die Einsatzdauer abnimmt. Dieser Effekt zeigt sich insbesondere in der Oberstufe, wo die analoge Vergleichsgruppe sich in der vierten Stunde der digitalen Interventionsgruppe stark annähert. Zusätzlich können neben der Usability auch noch weitere Variablen wie der Cognitive Load zur Klärung der Wirkmechanismen beitragen. Dabei sollte die Validität und Reliabilität der hier verwendeten Messinstrumente erneut geprüft werden. Gegebenenfalls sind andere kürzere Messinstrumente einzusetzen, um die gesamte Bearbeitungszeit der Fragebögen zu verringern. Ein Beispiel für die Erfassung der intrinsischen Motivation nach der Self-Determination Theory ist die Kurzskala Intrinsische Motivation (Wilde et al., 2009). Zusätzlich sollte explizit das situative und individuelle Interesse expliziter erhoben werden, um HyperDocs als catch- beziehungsweise hold-Komponente bestätigen zu können. In Bezug auf den Cognitive Load sollte der Einsatz von weiteren Messinstrumenten oder eine Verbesserung des hier verwendeten Instruments geprüft werden. Um stärker Schul- und Klasseneffekte berücksichtigen zu können, wäre eine Mehrebenenanalyse anzudenken. Diese erfordert allerdings große Stichproben auf Basis der einzelnen Ebenen (Pötschke, 2019), die in dieser Studie nicht erreicht werden konnten. In diesem Zusammenhang sollten in zukünftigen Studien die Erhebungszeiträume und Erhebungsumfänge angepasst werden. Um eine genauere isolierte Untersuchung der Motivation im Verlauf der Unterrichtsreihe zu untersuchen, sollten gleichbleibende Unterrichtsstunden in dem

jeweiligen Kompetenzschwerpunkt untersucht und der Anforderungsbereich kontrolliert werden. Idealerweise sollten die Stunden ebenfalls zur gleichen Zeit (Schulstunde) stattfinden, da ein Einfluss des Tageszeitpunktes möglich erscheint. Um eine bessere Differenzierung der Untersuchungsergebnisse in der Oberstufe zu erreichen, sollte zwischen Grund- und Leistungskursen unterschieden werden und diesbezüglich eine differenziertere Auswertung erfolgen.

Lehrkräfte sollten HyperDocs gezielt und reflektiert einsetzen. In der Mittelstufe profitieren die Schüler:innen generell vom Einsatz der digitalen Arbeitsblätter, auch noch nach vier Unterrichtsstunden. Aufgrund der unklaren Entwicklung im weiteren Verlauf, sollte diese Methode jedoch insbesondere für schwierige und gegebenenfalls theoretische Inhalte verwendet werden, um gerade in diesen Stunden die intrinsische Motivation zu steigern. Vor allem Schülerinnen profitieren, in Hinblick auf die Bearbeitung von Arbeitsblättern, vom Einsatz der HyperDocs. In der Oberstufe sind die Unterschiede zu der herkömmlichen Unterrichtsform weniger stark. Gleichzeitig nimmt mit zunehmender Einsatzdauer die intrinsische Motivation bei den weiblichen Lernenden stärker ab als bei den Schülern. Das zeigt sich auch in der Abnahme des Interesses am Fach Chemie bei den Schülerinnen. Digitale Arbeitsblätter sollten daher auch bei dieser Gruppe punktuell eingesetzt werden. Zudem sollte der Zugewinn, beispielsweise durch die Differenzierung, sorgfältig geprüft werden. Unabhängig von HyperDocs sollten Lehrkräfte vorab die Benutzerfreundlichkeit der eingesetzten digitalen Werkzeuge testen, da die Usability einen Einfluss auf die intrinsische Motivation und damit wahrscheinlich auch auf die Lernwirksamkeit nimmt.

#### **6.4.2 Usability**

In Bezug auf die Usability ergeben sich mehrere Empfehlungen für die Forschungspraxis. Aufgrund der signifikanten Abnahme der Benutzerfreundlichkeit über die Einsatzdauer, sollen Forschende die Usability einer Anwendung zu mehreren Zeitpunkten erheben, da ansonsten eine falsche Wahrnehmung der Benutzerfreundlichkeit entstehen kann. Dies kann gerade bei älteren Schüler:innen (Oberstufe) der Fall sein. Diese Tatsache setzt ein kurzes Erhebungsinstrument voraus, das zu keiner Überladung des Fragebogens führt. In diesem Zusammenhang konnte ein kurzes, reliables und valides Instrument auf Basis von Lund (2001) geschaffen werden. Das Instrument eignet sich daher für den Einsatz im Unterricht, wobei die hohe Korrelation zwischen der Einfachheit der Nutzung und der Einfachheit des Erlernens der Nutzung berücksichtigt werden muss. Der Fragebogen

bietet allerdings eine gute Alternative zum SUS, sollten beispielsweise verschiedene Aspekte der Usability untersucht werden. Neben der Usability ist außerdem die Erhebung weiterer Variablen sinnvoll, falls die Untersuchung der Anwendung im Vordergrund steht. Beispielsweise sollten der Cognitive Load und die tabletbezogene Selbstwirksamkeit miterhoben werden, um die Bewertung der Benutzerfreundlichkeit besser verstehen zu können. Sinnvoll wäre sicherlich auch die Betrachtung des Interesses, wobei dadurch der Befragungsumfang zunimmt.

Für Lehrkräfte bedeuten die Ergebnisse dieser Arbeit einen bedenkenlosen Einsatz in der Mittel- und Oberstufe, da die Schüler:innen HyperDocs intuitiv und schnell nutzen können. Aufgrund der hohen Benutzerfreundlichkeit ist auch ein Einsatz in der Unterstufe denkbar. Die Einstiegshürden sind damit gering. Gleichzeitig bewerten die Lernenden das System als nützlich und sind insgesamt damit zufrieden. Lehrer:innen können dadurch sehr effizient binnendifferenzierte Arbeitsmaterialien, auch für Hausaufgaben oder sonstige unterrichtsferne Arbeitsaufträge, erstellen.

### **6.4.3 Lernhilfen und Vertiefungsaufgaben**

Im Bereich der digitalen Binnendifferenzierung besteht zum aktuellen Zeitpunkt ein großer Forschungsbedarf. Diese Arbeit leistet einen ersten Schritt zur konzeptionellen digitalen Integration von Lernhilfen und Vertiefungsaufgaben. Mit HyperDocSystems wird eine flexible Erhebung der Nutzungshäufigkeiten gewährleistet, womit neue Forschungsfragen beantwortet werden können. Bei der Auswertung der Hilfenutzung können die hier angewandten Heuristiken umgesetzt werden. Demnach sollten Hilfen und Vertiefungsaufgaben unter Berücksichtigung der Öffnungsdauer und des Gesamtkontextes, wie der Anzahl der Hilfen und der verfügbaren Zeit, ausgewertet werden. Dabei sollten allerdings auch die Einschränkungen dieser Methode in Bezug auf Audio- und Videohilfen berücksichtigt werden. Je nach Forschungsfrage wäre aufgrund der geringen Nutzung dieser Formate darauf zu verzichten, um eine exaktere Auswertung zu ermöglichen. Dabei muss allerdings berücksichtigt werden, dass auch die Hilfen in Form eines Bildes nicht eindeutig als gelesen identifiziert werden können. Dieser Problematik könnte mit einer qualitativen Erhebung, zum Beispiel Videografie, begegnet werden. Zusätzlich sollte in diesem Zusammenhang verifiziert werden, ob eine Vertiefungsaufgabe tatsächlich bearbeitet wurde. Bei Studien, die den Einsatz von Lernhilfen untersuchen, sollte beachtet werden, dass die Lernenden ihre Kompetenzen im Bereich des Help-Seeking wahrscheinlich noch nicht vollständig ausprägen. Dadurch können die Ergebnisse verzerrt werden, weshalb

eine längere Nutzungsphase oder eine explizite Förderung dieser Kompetenz empfohlen wird. Gleichzeitig bietet es sich an, weitere Variablen, wie beispielsweise den Cognitive Load, zu erheben, um auf eine breite Datenbasis für die Interpretation der Nutzung des Differenzierungsangebots zurückgreifen zu können. Zudem könnten die jeweiligen Unterrichtsstunden an den voraussichtlich anzutreffenden Leistungsstand der Schüler:innen angepasst werden. Weiterhin sollte für eine genaue Erklärung der Hilfenutzung das Help-Seeking-Verhalten erhoben werden (z. B. Wolters et al., 2005), um so die Ergebnisse der Hilfenutzung besser einordnen zu können. In diesem Zusammenhang wäre auch die Erhebung des Selbstkonzepts im jeweiligen Fach wichtig, da hierbei zwischen den Geschlechtern zum Teil Unterschiede bestehen (Jansen et al., 2013).

Für Lehrkräfte zeigt die vorliegende Arbeit einige Umsetzungsanregungen für den Einsatz von Lernhilfen und Vertiefungsaufgaben im Unterricht. Aus organisatorischer Sicht eignet sich der Einsatz von HyperDocSystems, um differenziertes Material schnell zu erstellen und zu distribuieren, zum Beispiel auch in Form von Arbeitsaufträgen oder Hausaufgaben, bei denen die Lehrkraft nicht anwesend sein kann. Zusätzlich bieten HyperDocs die Integration von Audio- und Videoinhalten ohne großen Mehraufwand. Diese Medien werden von einem kleinen Teil der Schülerschaft verwendet. Aus den Nutzungshäufigkeiten kann möglicherweise auf Lernprobleme geschlossen werden. Sie sollten allerdings nicht alleinstehend zur Auswertung herangezogen werden. Die Daten müssten durch eine Rückmeldung von den Schüler:innen in einem qualitativen Setting verifiziert werden. Beim punktuellen oder einmaligen Einsatz von HyperDocSystems ist außerdem mit einer zunächst unreflektierten Verwendung der Lernhilfen zu rechnen, wobei das Interesse und die Neugier genutzt werden können, um das Interesse am Lerngegenstand zu erhöhen.

#### **6.4.4 Weiterentwicklung von HyperDocSystems und ähnlichen Systemen**

HyperDocSystems hat sich im Laufe der Entwicklung und der Erprobung als ein zuverlässiges und nützliches Werkzeug erwiesen, um binnendifferenzierte digitale Arbeitsmaterialien zu erstellen. Die umfangreichen Funktionen des Programms ermöglichen vielfältige Einsatzzwecke im Unterricht und darüber hinaus. Insbesondere die digitale Umsetzung von Arbeitsblättern bietet viele Vorteile, wie beispielsweise die Integration verschiedener Medien, das ortsunabhängige Lernen oder auch die unkomplizierte Integration von Lernhilfen.

An dieser Stelle sollen Empfehlungen für die Weiterentwicklung von HyperDocSystems oder ähnlichen Systemen gegeben werden. Diese Empfehlungen beruhen zum einen auf den empirisch gewonnenen Daten der Studie und zum anderen auf noch nicht belegten, aber anzunehmenden Wirkmechanismen, die sich aus den vielzählig gehaltenen Unterrichtsstunden mit HyperDocSystems ergeben. Wie sich zeigt, nimmt die Motivation beim Lernen mit den digitalen Arbeitsblättern ab. Dennoch bleibt die Motivation auch nach dem mehrmaligen Einsatz über dem Niveau der analogen Variante. Es ist zu vermuten, dass mit HyperDocSystems das situationale Interesse erhöht werden kann. Um dieses situationale Interesse weiter zu fördern und gegebenenfalls in ein langandauerndes individuelles Interesse zu überführen, beziehungsweise den Lerngegenstand in das Selbstbild der oder des Lernenden zu integrieren, könnte eine weitere hold-Komponente in HyperDocSystems integriert werden. Als hold-Komponente schlägt Mitchell (1993) die Bedeutsamkeit des Lerngegenstandes und die Eingebundenheit der Schüler:innen in die Lernumgebung vor. Diese Eingebundenheit kann zum einen in der aktiven Beteiligung und Steuerung des Lernprozesses oder zum anderen in der sozialen Eingebundenheit verstanden werden. Damit zeigen sich Parallelen zur Selbstbestimmungstheorie, die die Autonomie und die soziale Eingebundenheit als menschliche Grundbedürfnisse definieren. HyperDocSystems sollte daher Funktionen integrieren, die diese beiden Aspekte stärker berücksichtigt. Eine Funktion könnte darin bestehen, für jedes Arbeitsblatt automatisch einen passenden Chat bereitzustellen, in dem die Schüler:innen miteinander kommunizieren und gegebenenfalls auch untereinander Fragen stellen können. Sie hätten damit einerseits die Wahl, die Lehrkraft oder andere Schüler:innen zu fragen (Autonomie) und sind andererseits in den Lernprozess eingebunden (soziale Eingebundenheit). Eine weitere Möglichkeit, die soziale Eingebundenheit zu erhöhen, bestünde in der stärkeren Integration von interaktiven Aktivitäten, wie beispielsweise der Nutzung von Mindmaps, Padlets oder Etherpads. Diese Funktionen können größtenteils bereits in HyperDocSystems durch das Einfügen externer Inhalte genutzt werden, allerdings erfordert diese Umsetzung ein gewisses technisches Know-how. Zudem ist es denkbar, dass die jeweiligen Betreiber:innen der Webseiten die externe Integration über *iFrames* aus Sicherheitsgründen zunehmend unterbinden. Daher könnten einige einfache interaktive Arbeitsformen als Plugin für HyperDocSystems entwickelt werden. Eine weitere Idee, um die soziale Eingebundenheit und Autonomie zu fördern, wäre die Möglichkeit, dass sich Schüler:innen gegenseitig kontrollieren und beispielsweise auch die Lösung kommentieren oder „Liken“ können. Damit wird die Lebenswelt der Lernenden miteingebunden. Diese Funktion bietet



jedoch ebenso einige Nachteile. Die Lernenden müssen dazu die richtige Lösung kennen. Zudem können schwächere Schüler:innen wahrscheinlich nicht in einem ausreichenden Maß Korrekturen vornehmen und es besteht eine Gefahr für das Selbstbild, wenn beispielsweise einige Schüler:innen in diesen kollaborativen Prozess nicht eingebunden werden.

Neben der Autonomie und sozialen Eingebundenheit spielt das Kompetenzerleben ebenfalls eine wichtige Rolle bei der Entwicklung intrinsischer Motivation. Es stellt sich daher die Frage, wie dieses menschliche Bedürfnis in HyperDocSystems gefördert werden kann. Eine Möglichkeit besteht darin, *Prompts* einzubauen, die den Lernenden bei der Entwicklung von metakognitiven Kompetenzen unterstützen. Solche Beispiele existieren bereits für Tutoringsysteme (z. B. Aleven et al., 2016). Außerdem können auch einfache Hinweise eingeblendet werden, beispielsweise nach einer gewissen Zeit oder nach einem bestimmten Absatz, die die Schüler:innen für die bisherige Arbeit loben und zur weiteren Bearbeitung anregen, um das sogenannte *Growth-Mindset* zu fördern. Dadurch könnte eine positivere Einstellung gegenüber der Schule entwickelt werden (Zeng et al., 2016). Außerdem könnte ein Bewertungssystem integriert werden, das eine genaue Rückmeldung und Bepunktung der bearbeiteten Arbeitsblätter ermöglicht und Schwerpunkte für die weitere Entwicklung der Schülerin oder des Schülers setzt. Diese Rückmeldung könnte in einem speziellen Dashboard angezeigt werden und dem Lernenden eine Fortschrittskontrolle ermöglichen.

Weitere Entwicklungsfelder beziehen sich weniger auf lernpsychologische Aspekte, sondern umfassen den praktischen Nutzen und die bisherigen Erfahrungen beim Einsatz von HyperDocSystems. Das Layout betreffend wäre die Integration von einzelnen „Seiten“ möglich, um die HyperDocs stärker an analogen Arbeitsblättern zu orientieren und diese übersichtlicher zu gestalten, beziehungsweise einzelne Inhalte voneinander abgrenzen zu können. Für eine bessere Vernetzung und Verknüpfung der Lerninhalte wäre es zudem von Vorteil, verschiedene HyperDocs, beispielsweise über einen Link, miteinander in Verbindung setzen zu können. Umfangreichere Funktionen umfassen eine intelligente Autokorrektur der Lösungen, eine intelligentere Auswertung der Hilfenutzungen sowie eine App, um Inhalte auch offline bearbeiten zu können. Eine intelligente Auswertung der Schüler:innenlösungen würde Lehrkräfte bei der Korrektur entlasten. Zudem könnten die Lernenden ein unmittelbares Feedback nach Bearbeitung der Aufgabe erhalten. Solche Systeme sind jedoch äußerst kompliziert in der Entwicklung und gerade bei offenen

Antworten schwierig zu implementieren. Bei diesem Ansatz sollte allerdings die Relevanz der Lehrkraft im Unterrichtsgeschehen nicht missachtet werden (z. B. Conradt & Bogner, 2016). Ein weiterer Anwendungszweck von künstlicher Intelligenz könnte in einer intelligenten Auswertung der Hilfenutzung liegen, sodass Lehrkräften Empfehlungen für die weitere Entwicklung von Lernhilfen gegeben werden. Eine solche Auswertung könnte beispielsweise bevorzugte Hilfearten oder den Umfang von Hilfen berücksichtigen. Gleichzeitig könnte damit eine Funktion integriert werden, die den Schüler:innen erlaubt, Hilfen zu bewerten, zum Beispiel mit einem „Daumen hoch“ oder „Daumen runter“. Damit hätte die Lehrkraft eine unmittelbare Rückmeldung zu den eingesetzten Hilfen. Gegebenenfalls ist in der weiteren Implementation eine differenziertere Rückmeldung mit einer genauen Begründung anzudenken. Bewertet eine Schülerin oder ein Schüler die Hilfe als „schlecht“ beziehungsweise nicht hilfreich, könnte ein weiterer Dialog mit vorgegebenen Antworten („nicht hilfreich“, „zu umfangreich“ oder „unverständlich“) ein genaueres Feedback beinhalten. Zuletzt wäre darüber nachzudenken, die Funktion, Arbeitsblätter zu teilen, weiter auszubauen, um einen breiteren Austausch zwischen Lehrenden zu ermöglichen. Das könnte in Form einer externen Webseite oder eines Forums geschehen. Beispielsweise könnten Lehrkräfte ihre Arbeitsblätter zur Verfügung stellen und andere Nutzer:innen von HyperDocSystems könnten sich diese einfach in ihre bestehende Datenbank von Arbeitsblättern kopieren.

## **6.5 Forschungsbedarf**

Im Laufe der Durchführung der Studie und der Analyse der Daten haben sich weitere Fragen eröffnet. Weiterhin ließen sich nicht alle interessanten und relevanten Fragestellungen im Umgang mit HyperDocs in einer einzelnen Studie untersuchen, weshalb im Folgenden weitere Forschungsfragen offengelegt werden.

### **6.5.1 Interesse und Motivation**

Im Bereich der Interessens- und Motivationsforschung stehen die Unterschiede der Geschlechter im Vordergrund. Hierbei sollte weiter untersucht werden, weshalb Unterschiede zwischen den Geschlechtern bestehen und welche Wirkmechanismen anzunehmen sind. Ein Beispiel, das die Relevanz dieses Forschungsgebiets herausstellt, ist die starke Abnahme des Interesses beim Lernen mit HyperDocs der Schülerinnen in der Intervention der Oberstufe ab der zweiten Stunde. Anzumerken ist hierbei, dass das

Interesse der weiblichen Lernenden in der ersten Stunde auf einem höheren Niveau liegt als das der Jungen. Ein weiteres Forschungsfeld betrifft eine genauere Untersuchung des situativen Interesses und des Neuheitseffekts. Genauer müsste überprüft werden, wie lange HyperDocs zu einer erhöhten Motivation gegenüber einer analogen Variante führen. Die in dieser Studie gefundenen Ergebnisse gelten für eine maximale Einsatzdauer von vier Unterrichtsstunden, was sich insbesondere für eine Unterrichtsreihe eignet. Bereits in dieser Zeitspanne war eine Annäherung des Interesses festzustellen. HyperDocs müssten daher über mehrere Wochen eingesetzt und mit einer analogen Variante verglichen werden. In Bezug auf den interessens- und motivationsfördernden Effekt könnte überprüft werden, ob sich der Effekt durch eine geschicktere Einbindung der Funktionen von HyperDocSystems weiter fördern lässt. In der vorliegenden Studie wurden einige Funktionen des Systems aufgrund der Ausstattung an den Schulen nicht verwendet. Das betrifft beispielsweise die Einsicht von Schüler:innenlösungen. Aus Sicht der Selbstbestimmungstheorie ist es denkbar, dass sich Schüler:innen sozial eingebunden fühlen oder Kompetenz erleben, wenn sie ihre Lösungen in Kombination mit der Präsentation des HyperDocs vor der Klasse vorstellen können. Zudem bestehen weitere Funktionen, die in einer analogen Variante nicht genutzt werden können. Das betrifft den Datei-Upload, bei dem Schüler:innen eigene Inhalte und Medien in das HyperDoc integrieren können. Weiterhin können durch externe Inhalte, sofern eine Internetverbindung besteht, kollaborative Werkzeuge (Etherpad, Padlet, etc.) eingebunden werden. Diese wiederum könnten ebenso für eine höhere soziale Eingebundenheit sorgen. Die eben genannten Einsatzzwecke erfordern allerdings eine sehr genaue Planung der Vergleichsgruppe, um die Wirksamkeit tatsächlich auf das Medium zurückführen zu können.

Von weiterem Interesse wäre die genaue Untersuchung des Zusammenhangs zwischen Usability und intrinsischer Motivation, sodass zukünftige Anwendungen gezielt optimiert und angepasst werden können. Hierbei spielt die User Experience, die in der vorliegenden Arbeit nicht untersucht wurde, sicherlich ebenfalls eine größere Rolle. In diesem Zusammenhang können möglicherweise qualitative Untersuchungen einen tieferen Einblick gewähren, um zu identifizieren, welche Aspekte an HyperDocs die Schüler:innen besonders motivieren. Alle Ergebnisse der jetzigen Studie sollten weiterhin in der Unterstufe geprüft werden, sodass der Einsatz von HyperDocs in Bezug auf das motivationsfördernde Potential bedenkenlos für alle Klassenstufen empfohlen werden kann. Eine weitere wichtige Forschungsfrage betrifft die Lernwirksamkeit von HyperDocs. Es wurde bereits der Einfluss des Interesses und der Motivation auf das Lernen aufgezeigt. Daher ist durch eine

Förderung des Interesses als Mediator auf eine höhere Lernwirksamkeit zu schließen. Da nun die Usability der HyperDocs ausreichend geprüft wurde, wäre die Untersuchung der Lernwirksamkeit ein wichtiger nächster Schritt.

### **6.5.2 Usability**

Bei der Usability steht die Abnahme über den Untersuchungszeitraum im Vordergrund, da die Ergebnisse nicht eindeutig erklärt werden können. Hierbei wird ein Zusammenhang zu dem Interesse oder dem Cognitive Load vermutet. In weiteren Studien müsste daher der Einfluss dieser Variablen in einem angepassten Studiendesign näher überprüft werden, indem diese Variable bewusst manipuliert wird. Dabei müsste untersucht werden, ob es sich im schulischen Kontext um ein generelles Phänomen handelt oder eine Abhängigkeit vom Tool oder Thema vorliegt. Zudem müsste die Usability der HyperDocs für die Unterstufe erhoben werden, um auch dort eine hohe Funktionalität und Passung sicherzustellen. Zusätzlich ist das Konstrukt der Usability im Kontext der Schule neu zu beleuchten. Während die Usabilitystudien an Erwachsenen bereits eine Vielzahl von Wirkungsmechanismen und Einflussfaktoren identifizieren konnten, fehlen diese für Schüler:innen oder sind, insbesondere für den Schulkontext, nicht ausreichend erprobt. Von daher sind die Zusammenhänge, die in dieser Arbeit gefunden wurden, zu verifizieren und gegebenenfalls in die Bewertung von zukünftigen Usabilityuntersuchungen einzubinden. In Bezug auf die Usability sollte weiterführend geprüft werden, welche Oberflächen- und Funktionselemente als besonders nützlich, zufriedenstellend und einfach in der Nutzung erachtet wurden und welche nicht. Damit könnten zum einen HyperDocSystems optimiert und zum anderen generelle Empfehlungen für die Gestaltung von ähnlichen Softwaresystemen gegeben werden.

### **6.5.3 Lernhilfen und Vertiefungsaufgaben**

Eine wesentliche Funktion von HyperDocSystems und ein Forschungsschwerpunkt dieser Dissertation bildet das digitale Differenzierungsangebot. Diesbezüglich ergeben sich in Folge der Studie weitere Desiderata. Bisher konnten in der Literatur nur wenige Studien gefunden werden, die eine quantitative Betrachtung der Nutzung des Differenzierungsangebots vornahmen. Die vorliegende Arbeit bietet in diesem Zusammenhang wichtige Erkenntnisse. Diese beruhen zum Teil auf Annahmen, die aus dem aktuellen Forschungsstand gewonnen wurden. Ein Beispiel betrifft die Einordnung einer Hilfe als „gelesen“. Für diese Einordnung wurden die Lesegeschwindigkeit, die Abspieldauer einer

Audio- und Videodatei oder, im Falle von Bildern, eine feste Sekundenzahl angenommen. Das ist eine starke Vereinfachung. Daher müssten für zukünftige Projekte im Bereich der digitalen Differenzierung durch Lernhilfen genauere Möglichkeiten der Erfassung der vollständigen Bearbeitungsdauer gefunden werden. Dadurch wäre in Zukunft auch die Einbindung von künstlicher Intelligenz besser möglich, da diese Verfahren eine genaue Datenerfassung voraussetzen. Ein weiteres Forschungsfeld betrifft die Art des Differenzierungsangebots. In dieser Studie wurden keine gestuften Lernhilfen, sondern einfache Lernhilfen, unterteilt in verschiedene Sinnesmodalitäten, eingesetzt. Diese Lernwirksamkeit beziehungsweise die Annahme über verschiedene Lerntypen ist weiterhin umstritten (z. B. Pashler et al., 2008). Von daher wäre eine gezielte Untersuchung dieses Konzepts mit HyperDocSystems vorzunehmen. Beispielsweise könnten in qualitativen Untersuchungen genaue Nutzungsgründe für die bevorzugte Nutzung der verschiedenen Hilfen erforscht werden. In einem zweiten Schritt könnte die Lernwirksamkeit dieses Konzepts untersucht werden, indem digitale Arbeitsblätter mit einem unterschiedlich stark differenzierten Unterstützungsangebot eingesetzt werden. Es ist denkbar, dass beispielsweise in den Sprachen die Nützlichkeit von auditiven Hilfen größer ist als im naturwissenschaftlichen Unterricht. Allgemeiner betrachtet müsste zudem die Lernwirksamkeit für digitale Lernhilfen prinzipiell bestätigt werden. Während für analoge (gestufte) Lernhilfen solche Ergebnisse bereits vorliegen, ist zum jetzigen Zeitpunkt die Nützlichkeit der digitalen Variante noch nicht bestätigt. Es ist daher zum Beispiel nicht sicher, ob die Motivationsförderung von HyperDocs auf die Lernhilfen oder auf andere Faktoren zurückzuführen ist (vgl. nächsten Abschnitt). Aus Gründen der Vergleichbarkeit wurde die analoge Methode exakt an die digitale Methode, im Rahmen der Möglichkeiten, angeglichen. Daher wurden allen Schüler:innen der Vergleichsgruppe Hilfen in Umschlägen präsentiert. Dieses Verfahren stellt Lehrkräfte in der Praxis vor große organisatorische Herausforderungen. In diesem Zusammenhang könnten die Ergebnisse älterer Studien, die den Kontext des Help-Seeking näher untersuchten, in einer experimentellen Studie mit HyperDocSystems (Experimentalgruppe) und zentralen Hilfekarten am Pult (Kontrollgruppe) repliziert werden. Weiterhin wurden in der Studie in der Vergleichsgruppe keine audiovisuellen oder auditiven Hilfen verwendet. Damit wurde sowohl das Medium als auch die prinzipielle Auswahl des Differenzierungsangebots variiert. Ebenso lässt sich nicht eindeutig belegen, ob HyperDocs aufgrund ihrer digitalen Beschaffenheit oder aufgrund der einfachen Verfügbarkeit von Lernhilfen motivationsfördernd wirken. Für zukünftige Studien wird daher ein 2x2-Design mit den Varianten „HyperDoc mit / ohne Lernhilfe“ und

„Analoges Arbeitsblatt mit / ohne Lernhilfe“ vorgeschlagen. Einen weiteren Punkt betrifft das Nutzungsangebot der Lernhilfen. So ist zu prüfen, ob ein Zusammenhang zwischen dem Lerntyp und dem bevorzugten Darstellungstyp der Lernhilfen besteht. Dadurch wäre die Differenzierung in die verschiedenen Darstellungsformen gerechtfertigt. Zusätzlich ist das bereits oben beschriebene Prinzip der multimedialen Lernhilfen auf seine Wirksamkeit hin zu untersuchen, da es Lehrkräfte vor einen Mehraufwand stellt. In Bezug auf die Lernhilfen sollten zukünftige Studien die hier angewandten Heuristiken zur Hilfeauswertung weiter verfeinern und untersuchen, um die zukünftige Entwicklung der automatischen Erfassung und Verarbeitung der Nutzung von Lernhilfen voranzutreiben. Zudem wäre es vorzuziehen, die verschiedenen Variablen (Motivation, Cognitive Load, Help-Seeking) in einzelnen Studien zu untersuchen, um die Last durch die Beantwortung der Fragebögen zu reduzieren und damit auch eine gezieltere Analyse der Wirkmechanismen durch zusätzliche Variablen zu ermöglichen. Diese Analysen könnten zum Beispiel in Form von Strukturgleichungsmodellen durchgeführt werden.

Die erhobenen Daten lassen primär nur Rückschlüsse auf die Mittelstufe und Oberstufe des Chemieunterrichts von Gymnasien und Gesamtschulen zu, wobei auch keine ausreichenden Aussagen in Bezug auf die Oberstufe der Gesamtschule getroffen werden können. HyperDocs sollten daher in einem breiteren Rahmen in weiteren Fächern, in der Unterstufe und in anderen Schulformen eingesetzt werden. Beispielsweise ist es denkbar, dass Schüler:innen der Realschule stärker von einem breiten Differenzierungsangebot profitieren als Gymnasiast:innen oder dass der Nutzen des digitalen Differenzierungsangebots in den Sprachen größer ist.

## **6.6 Kritische Würdigung**

Die durchgeführte Studie liefert einen wichtigen Beitrag zum Lernen mit digitalen Arbeitsblättern mit Differenzierungsangebot. Dabei wurde ein realitätsnaher Forschungsansatz im Feld gewählt, um eine hohe Relevanz für den alltäglichen Schulgebrauch zu schaffen. Dieser Ansatz geht mit einigen Schwächen einher, die bereits in der Methodenkritik aufgeführt wurden. Positiv hervorzuheben ist der mehrstündige Einsatz der Arbeitsblätter im Regelunterricht Chemie von mehreren Schulen und Klassenstufen. Damit wurde eine breite Stichprobe geschaffen. Gleichzeitig fand die Erhebung im Rahmen der Corona-Pandemie statt, wodurch die Pilotierung und die Erhebung der Vergleichsgruppe verzögert wurden. In nachfolgenden Arbeiten sollte daher eine parallele Erhebung

zwischen Interventions- und Vergleichsgruppe stattfinden. Zudem müssen die eingesetzten Instrumente nochmal in Hinblick auf Reliabilität geprüft und gegebenenfalls umformuliert werden. Das betrifft insbesondere die Instrumente zur Erhebung der Motivation nach der Selbstbestimmungstheorie und die Erhebung des Cognitive Load. Das Instrument zur Erfassung der Usability hat sich hingegen als valide und reliabel erwiesen. Damit konnte ein wichtiger Beitrag zum Einsatz des USE und zur Beforschung der Usability im deutschsprachigen Bildungsraum geleistet werden. Weiterhin konnte ein wichtiger Beitrag zum Lernen mit einem digitalen differenzierten Lernangebot erbracht werden. Diese Erkenntnisse sind jedoch vor allem auf den gymnasialen Schulbereich beschränkt. Daher bleibt eine Verallgemeinerung noch offen, wobei die angenommenen Wirkmechanismen zur Motivation, zum Cognitive Load oder zur Usability allgemein gültig sein dürften und daher die Ergebnisse diesbezüglich nicht stark variieren sollten.

Wenige Fragen in Bezug auf die Nutzung von digitalen Lernhilfen und zum Lernen mit digitalen Arbeitsblättern unter Berücksichtigung der Motivation bleiben offen. Es ist nicht geklärt, inwieweit die Motivation auch über einen längeren Zeitraum aufrechterhalten werden kann. Dazu wäre eine umfangreichere Studie über mehrere Wochen und Monate nötig. Zudem konnte nicht geklärt werden, welche Aspekte der digitalen Arbeitsblätter für die Förderung der Motivation und des Interesses gegenüber der analogen Variante verantwortlich sind. Gleiches gilt für das Differenzierungsangebot. Es bleibt damit auch offen, ob die Lernwirksamkeit durch die digitalen Arbeitsblätter gesteigert werden kann und inwieweit die Lernhilfen dazu beitragen. Ein konsequenter Einsatz von HyperDocs im Regelunterricht könnte das motivationsfördernde Potential mindern oder gänzlich aufheben, falls das Potential auf dem Neuheitseffekt beruht.

Unter Berücksichtigung der erhobenen und ausgewerteten Daten können jedoch viele bisherige Ergebnisse bestätigt und erweitert werden. Der vermutete motivationsfördernde Aspekt von digitalen Medien kann auch für HyperDocs angenommen werden. Auch der Zusammenhang zwischen Usability und Interesse zeigte sich mit einem mittelstarken Effekt in der Erhebung. Dadurch konnten die Ergebnisse von Karapanos und Kolleg:innen (2018) repliziert werden. Im Themengebiet der digitalen Binnendifferenzierung konnte diese Arbeit neue Erkenntnisse und Daten gewinnen, die zu einer besseren Integration von digitalen Lernhilfen in den Unterricht beitragen.

## **7. Fazit**

Die vorliegende Arbeit versucht, einen schulnahen Beitrag zum Lernen mit digitalen Arbeitsblättern und Lernhilfen zu schaffen, indem unter Berücksichtigung der Motivations- und Interessensforschung sowie der theoretischen Grundlagen des Lernens mit digitalen Medien HyperDocs im Regelunterricht erprobt wurden. Die folgenden Unterkapitel fassen die Ergebnisse der Arbeit zusammen und geben einen Ausblick auf die zukünftige Entwicklung beim Lernen mit digitalen (differenzierten) Lernmaterialien.

### **7.1 Zusammenfassung**

Im Schwerpunkt des Interesses und der intrinsischen Motivation konnten viele der bisherigen Erkenntnisse gestützt werden, auch in Bezug auf die Unterschiede zwischen den Geschlechtern. Die Variablen der intrinsischen Motivation zeigen Korrelationen zu anderen Variablen, wie beispielsweise der Usability. Die Usability von HyperDocs wurde als hoch eingestuft, wobei Unterschiede zwischen den Messzeitpunkten bestehen. Im Forschungsschwerpunkt der Lernhilfen zeigen sich teilweise Ergebnisse, die nicht allein mit den vorliegenden Variablen zu erklären sind. Es konnten jedoch wichtige Erkenntnisse bezüglich der Nutzung von Lernhilfen gewonnen werden. In den folgenden Unterkapiteln werden die Ergebnisse der einzelnen Forschungsgebiete detaillierter zusammengefasst.

#### **7.1.1 Intrinsische Motivation und Interesse**

Der Einsatz von HyperDocs führt bei Mittel- und Oberstufenschüler:innen zu einer erhöhten intrinsischen Motivation gegenüber der analogen Arbeitsweise. Je nach Teilstichprobe und Unterrichtsstunde liegt zwischen den beiden verschiedenen Arbeitsformen mehr als eine Einheit des Mittelwertes, wodurch HyperDocs als motivationsfördernd angesehen werden können. Dieser Effekt ist jedoch endlich, da die Förderung der Motivation mit zunehmender Einsatzdauer abnimmt. Gerade in der Oberstufe zeigen sich am Ende der Reihe je nach Variable nur noch geringe Unterschiede zwischen der Interventions- und der Vergleichsgruppe. Teilweise profitieren die Schülerinnen der Oberstufe insbesondere in den ersten Stunden des Einsatzes, was eine gezieltere Verwendung im Unterricht erfordert. Auf der anderen Seite entstehen durch den Einsatz der HyperDocs keine signifikanten Nachteile, sodass selbst nach längerer Einsatzdauer die Motivation beim Lernen mit dieser Methode gegenüber der analogen Variante höher ist. In der Mittelstufe steigt das Fachinteresse Chemie beim Lernen mit HyperDocs insbesondere bei



den Schülern an, in der Oberstufe nimmt es bei den Schülerinnen ab. Die Ergebnisse sind jedoch nicht signifikant, was an dem geringen Zeithorizont liegen kann. Die Gründe für die Ergebnisse der weiblichen Lernenden in der Oberstufe sind nicht eindeutig identifizierbar. Verantwortlich für die Abnahme des Interesses könnte eine Kombination aus Überforderung aus fachlicher und technischer Sicht sein, die sich in einem erhöhten Cognitive Load und einer schlechteren Bewertung der Usability zeigt. Die Ergebnisse sind allerdings nicht immer eindeutig und signifikant, sodass weitere Untersuchungen bezüglich der Geschlechterunterschiede erforderlich sind. Dafür sollte ein größerer Zeithorizont für die Untersuchung eingeplant werden, um die langfristige Auswirkung des Einsatzes der HyperDocs näher untersuchen zu können. Im Schulformvergleich erfahren die Gesamtschüler:innen der Mittelstufe weniger wahrgenommene Wahlfreiheit (Autonomie), die Schülerinnen weniger Interesse (SDT) und das Fachinteresse Chemie fällt bei ihnen ebenfalls leicht ab. Die Schüler hingegen erfahren ein größeres Fachinteresse in der Gesamtschule. Begründet werden können die Schulformunterschiede durch den erhöhten Cognitive Load. HyperDocs sind damit insgesamt für den Einsatz in der Mittelstufe bei der Schulformen und in der Oberstufe des Gymnasiums zu empfehlen, sofern die technischen Voraussetzungen erfüllt sind. Für die Oberstufe der Gesamtschule können aufgrund der fehlenden Daten keine Aussagen getroffen werden. Allerdings sollten dort, unter Berücksichtigung der Ergebnisse der Gesamtschüler:innen der Mittelstufe, ähnliche Ergebnisse wie bei den Gymnasiast:innen beobachtet werden.

Die vorliegende Arbeit konnte außerdem den Bezug zwischen Usability und Motivation bekräftigen, wodurch der Benutzerfreundlichkeit ein hoher Stellenwert bei der Entwicklung und Evaluation digitaler Medien eingeräumt werden sollte. Zwischen nahezu allen Variablen der beiden Konstrukte zeigen sich für die Mittel- und Oberstufe signifikante positive Zusammenhänge.

### **7.1.2 Usability**

Die Ergebnisse der Usabilityerhebung zeigen eine hohe Benutzerfreundlichkeit von HyperDocs in der Mittel- und Oberstufe. Zwischen den Schulformen finden sich keine großen Unterschiede. Besonders positiv hervorzuheben sind die einfache Nutzung und das einfache Erlernen der Nutzung. Die Schüler:innen konnten bereits nach der ersten Stunde das System zielführend nutzen. Die Zufriedenheit und Nützlichkeit erfahren eine geringere, aber dennoch hohe Bewertung. Die Oberstufe zeigt sich in diesem Zusammenhang kritischer. Die tabletbezogene Selbstwirksamkeit und der Cognitive Load können mit der

Usability in Zusammenhang gebracht werden. Dabei kann jedoch nicht immer eine Kausalität hergestellt werden. Im Falle des Cognitive Load sind beide Wirkungsrichtungen denkbar, wonach eine erhöhte Last zu einer geringen Bewertung der Benutzerfreundlichkeit oder die geringe Benutzerfreundlichkeit zu einem höheren Cognitive Load führt. Die Kausalität der Zusammenhänge müsste daher, auch für die anderen Variablen, in einer weiteren Studie untersucht werden. Diese Faktoren sind daher bei der Bewertung der Benutzbarkeit gegebenenfalls, je nach Forschungsfrage, mitzubersichtigen. Die Arbeit konnte damit zur Schließung von Forschungslücken beitragen und unter anderem weitere Variablen identifizieren, die mit der Usability in Zusammenhang stehen. Zusätzlich konnten die Validität und die Reliabilität des Messinstruments nach Lund (2001) bestätigt werden. Dabei zeigt sich allerdings eine hohe Korrelation zwischen der Einfachheit der Nutzung und dem Erlernen der Nutzung. Das Instrument konnte zusätzlich erfolgreich auf weniger Items gekürzt werden, was einen Einsatz in Befragungen mit Schüler:innen erleichtert. Etwas überraschend ist der Abfall in der Bewertung der Usability über den Nutzungszeitraum. Der Grund dafür könnte in dem bereits erwähnten Abfall des Interesses liegen. Erstaunlich ist in diesem Zusammenhang jedoch der Abfall bei den Variablen „Einfachheit der Nutzung“ und „Einfachheit des Lernens der Nutzung“. Diese sollten über den Messzeitraum konstant bleiben, da keine neuen Funktionen eingeführt wurden. Wegen des abfallenden Interesses und der steigenden kognitiven Last in den letzten Stunden der Erhebung, könnte durch den Recency-Effekt die Minderung der Usability erklärt werden. Für die Forschungspraxis bedeutet das, die Benutzerfreundlichkeit nach der einmaligen und mehrmaligen Nutzung zu erheben, da mitunter ein größerer Abfall, wie hier in der Oberstufe, wahrnehmbar ist. Bei geringen Forschungskapazitäten wäre gegebenenfalls eine einzelne Erhebung am Ende der Untersuchungsreihe vorzuschlagen. In einer qualitativen Studie könnte untersucht werden, inwieweit die einzelnen Funktionen optimiert und ergänzt werden können, um die Nützlichkeit der HyperDocs und die Zufriedenheit der Nutzer:innen weiter zu steigern.

### **7.1.3 Lernhilfen- und Vertiefungsaufgaben**

In Hinblick auf die Differenzierung mit digitalen Lernhilfen konnte diese Arbeit Erkenntnisse liefern, wonach Lernende hauptsächlich Bild- und Texthilfen verwenden. Dabei veranlassen die Neugier und das Interesse die Schüler:innen zur Nutzung der Lernhilfen und Vertiefungsaufgaben. Das digitale Setting hat hierbei den Vorteil, dass dieses Interesse und die Neugier länger bestehen bleiben als bei analoger Nutzung. Zudem zeigt sich

die digitale Konzeption als flexibler, weil verschiedene Arten von Hilfen leicht eingebunden werden können. Dennoch muss die adäquate Verwendung der Lernhilfen zunächst erlernt werden, wie die Daten dieser Arbeit zeigen. Damit können bisherige Ergebnisse der Literatur (z. B. Aleven & Koedinger, 2000) weiter gefestigt werden. Das indirekte Schließen auf das Interesse oder den Cognitive Load der Schüler:innen aufgrund der Hilfenutzungen ist zum jetzigen Zeitpunkt nicht möglich. Zwar finden sich Hinweise darauf, dass Lernende bei höherer kognitiver Last zu einer Unterstützung greifen, allerdings sind die Daten nicht eindeutig interpretierbar. Es sind vermutlich mehrere Faktoren beteiligt, die auf die Verwendung Einfluss nehmen. Diese Erkenntnisse gelten sowohl für die Mittel- als auch für die Oberstufe und die Gesamtschule. Die Datenerhebung über HyperDocSystems eignet sich, um die Nutzung des Differenzierungsangebots erheben zu können. Bei der anschließenden Auswertung wurden allerdings Heuristiken angewandt, die weiterführend optimiert werden müssten. Beispielsweise wäre empirisch für die einzelnen Stufen (Unter,- Mittel,- Oberstufe) zu überprüfen, ab wann eine Hilfe als gelesen angesehen werden kann. Zusätzlich müsste ein qualitatives Studiendesign die Nutzungsgründe für das Differenzierungsangebot näher in den Blick nehmen und die Veränderung der Nutzungsmuster analysieren. Unter Berücksichtigung des Genderaspekts zeigt sich bei den Schülerinnen in der Mittelstufe und Oberstufe eine höhere Nutzung von Lernhilfen. In der deskriptiven Analyse konnten die bereits diskutierten Variablen als eine mögliche Ursache identifiziert werden. Da die Unterschiede zum Teil nicht signifikant sind und nicht alle möglichen Wirkmechanismen in dieser Arbeit untersucht wurden, sind weitere Studien anzustellen.

Die Vertiefungsaufgaben wurden bei allen Lerngruppen vor allem aus Interesse und als zeitlicher Puffer betrachtet. Ob die Schüler:innen die Aufgaben auch bearbeiteten, wurde nicht geprüft. Bezüglich der Nutzungshäufigkeiten liegen die Schüler des Gymnasiums in zwei der vier Stunden über den Schülerinnen, in der Gesamtschule ist der Effekt umgekehrt. In der Oberstufe nutzen tendenziell auch die Schüler häufiger eine Vertiefungsaufgabe. Über den gesamten Zeitraum nimmt das Aufrufen solcher Aufgaben ab, was sich durch die geringere zur Verfügung stehende Arbeitszeit und das abnehmende Interesse gut erklären lässt. Für eine genaue Analyse der Nutzung des Vertiefungsangebots sind jedoch qualitative Studien, die auch die Beantwortung der Fragen in den Fokus nehmen, zu berücksichtigen.

## 7.2 Ausblick

Aufgrund der bisherigen Entwicklungen im deutschen Schulsystem ist eine zunehmende Digitalisierung des Unterrichts zu erwarten. Dies ist als positiv zu erachten, da mit dem Einsatz digitaler Medien viele organisatorische und lernpsychologische Vorteile einhergehen. Insbesondere die Individualisierung des Lernprozesses wird durch den Einsatz von Tutoringsystemen und digitalen Arbeitsmaterialien weiter vorangetrieben. In diesem Zusammenhang ist mit einer zunehmenden Integration von künstlicher Intelligenz zu rechnen, die die Lehrkraft nicht ersetzen, sondern unterstützen soll. Das kann zum Beispiel in Form einer besseren Passung der Lernmittel auf die Lerngruppe erfolgen. In diesem Zusammenhang bildet HyperDocSystems ein grundlegendes Werkzeug, mithilfe dessen dieser Prozess vollzogen werden kann. Neben dem Vorteil der Individualisierung bietet der Einsatz der digitalen Medien auch motivationsfördernde Potentiale, die auch darin begründet sein könnten, dass sich die analoge Arbeitsvariante mit Papier und Stift immer mehr von der Lebenswelt der Schüler:innen entfernt. Damit steht die Entwicklung der Schule im Spannungsfeld zwischen alten, erprobten sowie zuverlässigen Unterrichtsstrukturen und neuen Werkzeugen und Methoden, die wiederum eine adäquate Ausstattung der Bildungseinrichtungen voraussetzen. Hierin dürfte wohl der größte Gelingensfaktor der Integration von digitalen Arbeitsmaterialien in den Schulen liegen. Ohne die nötige Ausstattung, auch in Bezug auf eine breite Internetanbindung, können diese Werkzeuge nicht eingesetzt werden. Der DigitalPakt stellt in Zusammenhang mit dem Ausbau der Digitalisierung einen wichtigen ersten Schritt dar. Ein weiterer Faktor für den Einsatz neuer digitaler Werkzeuge im Unterricht ist die in dieser Arbeit noch nicht betrachtete Rolle von Lehrkräften. Verschiedene Programme und Modelle zur Förderung der digitalen Kompetenzen bestehen bereits (z. B. DigCompEdu; Redecker & Punie, 2017). Für eine konsequente Integration von digitalen Werkzeugen in den Unterricht müssen die Akteur:innen daher weiter für das Thema ausgebildet und sensibilisiert werden, auch unter Berücksichtigung eines sinnvollen Umgangs mit den neuen Möglichkeiten des Unterrichts.

HyperDocSystems ist ein wichtiges Unterrichts- und Forschungswerkzeug. Einerseits bietet das Programm bereits eine Vielzahl sinnvoller und relevanter Funktionen, die einen breiten Einsatz im Unterricht ermöglichen. Das Tool hat sich bereits als reliabel und einfach in der Nutzung erwiesen. Andererseits wurden im Rahmen dieser Arbeit weitere Entwicklungen in Betracht gezogen, sodass HyperDocSystems zu einem

umfangreicheren und „intelligenteren“ System heranwachsen kann. Hierbei spielt die zukünftige Entwicklung der Schulen sicher eine große Rolle. Sollte die Akzeptanz des Einsatzes von digitalen Arbeitsmaterialien und intelligenten Lernsystemen weiter steigen, zeigt sich auch eine Notwendigkeit der Evolution von HyperDocs. In Ausblick zu stellen wäre außerdem, unabhängig von der zukünftigen Entwicklung im Bildungswesen, eine aktualisierte Version des Programms, das auf einer schlankeren und moderneren Entwicklungsbasis aufgebaut wird. HyperDocSystems wurde über eine Zeitspanne von vier Jahren entwickelt, erprobt und weiterentwickelt, wodurch einige Schwächen in der aktuellen Softwarebasis aufgedeckt werden konnten. Diese Schwächen könnten jedoch mit einem anderen Back- und Frontend weitestgehend aufgehoben werden. Zusätzlich wäre HyperDocSystems damit für die zukünftige Entwicklung gerüstet. Insgesamt bietet das Programm eine große Chance und Potential für die Integration von digitalen Arbeitsblättern mit Differenzierungsangebot in den Unterricht.

Die Themen der Motivation und Differenzierung werden sicherlich weiterhin eine große Rolle im Bereich der Bildung spielen. Die Geschichte der Motivationsforschung besitzt eine lange Tradition und Forschungshistorie. Es könnte der Eindruck gewonnen werden, dass die grundlegenden Annahmen bereits auszureichend ergründet sind und weitere Forschung lediglich andere Begriffe für die gleichen oder ähnliche Theorien einführt. Dennoch zeigen sich mit der stetigen Entwicklung der Lehr- und Lernmethoden neue Forschungsbedarfe in Hinblick auf den geeigneten Einsatz dieser neuen Möglichkeiten und ihre prinzipielle Nützlichkeit. Ohne Beachtung der Nützlichkeit in ihrer lehr-lernpsychologischen oder organisatorischen Dimension, dürften sich digitale Werkzeuge nicht durchsetzen. Die organisatorischen Vorteile des Einsatzes von digitalen Medien, zum Beispiel weniger Traglast oder die Einbindung von Messsystemen, sind zurzeit in der Schule nur in Teilen ersichtlich, unter anderem auch aufgrund der fehlenden Ausstattung in den Schulen. Umso wichtiger ist es daher, weiterhin die Vorteile von digitalen Arbeitsblättern und ähnlichem in Bezug auf den Lernprozess zu untersuchen, um den Einsatz dieser Medien zu rechtfertigen. Menschen werden außerdem immer unterschiedliche Voraussetzungen und personelle Hintergründe einbringen, die im Unterricht bedient werden müssen. Unter Berücksichtigung der zunehmenden Digitalisierung bietet HyperDocSystems eine Möglichkeit, mit geringem Aufwand die Bedürfnisse der Individuen zu bedienen. In Kombination mit intelligenten Systemen bietet die Zukunft sicher eine noch bessere Passung und Bedienung der Inhalte für unterschiedliche Personen. Dabei ist jedoch hervorzuheben, dass die Rolle der Lehrkraft keineswegs ersetzt, sondern lediglich ergänzt

und entlastet werden soll, zum Beispiel durch eine individuellere Passung der Arbeitsmaterialien für Lernende. Insbesondere in Momenten, in denen die Lehrkraft nicht anwesend sein kann, wie beispielsweise bei Hausaufgaben oder dem Lernen zuhause, bieten passgenaue Lernhilfen eine sinnvolle Unterstützung. Es ist zu hoffen, dass die Unterstützung über die Schule hinaus durch die neu gewonnen Möglichkeiten zunimmt.

## 8. Literaturverzeichnis

- Abuhlfaia, K. & Quincey, E. de. (2018). The usability of E-learning platforms in higher education: A Systematic Mapping Study. *Proceedings of British HCI 2018*, 1–13. <https://doi.org/10.14236/ewic/HCI2018.7>
- Affeldt, F., Siol, A., Markic, S. & Eilks, I. (2018). Neue Ansätze zur Differenzierung im Schülerlabor. *CHEMKON*, 25(7), 255–262. <https://doi.org/10.1002/ckon.201700502>
- Aleven, V. & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson & K. VanLehn (Hrsg.), *Intelligent Tutoring Systems* (S. 292–303). Springer. [https://doi.org/10.1007/3-540-45108-0\\_33](https://doi.org/10.1007/3-540-45108-0_33)
- Aleven, V., McLaren, B. & Koedinger, K. (2006). Towards Computer-Based Tutoring of Help-Seeking Skills. In S. A. Karabenick & R. S. Newman (Hrsg.), *Help seeking in academic settings: Goals, groups, and contexts* (1. Auflage). Routledge.
- Aleven, V., Roll, I., McLaren, B. M. & Koedinger, K. R. (2016). Help Helps, But Only So Much: Research on Help Seeking with Intelligent Tutoring Systems. *International Journal of Artificial Intelligence in Education*, 26(1), 205–223. <https://doi.org/10.1007/s40593-015-0089-1>
- Aleven, V., Stahl, E., Schworm, S., Fischer, F. & Wallace, R. (2003). Help Seeking and Help Design in Interactive Learning Environments. *Review of Educational Research*, 73(3), 277–320. <https://doi.org/10.3102/00346543073003277>
- Algharaibeh, S. A. S. (2020). Should I ask for help? The role of motivation and help-seeking in students' academic achievement: A path analysis model. *Cypriot Journal of Educational Sciences*, 15(5). <https://doi.org/10.18844/cjes.v15i5.5193>
- Algina, J., Keselman, H. J. & Penfield, R. D. (2005). An Alternative to Cohen's Standardized Mean Difference Effect Size: A Robust Parameter and Confidence Interval in the Two Independent Groups Case. *Psychological Methods*, 10(3), 317–328. <https://doi.org/10.1037/1082-989X.10.3.317>
- Allison, P. D. (2010). Missing Data. In P. V. Marsden & J. D. Wright (Hrsg.), *Handbook of Survey Research* (2. Auflage, S. 631–657). Emerald Group Publishing Limited.

- Allison, P. D. (2012). *Handling Missing Data by Maximum Likelihood*. <https://statisticalhorizons.com/wp-content/uploads/2022/01/MissingDataByML.pdf> (letzter Zugriff: 22.09.2023)
- Altrichter, H., Trautmann, M., Wischer, B., Sommerauer, S. & Doppler, B. (2009). Unterrichten in heterogenen Gruppen: Das Qualitätspotenzial von Individualisierung, Differenzierung und Klassenschülerzahl. In W. Specht (Hrsg.), *Nationaler Bildungsbericht Österreich 2009* (Bd. 2, S. 341–360). Leykam.
- An, Q., Xu, D. & Brooks, G. (2013). Type I Error Rates and Power of Multiple Hypothesis Testing Procedures in Factorial ANOVA. *Multiple Linear Regression Viewpoints*, 39(2), 1–16.
- Apache Friends (2019). *XAMPP Installers and Downloads for Apache Friends*. <https://www.apachefriends.org/index.html> (letzter Zugriff: 22.09.2023)
- Arnold, J., Kremer, K. & Mayer, J. (2016). Scaffolding beim Forschenden Lernen: Eine empirische Untersuchung zur Wirkung von Lernunterstützungen. *Zeitschrift für Didaktik der Naturwissenschaften*, 23(1), 21–37. <https://doi.org/10.1007/s40573-016-0053-0>
- Averina, F. E. & Kuswandono, P. (2022). High School Students' Motivation and Help-Seeking Strategies in English Language Learning. *Journal of English Language Teaching and Linguistics*, 7(1), 63–81. <https://doi.org/10.21462/jeltl.v7i1.719>
- Ayres, P. & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (2. Auflage, S. 206–226). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.011>
- Bak, P. M. (2019). *Lernen, Motivation und Emotion: Allgemeine Psychologie II – das Wichtigste, prägnant und anwendungsorientiert*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-59691-3>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W .H. Freemanand Company.



- Barnum, C., Bevan, N., Cockton, G., Nielsen, J., Spool, J. & Wixon, D. (2003). The „magic number 5“: Is it enough for web testing? *Conference Proceedings, Conference on Human Factors in Computing Systems*, 698–699.  
<https://doi.org/10.1145/765891.765936>
- Barnum, C. M. (2021). *Usability testing essentials: Ready, set...test!* (2. Auflage). Morgan Kaufmann.
- Bartholomé, T., Stahl, E., Pieschl, S. & Bromme, R. (2006). What matters in help-seeking? A study of help effectiveness and learner-related factors. *Computers in Human Behavior*, 22(1), 113–129. <https://doi.org/10.1016/j.chb.2005.01.007>
- Bätz, K., Beck, L., Kramer, L., Niestradt, J. & Wilde, M. (2009). Wie beeinflusst Schülermitbestimmung im Biologieunterricht intrinsische Motivation und Wissenserwerb? *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 307–323.
- Baumert, J., Maaz, K., Stanat, P. & Watermann, R. (2009). Schulkomposition oder Institution – was zählt? *DDS – Die Deutsche Schule*, 101(1), 33–46.
- Beal, C. R., Qu, L. & Lee, H. (2008). Mathematics motivation and achievement as predictors of high school students’ guessing and help-seeking with instructional software: Mathematics motivation and achievement. *Journal of Computer Assisted Learning*, 24(6), 507–514. <https://doi.org/10.1111/j.1365-2729.2008.00288.x>
- Bescherer, C. & Spannagel, C. (2011). *CUSE-D-r Fragebogen zur computerbezogenen Selbstwirksamkeit—Reduziert*. <https://csp.uber.space/phhd/CUSE-D-r.pdf> (letzter Zugriff: 22.09.2023)
- Bevan, N. (2009). What is the difference between the purpose of usability and user experience evaluation methods? *Proceedings of the Workshop UXEM*, 1–4.
- Bevan, N. (1995a). Usability is Quality of Use. In Y. Anzai, K. Ogawa & H. Mori (Hrsg.), *Advances in Human Factors/Ergonomics* (Bd. 20, S. 349–354). Elsevier.  
[https://doi.org/10.1016/S0921-2647\(06\)80241-8](https://doi.org/10.1016/S0921-2647(06)80241-8)
- Bevan, N., Carter, J., Earthy, J., Geis, T. & Harker, S. (2016). New ISO Standards for Usability, Usability Reports and Usability Measures. In M. Kurosu (Hrsg.), *Human-Computer Interaction. Theory, Design, Development and Practice* (S. 268–278). Springer International Publishing. [https://doi.org/10.1007/978-3-319-39510-4\\_25](https://doi.org/10.1007/978-3-319-39510-4_25)

- Bevan, N., Carter, J. & Harker, S. (2015). ISO 9241-11 Revised: What Have We Learnt About Usability Since 1998? In M. Kurosu (Hrsg.), *Human-Computer Interaction: Design and Evaluation* (Bd. 9169, S. 143–151). Springer International Publishing. [https://doi.org/10.1007/978-3-319-20901-2\\_13](https://doi.org/10.1007/978-3-319-20901-2_13)
- Bevan, N., Kirakowski, J. & Maissel, J. (1991). What is Usability? *Proceedings of the 4th International Conference on HCI*. HCI International, Stuttgart.
- Billestrup, J., Bruun, A. & Stage, J. (2016). Usability Problems Experienced by Different Groups of Skilled Internet Users: Gender, Age, and Background. In C. Bogdan, J. Gulliksen, S. Sauer, P. Forbrig, M. Winckler, C. Johnson, P. Palanque, R. Bernhaupt & F. Kis (Hrsg.), *Human-Centered and Error-Resilient Systems Development* (S. 45–55). Springer International Publishing. [https://doi.org/10.1007/978-3-319-44902-9\\_4](https://doi.org/10.1007/978-3-319-44902-9_4)
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler ; mit 87 Tabellen* (4., überarb. Aufl., [Nachdr.]). Springer.
- Breidenstein, G., Menzel, C. & Rademacher, S. (2013). Legitime und illegitime Differenzen im individualisierten Unterricht. Beobachtungen aus einer Montessori-Schule. In J. Budde (Hrsg.), *Unscharfe Einsätze: (Re-)Produktion von Heterogenität im schulischen Feld* (Bd. 42, S. 153–167). Springer Fachmedien Wiesbaden. [https://doi.org/10.1007/978-3-531-19039-6\\_7](https://doi.org/10.1007/978-3-531-19039-6_7)
- Brooke, J. (1996). SUS: A „Quick and Dirty“ Usability Scale. In P. W. Jordan, B. Thomas, I. L. McClelland & B. Weerdmeester (Hrsg.), *Usability Evaluation In Industry* (1. Auflage, S. 189–194). CRC Press.
- Bruckermann, T., Aschermann, E., Bresges, A. & Schlüter, K. (2017). Metacognitive and multimedia support of experiments in inquiry learning for science teacher preparation. *International Journal of Science Education*, 39(6), 701–722. <https://doi.org/10.1080/09500693.2017.1301691>
- Bruckermann, T., Diederich, A., Schlueter, K. & Edelman, H. (2016). Does the use of mobile multimedia devices in practical lessons affect the motivation of pupils? *School Science Review*, 97(361), 113–120.
- Bru-nO (2017, 16. März). *Gasflasche*. <https://pixabay.com/de/photos/gas-schweißen-schweißer-schutzgas-2147493/> (letzter Zugriff: 22.09.2023)

- Bruun, A. & Ahm, S. (2015). Mind the Gap! Comparing Retrospective and Concurrent Ratings of Emotion in User Experience Evaluation. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque & M. Winckler (Hrsg.), *Human-Computer Interaction – INTERACT 2015* (Bd. 9296, S. 237–254). Springer International Publishing. [https://doi.org/10.1007/978-3-319-22701-6\\_17](https://doi.org/10.1007/978-3-319-22701-6_17)
- Butler, R. (1998). Determinants of Help Seeking: Relations Between Perceived Reasons for Classroom Help-Avoidance and Help-Seeking Behaviors in an Experimental Context. *Journal of Educational Psychology*, *90*(4), 630–643.
- Cassidy, S. & Eachus, P. (2002). Developing the Computer User Self-Efficacy (Cuse) Scale: Investigating the Relationship between Computer Self-Efficacy, Gender and Experience with Computers. *Journal of Educational Computing Research*, *26*(2), 133–153. <https://doi.org/10.2190/JGJR-0KVL-HRF7-GCNV>
- Cheong, Y. F., Pajares, F. & Oberman, P. S. (2004). Motivation and Academic Help-Seeking in High School Computer Science. *Computer Science Education*, *14*(1), 3–19. <https://doi.org/10.1076/csed.14.1.3.23501>
- Collins, L. M., Schafer, J. L. & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351. <https://doi.org/10.1037/1082-989X.6.4.330>
- Conderman, G. & Hedin, L. (2011). Cue Cards: A Self-Regulatory Strategy for Students With Learning Disabilities. *Intervention in School and Clinic*, *46*(3), 165–173. <https://doi.org/10.1177/1053451210378745>
- Conderman, G. & Hedin, L. R. (2015). Using Cue Cards in Inclusive Middle School Classrooms. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *88*(5), 155–160. <https://doi.org/10.1080/00098655.2015.1061971>
- Conradty, C. & Bogner, F. (2016). Hypertext or Textbook: Effects on Motivation and Gain in Knowledge. *Education Sciences*, *6*(4), 29. <https://doi.org/10.3390/educsci6030029>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. <https://doi.org/10.1037/0021-9010.78.1.98>

- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.  
<https://doi.org/10.1017/S0140525X01003922>
- Crocker, S. & Maratos, F. A. (2011). Visual Processing Speeds in Children. *Child Development Research*, 2011, 1–6. <https://doi.org/10.1155/2011/450178>
- CSDT (o. D.). *Intrinsic Motivation Inventory (IMI)*. <https://selfdeterminationtheory.org/intrinsic-motivation-inventory/> (letzter Zugriff: 22.09.2023)
- Czubatinski, L., Hornung, G. & Nieß, C. (2020). Quantitative Analysen mit dem Smartphone oder Tablet zur Einführung des Konzentrationsbegriffs—Ein Beispiel für wissenschaftspropädeutisches Arbeiten in der SEK I. In *Chemie konkret* (Bd. 27, Nummer 6, S. 295–299).
- Davids, M. R., Chikte, U. M. E. & Halperin, M. L. (2014). Effect of improving the usability of an e-learning resource: A randomized trial. *Advances in Physiology Education*, 38(2), 155–160. <https://doi.org/10.1152/advan.00119.2013>
- Davids, R., Halperin, M. & Chikte, U. (2015). Optimising cognitive load and usability to improve the impact of e-learning in medical education. *African Journal of Health Professions Education*, 7(2), 147–152. <https://doi.org/10.7196/AJHPE.659>
- Dávila-Acedo, M. A., Airado-Rodríguez, D., Cañada-Cañada, F. & Sánchez-Martín, J. (2021). Detailed Emotional Profile of Secondary Education Students Toward Learning Physics and Chemistry. *Frontiers in Psychology*, 12, 1–11.
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319–340.  
<https://doi.org/10.2307/249008>
- Davison, K., Malmberg, L.-E. & Sylva, K. (2022). Academic help-seeking interactions in the classroom: A microlongitudinal study. *British Journal of Educational Psychology*, 93(1), 33–55. <https://doi.org/10.1111/bjep.12538>
- De Bruin, A. B. H. & van Merriënboer, J. J. G. (2017). Bridging Cognitive Load and Self-Regulated Learning Research: A complementary approach to contemporary issues in educational research. *Learning and Instruction*, 51, 1–9.  
<https://doi.org/10.1016/j.learninstruc.2017.06.001>

- De Giorgi, G., Woolston, W. G. & Pellizzari, M. (2012). Class Size and Class Heterogeneity. *Journal of the European Economic Association*, 10(4), 795–830.
- De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, 38(2), 105–134.  
<https://doi.org/10.1007/s11251-009-9110-0>
- Deci, E. (1971). The Effects of Externally Mediated Rewards on Intrinsic Motivation. *Journal of Personality and Social Psychology*, 18(1), 105–115.  
<https://doi.org/10.1037/h0030644>
- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Springer New York.
- Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39(2), 223–238.  
<https://doi.org/10.25656/01:11173>
- Deci, E. L. & Ryan, R. M. (2000). The „What“ and „Why“ of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, 11(4), 227–268. [https://doi.org/10.1207/S15327965PLI1104\\_01](https://doi.org/10.1207/S15327965PLI1104_01)
- Decristan, J., Fauth, B., Kunter, M., Büttner, G. & Klieme, E. (2017). The interplay between class heterogeneity and teaching quality in primary school. *International Journal of Educational Research*, 86, 109–121.  
<https://doi.org/10.1016/j.ijer.2017.09.004>
- Delorme, A., Rousselet, G. A., Macé, M. J.-M. & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Cognitive Brain Research*, 19(2), 103–113. <https://doi.org/10.1016/j.cog-brainres.2003.11.010>
- Demirtas, H. & Schafer, J. L. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine*, 22(16), 2553–2575. <https://doi.org/10.1002/sim.1475>
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>

- Deunk, M. I., Smale-Jacobse, A. E., de Boer, H., Doolaard, S. & Bosker, R. J. (2018). Effective differentiation Practices: A systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education. *Educational Research Review*, 24, 31–54. <https://doi.org/10.1016/j.edu-rev.2018.02.002>
- DIN (2019). *Ergonomie der Mensch-System-Interaktion Teil 210: Menschzentrierte Gestaltung interaktiver Systeme (ISO 9241-210: 2019)*.
- Djalev, L. & Bogdanov, S. (2019). Age and Gender Differences in Evaluating the Pedagogical Usability of E-Learning Materials. *English Studies at NBU*, 5(2). <https://doi.org/10.33919/esnbu.19.2.0>
- Dong, A., Jong, M. S.-Y. & King, R. B. (2020). How Does Prior Knowledge Influence Learning Engagement? The Mediating Roles of Cognitive Load and Help-Seeking. *Frontiers in Psychology*, 11, 1–10. <https://doi.org/10.3389/fpsyg.2020.591203>
- Eitel, A., Endres, T. & Renkl, A. (2020). Self-management as a Bridge Between Cognitive Load and Self-regulated Learning: The Illustrative Case of Seductive Details. *Educational Psychology Review*, 32, 1073–1087. <https://doi.org/10.1007/s10648-020-09559-5>
- Engeln, K. (2004). *Schülerlabors. Authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken*. (Bd. 36). Logos.
- Evertson, C. M., Sanford, J. P. & Emmer, E. T. (1981). Effects of Class Heterogeneity in Junior High School. *American Educational Research Journal*, 18(2), 219–232. <https://doi.org/10.2307/1162383>
- Eysink, T. H. S., de Jong, T., Berthold, K., Kolloffel, B., Opfermann, M. & Wouters, P. (2009). Learner Performance in Multimedia Learning Arrangements: An Analysis Across Instructional Approaches. *American Educational Research Journal*, 46(4), 1107–1149. <https://doi.org/10.3102/0002831209340235>
- Fabian, K. & Topping, K. J. (2019). Putting “mobile” into mathematics: Results of a randomised controlled trial. *Contemporary Educational Psychology*, 59, 1–12. <https://doi.org/10.1016/j.cedpsych.2019.101783>

- Fachdidaktik Chemie TUK (2021, 10. Mai). *HyperDocSystems*. <https://didaktik.chemie.uni-kl.de> (letzter Zugriff: 22.09.2023)
- Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z. & Yuan, Q. (2018). The Big-Fish-Little-Pond Effect on Academic Self-Concept: A Meta-Analysis. *Frontiers in Psychology*, 9, 1–11. <https://doi.org/10.3389/fpsyg.2018.01569>
- Feise, R. J. (2002). Do multiple outcome measures require p-value adjustment? *BMC Medical Research Methodology*, 2(1), 1–4. <https://doi.org/10.1186/1471-2288-2-8>
- Fernandez, A., Insfran, E. & Abrahão, S. (2011). Usability evaluation methods for the web: A systematic mapping study. *Information & Software Technology*, 53(8), 789–817. <https://doi.org/10.1016/j.infsof.2011.02.007>
- Field, A. P., Miles, J. & Field, Z. (2012). *Discovering statistics using R*. Sage.
- Fitting, N. (2018). *Untersuchung des Nutzungsverhaltens einer heterogenen Lerngruppe bezüglich digitaler Hilfen und Vertiefungsaufgaben im Chemieunterricht*. TU Kaiserslautern.
- Fitting, N. & Hornung, G. (2022). Verwendung digitaler Lernhilfen als ein neuer Zugang zur Differenzierung im Unterricht Wie nutzen Schüler\*innen digitale Lernhilfen? In M. Veber, P. Gollub, S. Greiten & T. Schkade (Hrsg.), *Umgang mit Heterogenität und Inklusion. Chancen und Herausforderungen für die schupraktische Professionalisierung* (S. 172–189). Julius Klinkhardt.
- Fitting, N., Hornung, G. & Hemm, H. (2018). „Fördern und Fordern“ mit digitalisierten multimedialen Lernzugängen (Poster). 35. Fortbildungs- und Vortragstagung der Fachgruppe Chemieunterricht, Karlsruhe.
- Fitting, N., Ulber, R., Czubatinski, L. & Hornung, G. (2023). Evaluation digitaler Arbeitsblätter im Chemieunterricht in Hinblick auf Usability und Interesse. In J. Roth, M. Baum, K. Eilerts, G. Hornung & T. Trefzger (Hrsg.), *Die Zukunft des MINT-Lernens – Band 2: Digitale Tools und Methoden für das Lehren und Lernen* (S. 93–108). Springer. [https://doi.org/10.1007/978-3-662-66133-8\\_7](https://doi.org/10.1007/978-3-662-66133-8_7)

- Franke-Braun, G., Schmidt-Weigand, F., Stäudel, L. & Wodzinski, R. (2008). Aufgaben mit gestuften Lernhilfen – ein besonderes Aufgabenformat zur kognitiven Aktivierung der Schülerinnen und Schüler und zur Intensivierung der sachbezogenen Kommunikation. In Kasseler Forschergruppe (Hrsg.), *Forschergruppe Empirische Bildungsforschung* (Bd. 2, S. 27–42).
- Gao, M., Kortum, P. & Oswald, F. (2018). Psychometric Evaluation of the USE (Usefulness, Satisfaction, and Ease of use) Questionnaire for Reliability and Validity. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62, 1414–1418. <https://doi.org/10.1177/1541931218621322>
- Gibbens, B. (2019). Measuring Student Motivation in an Introductory Biology Class. *The American Biology Teacher*, 81(1), 20–26. <https://doi.org/10.1525/abt.2019.81.1.20>
- González-Cutre, D., Romero-Elías, M., Jiménez-Loaisa, A., Beltrán-Carrillo, V. J. & Hagger, M. S. (2020). Testing the need for novelty as a candidate need in basic psychological needs theory. *Motivation and Emotion*, 44, 295–314. <https://doi.org/10.1007/s11031-019-09812-7>
- González-Cutre, D. & Sicilia, Á. (2019). The importance of novelty satisfaction for multiple positive outcomes in physical education. *European Physical Education Review*, 25(3), 859–875. <https://doi.org/10.1177/1356336X18783980>
- González-Cutre, D., Sicilia, Á., Sierra, A. C., Ferriz, R. & Hagger, M. S. (2016). Understanding the need for novelty from the perspective of self-determination theory. *Personality and Individual Differences*, 102, 159–169. <https://doi.org/10.1016/j.paid.2016.06.036>
- Goswami, A. & Dutta, S. (2016). Gender Differences in Technology Usage—A Literature Review. *Open Journal of Business and Management*, 4(1), 51–59. <https://doi.org/10.4236/ojbm.2016.41006>
- Graham, J. W. (2009). Missing Data Analysis: Making It Work in the Real World. *Annual Review of Psychology*, 60(1), 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>



- Graham, J. W., Hofer, S. M., Donaldson, S. I., MacKinnon, D. P. & Schafer, J. L. (1997). Analysis with missing data in prevention research. In K. J. Bryant, M. Windle & S. G. West (Hrsg.), *The science of prevention: Methodological advances from alcohol and substance abuse research*. (S. 325–366). American Psychological Association. <https://doi.org/10.1037/10222-010>
- Graham, J. W., Hofer, S. M. & MacKinnon, D. P. (1996). Maximizing the Usefulness of Data Obtained with Planned Missing Value Patterns: An Application of Maximum Likelihood Procedures. *Multivariate Behavioral Research*, 31(2), 197–218. [https://doi.org/10.1207/s15327906mbr3102\\_3](https://doi.org/10.1207/s15327906mbr3102_3)
- Großmann, C. & Woest, V. (2015). Unstrukturierte Lernhilfen – Erste Ergebnisse einer Längsschnittstudie. *Heterogenität und Diversität – Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Bremen 2014*, 576–578.
- Großmann, C. & Woest, V. (2014). *Unstrukturierte Lernhilfen – Erste Ergebnisse einer Längsschnittstudie (Poster)*. Heterogenität und Diversität – Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Bremen 2014, Bremen.
- Hackbarth, G., Grover, V. & Yi, M. Y. (2003). Computer playfulness and anxiety: Positive and negative mediators of the system experience effect on perceived ease of use. *Information & Management*, 40(3), 221–232. [https://doi.org/10.1016/S0378-7206\(02\)00006-X](https://doi.org/10.1016/S0378-7206(02)00006-X)
- Hamers, P., Bekel-Kastrup, H., Kleinert, S. I., Haunhorst, D. & Wilde, M. (2022). Schüler\*innen lernen selbstständig das Modellieren mit Exponentialfunktionen. *Die Materialwerkstatt. Zeitschrift für Konzepte und Arbeitsmaterialien für Lehrer\*innenbildung und Unterricht.*, 73–78. <https://doi.org/10.11576/DIMAWE-5474>
- Hänze, M., Schmidt-Weigand, F. & Blum, S. (2007). Mit gestuften Lernhilfen im naturwissenschaftlichen Unterricht selbstständig lernen und arbeiten. In K. Rabenstein & S. Reh (Hrsg.), *Kooperatives und selbstständiges Arbeiten von Schülern: Zur Qualitätsentwicklung von Unterricht* (S. 197–208). VS Verlag für Sozialwissenschaften. [https://doi.org/10.1007/978-3-531-90418-4\\_10](https://doi.org/10.1007/978-3-531-90418-4_10)

- Hänze, M., Schmidt-Weigand, F. & Stäudel, L. (2010). Gestufte Lernhilfen. In S. Boller & R. Lau (Hrsg.), *Individuelle Förderung durch innere Differenzierung. Ein Praxishandbuch für Lehrerinnen und Lehrer der Sekundarstufe II* (S. 63–73). Beltz.
- Hariyanto, D., Triyono, Moch. B. & Köhler, T. (2020). Usability evaluation of personalized adaptive e-learning system using USE questionnaire. *Knowledge Management & E-Learning: An International Journal*, 12(1), 85–105.  
<https://doi.org/10.34105/j.kmel.2020.12.005>
- Harrison, R., Flood, D. & Duce, D. (2013). Usability of mobile applications: Literature review and rationale for a new usability model. *Journal of Interaction Science*, 1(1), 1–16. <https://doi.org/10.1186/2194-0827-1-1>
- Hassenzahl, M. (2001). The Effect of Perceived Hedonic Quality on Product Appealingsness. *International Journal of Human-Computer Interaction*, 13(4), 481–499.  
[https://doi.org/10.1207/S15327590IJHC1304\\_07](https://doi.org/10.1207/S15327590IJHC1304_07)
- Hassenzahl, M. (2004). The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, 19(4), 319–349.  
[https://doi.org/10.1207/s15327051hci1904\\_2](https://doi.org/10.1207/s15327051hci1904_2)
- Hassenzahl, M., Diefenbach, S. & Göritz, A. (2010). Needs, affect, and interactive products – Facets of user experience. *Interacting with Computers*, 22(5), 353–362.  
<https://doi.org/10.1016/j.intcom.2010.04.002>
- Hassenzahl, M., Platz, A., Burmester, M. & Lehner, K. (2000). Hedonic and ergonomic quality aspect determine a software's appeal. *CHI '00: Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 2, 201–208.  
<https://doi.org/10.1145/332040.332432>
- Hassenzahl, M. & Sandweg, N. (2004). From mental effort to perceived usability: Transforming experiences into summary assessments. *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, 1283–1286.  
<https://doi.org/10.1145/985921.986044>
- Hassenzahl, M. & Tractinsky, N. (2006). User experience—A research agenda. *Behaviour and Information Technology*, 25, 91–97.  
<https://doi.org/10.1080/01449290500330331>

- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* (Reprinted). Routledge.
- Heckhausen, H. (1977a). Achievement motivation and its constructs: A cognitive model. *Motivation and Emotion*, 1(4), 283–329. <https://doi.org/10.1007/BF00992538>
- Heckhausen, H. (1977b). Motivation: Kognitionspsychologische Aufspaltung eines summarischen Konstrukts. In T. Hermann, P. R. Hofstätter, H. P. Hubert & F. E. Weinert (Hrsg.), *Handbuch psychologischer Grundbegriffe* (S. 175–189). Kösel.
- Heckhausen, H. & Rheinberg, F. (1980). Lernmotivation im Unterricht, erneut betrachtet. *Unterrichtswissenschaft*, 8, 7–47.
- Heckhausen, J. (2000). Evolutionary Perspectives on Human Motivation. *American Behavioral Scientist*, 43(6), 1015–1029. <https://doi.org/10.1177/00027640021955739>
- Heckhausen, J. & Heckhausen, H. (2018). Motivation und Handeln: Einführung und Überblick. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (S. 1–11). Springer. [https://doi.org/10.1007/978-3-662-53927-9\\_1](https://doi.org/10.1007/978-3-662-53927-9_1)
- Helbig, M., Edelstein, B., Fickermann, D. & Zink, C. (2022). *Aufholen nach Corona? Maßnahmen der Länder im Kontext des Aktionsprogramms von Bund und Ländern*. Waxmann Verlag GmbH. <https://doi.org/10.31244/9783830996033>
- Hellman, M. & Rönkkö, K. (2008). Controlling user experience through policing in the software development process. In S. M. Abrahão, E. L.-C. Law, J. Stage, K. Hornbæk & N. J. Juzgado (Hrsg.), *Proceedings of the first workshop on the interplay between usability evaluation and software development, pisa, italy, september 24, 2008* (Bd. 407). CEUR-WS.org.
- Helmke, A. (2003). *Unterrichtsqualität. Erfassen. Bewerten. Verbessern*. Kallmeyer.
- Hendra, S.Kom., M.T. & Arifin, S.Kom., M.M., Y. (2018). Web-based Usability Measurement for Student Grading Information System. *Procedia Computer Science*, 135, 238–247. <https://doi.org/10.1016/j.procs.2018.08.171>
- Herbst, M., Fürtbauer, E.-M. & Strahl, A. (2016). Interesse an Physik in Salzburg. *Phy-Did B - Didaktik Der Physik - Beiträge Zur DPG-Frühjahrstagung*, 1–7.

- Hidi, S. & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist, 41*(2), 111–127.  
[https://doi.org/10.1207/s15326985ep4102\\_4](https://doi.org/10.1207/s15326985ep4102_4)
- Higgins, K., Huscroft-D'Angelo, J. & Crawford, L. (2019). Effects of Technology in Mathematics on Achievement, Motivation, and Attitude: A Meta-Analysis. *Journal of Educational Computing Research, 57*(2), 283–319.  
<https://doi.org/10.1177/0735633117748416>
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I. & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education, 153*, 1–25.  
<https://doi.org/10.1016/j.compedu.2020.103897>
- Hollender, N., Hofmann, C., Deneke, M. & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior, 26*(6), 1278–1288. <https://doi.org/10.1016/j.chb.2010.05.031>
- Holstermann, N. & Bögeholz, S. (2007). Interesse von Jungen und Mädchen an naturwissenschaftlichen Themen am Ende der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften, 13*, 71–86.
- Holzinger, A. (2005). Usability engineering methods for software developers. *Communications of the ACM, 48*(1), 71–74. <https://doi.org/10.1145/1039539.1039541>
- Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *Int. J. Hum. Comput. Stud., 64*, 79–103.  
<https://doi.org/10.1016/j.ijhcs.2005.06.002>
- Hornbæk, K. & Lai-Chong Law, E. (2007). Meta-analysis of correlations among usability measures. *CHI '07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 617–626. <https://doi.org/10.1145/1240624.1240722>
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.  
<https://doi.org/10.1080/10705519909540118>

- Huang, Y.-M., Liang, T.-H. & Chiu, C.-H. (2013). Gender Differences in the Reading of E-books: Investigating Children's Attitudes, Reading Behaviors and Outcomes. *Journal of Educational Technology & Society*, 16(4), 97–110.
- Huang, Z. (2019). Developing Usability Heuristics for Recommendation Systems Within the Mobile Context. In A. Marcus & W. Wang (Hrsg.), *Design, User Experience, and Usability. Practice and Case Studies* (S. 143–151). Springer International Publishing. [https://doi.org/10.1007/978-3-030-23535-2\\_10](https://doi.org/10.1007/978-3-030-23535-2_10)
- HyperDocs Admin. (2019, 12. August). *What is a HyperDoc?* <https://hyperdocs.co/blog/posts/what-is-a-hyperdoc> (letzter Zugriff: 22.09.2023)
- Ibili, E. & Billingham, M. (2019). Assessing the Relationship between Cognitive Load and the Usability of a Mobile Augmented Reality Tutorial System: A Study of Gender Effects. *International Journal of Assessment Tools in Education*, 6(3). <https://doi.org/10.21449/ijate.594749>
- ISO (2018). *ISO 9241-11:2018(en), Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts*. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en> (letzter Zugriff: 22.09.2023)
- Jansen, M., Schroeders, U. & Stanat, P. (2013). Motivationale Schülermerkmale in Mathematik und den Naturwissenschaften. In H. A. Pant, P. Stanat, U. Schroeders, A. Roppelt, T. Siegle & C. Pöhlmann (Hrsg.), *IQB-Ländervergleich 2012. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I* (S. 348–365). Waxmann.
- Jeno, L. M., Adachi, P. J. C., Grytnes, J.-A., Vandvik, V. & Deci, E. L. (2019b). The effects of m-learning on motivation, achievement and well-being: A Self-Determination Theory approach: Effect of m-learning on motivation and well-being. *British Journal of Educational Technology*, 50(2), 669–683. <https://doi.org/10.1111/bjet.12657>
- Jeno, L. M., Grytnes, J.-A. & Vandvik, V. (2017). The effect of a mobile-application tool on biology students' motivation and achievement in species identification: A Self-Determination Theory perspective. *Computers & Education*, 107, 1–12. <https://doi.org/10.1016/j.compedu.2016.12.011>

- Jeno, L. M., Vandvik, V., Eliassen, S. & Grytnes, J.-A. (2019a). Testing the novelty effect of an m-learning tool on internalization and achievement: A Self-Determination Theory approach. *Computers & Education*, 128, 398–413.  
<https://doi.org/10.1016/j.compedu.2018.10.008>
- JJuni (2016, 12. Januar). *Waage*. <https://pixabay.com/de/vectors/die-skala-gewicht-masse-skala-diät-1133910/> (letzter Zugriff: 22.09.2023)
- Johansen, S. (1980). The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika*, 67(1), 85–92.  
<https://doi.org/10.1093/biomet/67.1.85>
- Joller-Graf, K. (2010). Binnendifferenziert unterrichten. In A. Buholzer & A. Kummer Wyss (Hrsg.), *Alle gleich—Alle unterschiedlich! Zum Umgang mit Heterogenität in Schule und Unterricht* (S. 122–151). Klett.
- Jung, Y. J., Zimmerman, H. T. & Land, S. M. (2019). Emerging and developing situational interest during children’s tablet-mediated biology learning activities at a nature center. *Science Education*, 103(4), 900–922.  
<https://doi.org/10.1002/sce.21514>
- Kandler, M. (2002). *Lernsoftware aus der Sicht von Schülerinnen und Schülern: Interesse- und motivationsfördernde Aspekte*. Peter Lang Verlag.
- Karabenick, S. A. & Berger, J.-L. (2013). Help seeking as a self-regulated learning strategy. In H. Bembenuity, T. J. Cleary & A. Kitsantas (Hrsg.), *Applications of self-regulated learning across diverse disciplines: A tribute to Barry J. Zimmerman* (S. 237–261). IAP Information Age Publishing.
- Karabenick, S. A. & Dembo, M. H. (2011). Understanding and facilitating self-regulated help seeking. *New Directions for Teaching and Learning*, 2011(126), 33–43.  
<https://doi.org/10.1002/tl.442>
- Karabenick, S. A. & Knapp, J. R. (1988). Effects of Computer Privacy on Help-Seeking. *Journal of Applied Social Psychology*, 18(6), 461–472.  
<https://doi.org/10.1111/j.1559-1816.1988.tb00029.x>

- Karapanos, M., Becker, C. & Christophel, E. (2018). Die Bedeutung der Usability für das Lernen mit digitalen Medien. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 36–57.  
<https://doi.org/10.21240/mpaed/00/2018.05.09.X>
- Kessels, U. & Steinmayr, R. (2013). Macho-man in school: Toward the role of gender role self-concepts and help seeking in school performance. *Learning and Individual Differences*, 23, 234–240. <https://doi.org/10.1016/j.lindif.2012.09.013>
- Khedr, A. E., Hana, M. A. & Shollar, D. Z. (2011). Investigating Learners' Acceptance of e-Learning Courses Using Adopted Technology Acceptance Model. *Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE)*.
- Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, 41(2), 75–86. [https://doi.org/10.1207/s15326985ep4102\\_1](https://doi.org/10.1207/s15326985ep4102_1)
- Kleinert, S. I., Besa, K.-S. & Wilde, M. (2022). Der Einsatz gestufter Lernhilfen als Unterstützung für Lernende im Kontext des biologischen Experimentierens: Einfluss auf die Schüler\*innen-Motivation im Naturwissenschafts- und Mathematikunterricht. *Zeitschrift für Didaktik der Biologie (ZDB) - Biologie Lehren und Lernen*, 26, 1–18. <https://doi.org/10.11576/zdb-5205>
- Kleinert, S. I., Isaak, R. C., Textor, A. & Wilde, M. (2021). Die Nutzung gestufter Lernhilfen zur Unterstützung des Experimentierprozesses im Biologieunterricht – eine qualitative Studie. *Zeitschrift für Didaktik der Naturwissenschaften*, 27(1), 59–71. <https://doi.org/10.1007/s40573-021-00126-1>
- Klepsch, M., Schmitz, F. & Seufert, T. (2017). Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in Psychology*, 8, 1–18. <https://doi.org/10.3389/fpsyg.2017.01997>
- Klepsch, M. & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instructional Science*, 48(1), 45–77. <https://doi.org/10.1007/s11251-020-09502-9>

- Koć-Januchta, M. M., Schönborn, K. J., Tibell, L. A. E., Chaudhri, V. K. & Heller, H. C. (2020). Engaging With Biology by Asking Questions: Investigating Students' Interaction and Learning With an Artificial Intelligence-Enriched Textbook. *Journal of Educational Computing Research*, 58(6), 1190–1224.  
<https://doi.org/10.1177/0735633120921581>
- Koć-Januchta, M., Schönborn, K., Roehrig, C., Chaudhri, V., Tibell, L. & Heller, C. (2022). “Connecting concepts helps put main ideas together”: Cognitive load and usability in learning biology with an AI-enriched textbook. *International Journal of Educational Technology in Higher Education*, 19(11), 1–22.  
<https://doi.org/10.1186/s41239-021-00317-3>
- Kopinska, M. (2020). Beyond the novelty effect: EFL learners' attitudes towards ICT use in the classroom. *Hungarian Educational Research Journal*, 10(1), 1–15.  
<https://doi.org/10.1556/063.2020.00001>
- Korner, M., Urban-Woldron, H. & Hopf, M. (2012). Entwicklung eines Messinstrumentes zur Motivation. In S. Bernholt (Hrsg.), *Konzepte fachdidaktischer Strukturierung für den Unterricht* (S. 98–100). LIT.
- Kortum, P. & Johnson, M. (2013). The Relationship Between Levels of User Experience with a Product and Perceived System Usability. *Proceedings of the Human Factors and Ergonomics Society 57th Annual Meeting*, 57 (1), 197–201.  
<https://doi.org/10.1177/1541931213571044>
- Kousa, P. & Aksela, M. (2019). The needs for successful chemistry teaching in diverse classes: Teachers' beliefs and practices. *LUMAT: International Journal on Math, Science and Technology Education*, 7(1). <https://doi.org/10.31129/LUMAT.7.1.390>
- Krapp, A. (1992a). Interesse, Lernen und Leistung. Neue Forschungsansätze in der Pädagogischen Psychologie. *Zeitschrift für Pädagogik*, 38(5), 747–770.  
<https://doi.org/10.25656/01:13977>
- Krapp, A. (1992b). Das Interessenkonstrukt. Bestimmungsmerkmale der Interessenhandlung und des individuellen Interesses aus der Sicht einer Person-Gegenstands-Konzeption. In A. Krapp & M. Prenzel (Hrsg.): *Interesse, Lernen, Leistung* (S. 297–329). Verlag Aschendorf.



- Krapp, A. (1998). Entwicklung und Förderung von Interessen im Unterricht. *Psychologie in Erziehung und Unterricht*, 45(3), 186–203.
- Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, 12, 383–409. [https://doi.org/10.1016/S0959-4752\(01\)00011-1](https://doi.org/10.1016/S0959-4752(01)00011-1)
- Krapp, A. (2007). An educational–psychological conceptualisation of interest. *International Journal for Educational and Vocational Guidance*, 7(1), 5–21. <https://doi.org/10.1007/s10775-007-9113-9>
- Kultusministerkonferenz, KMK (2004a). *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss*. [https://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Bildungsstandards-Chemie.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Bildungsstandards-Chemie.pdf) (letzter Zugriff: 22.09.2023)
- Kultusministerkonferenz, KMK (2004b). *Einheitliche Prüfungsanforderungen in der Abiturprüfung Chemie*. [https://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/1989/1989\\_12\\_01-EPA-Chemie.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1989/1989_12_01-EPA-Chemie.pdf) (letzter Zugriff: 22.09.2023)
- Kultusministerkonferenz, KMK (2019). *Standards für die Lehrerbildung: Bildungswissenschaften*. [https://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/2004/2004\\_12\\_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf](https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf) (letzter Zugriff: 22.09.2023)
- Kurosu, M. & Kashimura, K. (1995). Apparent usability vs. inherent usability: Experimental analysis on the determinants of the apparent usability. *Conference Companion on Human Factors in Computing Systems*, 292–293. <https://doi.org/10.1145/223355.223680>
- Laugwitz, B., Held, T. & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger (Hrsg.), *HCI and Usability for Education and Work* (S. 63–76). Springer. [https://doi.org/10.1007/978-3-540-89350-9\\_6](https://doi.org/10.1007/978-3-540-89350-9_6)
- Lazowski, R. A. & Hulleman, C. S. (2016). Motivation Interventions in Education: A Meta-Analytic Review. *Review of Educational Research*, 86(2), 602–640.
- LearningApps (o. D.). *LearningApps.org—Interaktive und multimediale Lernbausteine*. LearningApps. <https://learningapps.org/> (letzter Zugriff: 22.09.2023)

- Lee, S. & Ha, T. (2019). Changes in perceived usability and aesthetics with repetitive use in the first use session. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 29(6), 517–528. <https://doi.org/10.1002/hfm.20814>
- Lee, T. T. & Osman, K. (2012). Interactive Multimedia Module in the Learning of Electrochemistry: Effects on Students' Understanding and Motivation. *Procedia - Social and Behavioral Sciences*, 46, 1323–1327. <https://doi.org/10.1016/j.sbspro.2012.05.295>
- Leisen, J. (1999). *Methoden-Handbuch deutschsprachiger Fachunterricht (DFU)*. Varus.
- Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T. & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, 45(4), 1058–1072. <https://doi.org/10.3758/s13428-013-0334-1>
- Letzel, V. & Otto, J. (2019). Binnendifferenzierung und deren konkrete Umsetzung in der Schulpraxis – eine qualitative Studie. *Zeitschrift für Bildungsforschung*, 9(3), 375–393. <https://doi.org/10.1007/s35834-019-00256-0>
- Lewis, J. (1993). Sample Sizes for Usability Studies: Additional Considerations. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 36(2). <https://doi.org/10.1177/001872089403600215>
- Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human–Computer Interaction*, 7(1), 57–78. <https://doi.org/10.1080/10447319509526110>
- Lewis, J. R. (2014). Usability: Lessons Learned ... and Yet to Be Learned. *International Journal of Human–Computer Interaction*, 30(9), 663–684. <https://doi.org/10.1080/10447318.2014.930311>
- Lewis, J. R. & Sauro, J. (2021). Usability and User Experience: Design and Evaluation. In G. Salvendy & W. Karwowski (Hrsg.), *Handbook Of Human Factors and Ergonomics* (5. Auflage, S. 972–1015). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119636113.ch38>
- LF Projects, LLC. (2022). *Laminas Project—Enterprise PHP Framework*. <https://getlaminas.org/> (letzter Zugriff: 22.09.2023)

- Lin Gao, X. W., Murillo, B. & Paz, F. (2019). A Systematic Literature Review of Usability Evaluation Guidelines on Mobile Educational Games for Primary School Students. In A. Marcus & W. Wang (Hrsg.), *Design, User Experience, and Usability. Practice and Case Studies* (S. 172–182). Springer International Publishing. [https://doi.org/10.1007/978-3-030-23535-2\\_13](https://doi.org/10.1007/978-3-030-23535-2_13)
- Lipowsky, F. & Lotz, M. (2015). Ist Individualisierung der Königsweg zum erfolgreichen Lernen? In G. Mehlhorn, K. Schöppe & F. Schulz (Hrsg.), *Begabungen entwickeln & Kreativität fördern* (S. 155–219). kopaed.
- Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, 83(404), 1198–1202. <https://doi.org/10.2307/2290157>
- Little, R. J. A. & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2. Auflage). John Wiley & Sons, Inc.
- Liu, M., Olmanson, J., Horton, L. & Toprac, P. (2011). Motivational Multimedia: Examining Students' Learning and Motivation as They Use a Multimedia Enriched Learning Environment. *Online Submission*, 1–14.
- Loomatix (2019). *Color Grab (color detection)*. <https://play.google.com/store/apps/details?id=com.loomatix.colorgrab&hl=de> (letzter Zugriff: 22.09.2023)
- Loth, G. (2023). Teilhabe am Argumentieren und Darstellen im digital-gestützten Mathematikunterricht. In IDMI-Primar Goethe-Universität Frankfurt (Hrsg.), *Beiträge zum Mathematikunterricht 2022. 56. Jahrestagung der Gesellschaft für Didaktik der Mathematik* (S. 315–318). WTM.
- Loth, G. & Döhrmann, M. (2022). Teilhabe am digital-gestützten Mathematikunterricht: Entwicklung und Evaluation einer Lernumgebung zur Förderung der Datenkompetenz. *Digitales Lernen in Distanz und Präsenz : Herbsttagung 2021 des Arbeitskreises Mathematikunterricht und digitale Werkzeuge in der Gesellschaft für Didaktik der Mathematik*, 81–88. <https://doi.org/10.17185/dupublico/76036>
- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B. & d'Apollonia, S. (1996). Within-Class Grouping: A Meta-Analysis. *Review of Educational Research*, 66(4), 423–458. <https://doi.org/10.2307/1170650>

- Lund, A. M. (2001). Measuring Usability with the USE Questionnaire. *Usability Interface*, 8(2), 3–6.
- Lynch, R. & Dembo, M. (2004). The Relationship Between Self-Regulation and Online Learning in a Blended Learning Context. *The International Review of Research in Open and Distributed Learning*, 5(2), 1–16.  
<https://doi.org/10.19173/irrodl.v5i2.189>
- Maaz, K., Pfafferott, M. & Stichler, M. (2021). *Lehren aus der Pandemie: Gleiche Chancen für alle Kinder und Jugendlichen sichern*. <https://library.fes.de/pdf-files/a-p-b/17249.pdf> (letzter Zugriff: 22.09.2023)
- Mair, P. & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, 52(2), 464–488. <https://doi.org/10.3758/s13428-019-01246-w>
- Mair, P., Wilcox, R. & Patil, I. (2022). Package ‘WRS2’. *A Collection of Robust Statistical Methods*. <https://cran.r-project.org/web/packages/WRS2/WRS2.pdf> (letzter Zugriff: 22.09.2023)
- MariaDB Foundation (2022). *MariaDB Foundation*. MariaDB.Org. <https://mariadb.org/> (letzter Zugriff: 22.09.2023)
- Markland, D. & Hardy, L. (1997). On the Factorial and Construct Validity of the Intrinsic Motivation Inventory: Conceptual and Operational Concerns. *Research Quarterly for Exercise and Sport*, 68(1), 20–32.  
<https://doi.org/10.1080/02701367.1997.10608863>
- Martín-Arbós, S., Castarlenas, E. & Dueñas, J.-M. (2021). Help-Seeking in an Academic Context: A Systematic Review. *Sustainability*, 13(8).  
<https://doi.org/10.3390/su13084460>
- Maxwell, S. E., Delaney, H. D. & Kelley, K. (2017). *Designing Experiments and Analyzing Data: A Model Comparison Perspective*, 3. Auflage. Routledge.
- Mayer, R. E. (2014b). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (2. Auflage, S. 52–71). Cambridge University Press.

- Mayer, R. E. (2014a). Introduction to Multimedia Learning. In R. E. Mayer (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (2. Auflage, S. 1–24). Cambridge University Press.
- Mayer, R. E., Mautone, P. & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology*, 94(1), 171–185. <https://doi.org/10.1037/0022-0663.94.1.171>
- Mayer, R. E. & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1), 43–52. [https://doi.org/10.1207/S15326985EP3801\\_6](https://doi.org/10.1207/S15326985EP3801_6)
- McAuley, E., Duncan, T. & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, 60(1), 48–58. <https://doi.org/10.1080/02701367.1989.10607413>
- McClelland, D. C., Koestner, R. & Weinberger, J. (1989). How Do Self-Attributed and Implicit Motives Differ? *Psychological Review*, 96(4), 690–702. <https://doi.org/10.1037/0033-295X.96.4.690>
- McNeish, D. (2017). Missing data methods for arbitrary missingness with small samples. *Journal of Applied Statistics*, 44(1), 24–39. <https://doi.org/10.1080/02664763.2016.1158246>
- Medienpädagogischer Forschungsverbund Südwest, mpfs (2022). *JIM-Studie 2022*. [https://www.mpfs.de/fileadmin/files/Studien/JIM/2022/JIM\\_2022\\_Web\\_final.pdf](https://www.mpfs.de/fileadmin/files/Studien/JIM/2022/JIM_2022_Web_final.pdf) (letzter Zugriff: 22.09.2023)
- Meiselwitz, G. & Sadera, W. (2008). Investigating the Connection between Usability and Learning Outcomes in Online Learning Environments. *Journal of Online Learning and Teaching*, 4(2), 234–242.
- Mézes, C. (2016). *Zur Motivation beim Experimentieren im Physikunterricht*. Pädagogischen Hochschule Schwäbisch Gmünd.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97. <https://doi.org/10.1037/h0043158>

- Ministerium für Bildung, Wissenschaft und Weiterbildung Rheinland-Pfalz, MBBWK (1998). *Lehrplan Chemie Sekundarstufe II*. [https://static.bildung-rp.de/lehrplaene/naturwissenschaften/Chemie\\_LP\\_SekII.pdf](https://static.bildung-rp.de/lehrplaene/naturwissenschaften/Chemie_LP_SekII.pdf) (letzter Zugriff: 22.09.2023)
- Ministerium für Bildung, Wissenschaft, Weiterbildung und Kultur Rheinland-Pfalz, MBBWK (2014). *Lehrpläne für die naturwissenschaftlichen Fächer für die weiterführenden Schulen in Rheinland-Pfalz*. [https://naturwissenschaften.bildung-rp.de/fileadmin/user\\_upload/naturwissenschaften.bildung-rp.de/\\_Alt/pdf-download/Lehrplaene\\_Endversion\\_online.pdf](https://naturwissenschaften.bildung-rp.de/fileadmin/user_upload/naturwissenschaften.bildung-rp.de/_Alt/pdf-download/Lehrplaene_Endversion_online.pdf) (letzter Zugriff: 22.09.2023)
- Mitchell, M. (1992). *Situational Interest: Its Multifaceted Structure in the Secondary Mathematics Classroom*. 1–26. <https://files.eric.ed.gov/fulltext/ED353157.pdf> (letzter Zugriff: 22.09.2023)
- Mitchell, M. (1997). *Situational Interest in the Statistics Classroom*. 1–40. <https://files.eric.ed.gov/fulltext/ED413345.pdf> (letzter Zugriff: 22.09.2023)
- Möhring, W. & Schlütz, D. (2019). *Die Befragung in der Medien- und Kommunikationswissenschaft: Eine praxisorientierte Einführung*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-25865-8>
- Molich, R. (2018). Are usability evaluations reproducible? *Interactions*, 25(6), 82–85. <https://doi.org/10.1145/3278154>
- Monteiro, V., Mata, L. & Peixoto, F. (2015). Intrinsic Motivation Inventory: Psychometric Properties in the Context of First Language and Mathematics Learning. *Psicologia: Reflexão e Crítica*, 28(3), 434–443. <https://doi.org/10.1590/1678-7153.201528302>
- Moos, D. C. & Marroquin, E. (2010). Multimedia, hypermedia, and hypertext: Motivation considered and reconsidered. *Computers in Human Behavior*, 26(3), 265–276. <https://doi.org/10.1016/j.chb.2009.11.004>
- Moser, C. (2012). *User Experience Design*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-13363-3>
- Nelson-Le Gall, S. (1985). Help-Seeking Behavior in Learning. *Review of Research in Education*, 12, 55–90. <https://doi.org/10.2307/1167146>

- Nerlich, B. (2013). *Vergleichende Studie der L1-Lesekompetenz bilingual unterrichteter Schüler der Klassenstufen 7 und 10 an schleswig-holsteinischen Gymnasien*. Christian-Albrechts-Universität zu Kiel.
- Newman, R. S. (1990). Children's Help-Seeking in the Classroom: The Role of Motivational Factors and Attitudes. *Journal of Educational Psychology*, 82(1), 71–80.
- Newman, R. S. & Schwager, M. T. (1995). Students' Help Seeking During Problem Solving: Effects of Grade, Goal, and Prior Achievement. *American Educational Research Journal*, 32(2), 352–376.  
<https://doi.org/10.3102/00028312032002352>
- Nieding, G., Ohler, P. & Rey, G. D. (2015). *Lernen mit Medien*. UTB.
- Nielsen, C. M., Overgaard, M., Pedersen, M. B., Stage, J. & Stenild, S. (2006). It's worth the hassle!: The added value of evaluating the usability of mobile systems in the field. *Proceedings of the 4th Nordic Conference on Human-Computer Interaction Changing Roles - NordiCHI '06*, 272–280.  
<https://doi.org/10.1145/1182475.1182504>
- Nielsen, J. (1993). *Usability Engineering*. Ap Professional.
- Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 152–158. <https://doi.org/10.1145/191666.191729>
- Nieß, C., Czubatinski, L. & Hornung, G. (2020). Die Konzentration eines Farbstoffs bestimmen. *Unterricht Chemie*, 177 / 178, 32–34.
- Nieß, C., Czubatinski, L. & Hornung, G. (o. D.). *Die photometrische Untersuchung von Azofarbstoffen in Erfrischungsgetränken*. Verlaufsplan und Material. Unveröffentlicht.
- Nokelainen, P. (2004). *Conceptual Definition of the Technical and Pedagogical Usability Criteria for Digital Learning Material*. 4249–4254.
- Nokelainen, P. (2006). An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students. *Educational Technology & Society*, 9(2), 178–197.

- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education, 15*(5), 625–632. <https://doi.org/10.1007/s10459-010-9222-y>
- NumFOCUS (2021). *pandas—Python Data Analysis Library*. <https://pandas.pydata.org/> (letzter Zugriff: 22.09.2023)
- Ochsen, S., Bernholt, S., Bernholt, A. & Parchmann, I. (2021). Eine Mikroanalyse von Chemieunterricht – Einsatz und Perzeption von Triggern für situationales Interesse. *Zeitschrift für Didaktik der Naturwissenschaften, 27*(1), 1–15. <https://doi.org/10.1007/s40573-020-00122-x>
- OmarMedinaFilms (2016, 6. Juni). *Kleinkind*. <https://pixabay.com/de/photos/alegre-spielzeug-kind-teddy-1434720/> (letzter Zugriff: 22.09.2023)
- Ong, C.-S. & Lai, J.-Y. (2006). Gender differences in perceptions and relationships among dominants of e-learning acceptance. *Computers in Human Behavior, 22*(5), 816–829. <https://doi.org/10.1016/j.chb.2004.03.006>
- OpenClipart-Vectors (2016, 1. April). *Cocktail*. <https://pixabay.com/de/vectors/getränk-cocktail-trinken-glas-1300288/> (letzter Zugriff: 22.09.2023)
- OpenClipart-Vectors (2017, 1. Februar). *Personen*. <https://pixabay.com/de/vectors/figuren-komische-charaktere-2029374/> (letzter Zugriff: 22.09.2023)
- OpenJS Foundation (2022). *jQuery*. <https://jquery.com/> (letzter Zugriff: 22.09.2023)
- Orfanou, K., Tselios, N. & Katsanos, C. (2015). Perceived usability evaluation of learning management systems: Empirical evaluation of the System Usability Scale. *The International Review of Research in Open and Distributed Learning, 16*(2), 227–246. <https://doi.org/10.19173/irrodl.v16i2.1955>
- Osborne, J., Simon, S. & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049–1079. <https://doi.org/10.1080/0950069032000032199>
- Osman, K. & Lee, T. T. (2014). Impact of Interactive Multimedia Module with Pedagogical Agents on Students’ Understanding and Motivation in the Learning of Electrochemistry. *International Journal of Science and Mathematics Education, 12*(2), 395–421. <https://doi.org/10.1007/s10763-013-9407-y>



- Ostrow, K. S. & Heffernan, N. T. (2018). Testing the Validity and Reliability of Intrinsic Motivation Inventory Subscales Within ASSISTments. In C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren & B. du Boulay (Hrsg.), *Artificial Intelligence in Education* (S. 381–394). Springer International Publishing.  
[https://doi.org/10.1007/978-3-319-93843-1\\_28](https://doi.org/10.1007/978-3-319-93843-1_28)
- Paas, F. (1992). Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive-Load Approach. *Journal of Educational Psychology*, 84(4), 429–434. <https://doi.org/10.1037/0022-0663.84.4.429>
- Paas, F., Sweller, J. & Mayer, R. E. (2014). Implications of cognitive load theory for multimedia learning. In *The Cambridge handbook of multimedia learning* (2. Auflage, S. 27–42). Cambridge University Press.  
<https://doi.org/10.1017/CBO9781139547369.004>
- Padlet (2022). *Padlet*. Padlet. <https://de.padlet.com/> (letzter Zugriff: 22.09.2023)
- Pagefact (2014, 9. April). *Handschuhe*. <https://pixabay.com/de/photos/handschuhe-gummi-schwarz-hygiene-319838/> (letzter Zugriff: 22.09.2023)
- Palmer, D. (2004). Research Report: Situational interest and the attitudes towards science of primary teacher education students. *International Journal of Science Education*, 26(7), 895–908. <https://doi.org/10.1080/0950069032000177262>
- Palmer, D., Dixon, J. & Archer, J. (2016). Using Situational Interest to Enhance Individual Interest and Science-Related Behaviours. *Research in Science Education*, 47(4), 731–753. <https://doi.org/10.1007/s11165-016-9526-x>
- Palmer, D. H. (2009). Student interest generated during an inquiry skills lesson. *Journal of Research in Science Teaching*, 46(2), 147–165.  
<https://doi.org/10.1002/tea.20263>
- Pashler, H., McDaniel, M., Rohrer, D. & Bjork, R. (2008). Learning Styles: Concepts and Evidence. *Psychological Science in the Public Interest*, 9(3), 105–119.  
<https://doi.org/10.1111/j.1539-6053.2009.01038.x>

- Patall, E. A., Steingut, R. R., Freeman, J. L., Pituch, K. A. & Vasquez, A. C. (2018). Gender disparities in students' motivational experiences in high school science classrooms. *Science Education*, 102(5), 951–977.  
<https://doi.org/10.1002/sce.21461>
- Patall, E., Cooper, H. & Robinson, J. (2008). The Effects of Choice on Intrinsic Motivation and Related Outcomes: A Meta-Analysis of Research Findings. *Psychological Bulletin*, 134(2), 270–300. <https://doi.org/10.1037/0033-2909.134.2.270>
- Pawek, C. (2009). *Schülerlabore als interessefördernde außerschulische Lernumgebungen für Schülerinnen und Schüler aus der Mittel- und Oberstufe*. Christian-Albrechts-Universität zu Kiel.
- Perforce (2022). *Home—Zend Framework*. <https://framework.zend.com/> (letzter Zugriff: 22.09.2023)
- Peterson, L. & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58(2), 193–198.  
<https://doi.org/10.1037/h0049234>
- phpMyAdmin contributors (2022). *Bringing MySQL to the web*. phpMyAdmin.  
<https://www.phpmyadmin.net/> (letzter Zugriff: 22.09.2023)
- Pouw, W. T. J. L., Eielts, C., van Gog, T., Zwaan, R. A. & Paas, F. (2016). Does (Non-)Meaningful Sensori-Motor Engagement Promote Learning With Animated Physical Systems?: Motor Engagement and Learning From Animation. *Mind, Brain, and Education*, 10(2), 91–104. <https://doi.org/10.1111/mbe.12105>
- Pozas, M. & Schneider, C. (2019). Shedding Light on the Convoluted Terrain of Differentiated Instruction (DI): Proposal of a DI Taxonomy for the Heterogeneous Classroom. *Open Education Studies*, 1, 73–90. <https://doi.org/10.1515/edu-2019-0005>
- Prenzel, M., Krapp, A. & Schiefele, H. (1986). Grundzüge einer pädagogischen Interestheorie. *Zeitschrift für Pädagogik*, 32(2), 163–173.  
<https://doi.org/10.25656/01:14383>
- Quesenbery, W. (2003). Dimensions of Usability: Defining the Conversation, Driving the Process. *Proceedings of the UPA 2003 Conference*, 1–8. <http://www.wqusability.com/articles/5es-upa2003.pdf> (letzter Zugriff: 22.09.2023)

- Ran, H., Kasli, M. & Secada, W. G. (2021). A Meta-Analysis on Computer Technology Intervention Effects on Mathematics Achievement for Low-Performing Students in K-12 Classrooms. *Journal of Educational Computing Research*, 59(1), 119–153. <https://doi.org/10.1177/0735633120952063>
- Ran, H., Kim, N. J. & Secada, W. G. (2022). A meta-analysis on the effects of technology's functions and roles on students' mathematics achievement in K-12 classrooms. *Journal of Computer Assisted Learning*, 38(1), 258–284. <https://doi.org/10.1111/jcal.12611>
- Rasch, D. & Guiard, V. (2004). The robustness of parametric statistical methods. *Psychology Science*, 46, 175–208.
- Redecker, C. & Punie, Y. (2017). *European Framework for the Digital Competence of Educators: DigCompEdu*. Publications Office of the European Union.
- Reisinger, C.-M., Svecnik, E. & Schwetz, H. (2012). *Fehlende Werte und keine Normalverteilung?: Tipps und Abhilfen für das quantitativ orientierte Forschen*. Facultas.WUV.
- Renninger, K. A. & Bachrach, J. E. (2015). Studying Triggers for Interest and Engagement Using Observational Methods. *Educational Psychologist*, 50(1), 58–69. <https://doi.org/10.1080/00461520.2014.999920>
- Renninger, K., Bachrach, J. & Hidi, S. (2018). Triggering and maintaining interest in early phases of interest development. *Learning, Culture and Social Interaction*, 23, 1–17. <https://doi.org/10.1016/j.lcsi.2018.11.007>
- Rheinberg, F. & Engeser, S. (2018). Intrinsische Motivation und Flow-Erleben. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (S. 423–450). Springer. [https://doi.org/10.1007/978-3-662-53927-9\\_14](https://doi.org/10.1007/978-3-662-53927-9_14)
- Richter, M. & Flückiger, M. D. (2016). *Usability und UX kompakt*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-49828-6>
- Roegger (2014, 2. Dezember). *Sprühdose*. <https://pixabay.com/de/photos/spray-nacht-dose-sprühen-sprühdose-549085/> (letzter Zugriff: 22.09.2023)
- Root, R. W. & Draper, S. (1983). Questionnaires as a software evaluation tool. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '83*, 83–87. <https://doi.org/10.1145/800045.801586>

- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rothermund, K. & Eder, A. (2011). *Allgemeine Psychologie: Motivation und Emotion*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-93420-4>
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.
- Rubin, M. (2021). When to adjust alpha during multiple testing: A consideration of disjunction, conjunction, and individual testing. *Synthese*, 199, 10969–11000. <https://doi.org/10.1007/s11229-021-03276-4>
- Ruf, T. & Ploetzner, R. (2014). One click away is too far! How the presentation of cognitive learning aids influences their use in multimedia learning environments. *Computers in Human Behavior*, 38, 229–239. <https://doi.org/10.1016/j.chb.2014.06.002>
- Ryan, A., Anderman, L. & Midgley, C. (1997). Social Goals, Academic Goals, and Avoiding Seeking Help in the Classroom. *Journal of Early Adolescence*, 17(2), 152–171. <https://doi.org/10.1177/0272431697017002003>
- Ryan, A. M., Shim, S. S., Lampkins-uThando, S. A., Kiefer, S. M. & Thompson, G. N. (2009). Do gender differences in help avoidance vary by ethnicity? An examination of African American and European American students during early adolescence. *Developmental Psychology*, 45(4), 1152–1163. <https://doi.org/10.1037/a0013916>
- Ryan, A. M. & Shin, H. (2011). Help-seeking tendencies during early adolescence: An examination of motivational correlates and consequences for achievement. *Learning and Instruction*, 21(2), 247–256. <https://doi.org/10.1016/j.learninstruc.2010.07.003>
- Ryan, A., Patrick, H. & Shim, S. (2005). Differential Profiles of Students Identified by Their Teacher as Having Avoidant, Appropriate, or Dependent Help-Seeking Tendencies in the Classroom. *Journal of Educational Psychology*, 97(2), 275–285. <https://doi.org/10.1037/0022-0663.97.2.275>
- Ryan, A. & Pintrich, P. R. (1997). „Should I Ask for Help?“ The Role of Motivation and Attitudes in Adolescents’ Help Seeking in Math Class. *Journal of Educational Psychology*, 89(2), 326–341. <https://doi.org/10.1037//0022-0663.89.2.329>

- Ryan, R. M. (1995). Psychological Needs and the Facilitation of Integrative Processes. *Journal of Personality*, 63(3), 397–427. <https://doi.org/10.1111/j.1467-6494.1995.tb00501.x>
- Ryan, R. M. & Deci, E. L. (2000). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Ryan, R. M. & Deci, E. L. (2017). *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. Guilford Publications.
- Ryan, R. M. & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, 1–11. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Sanchez, E. & Garcia-Rodicio, H. (2008). The use of modality in the design of verbal aids in computer-based learning environments. *Interacting with Computers*, 20(6), 545–561. <https://doi.org/10.1016/j.intcom.2008.08.001>
- Sary, S. P., Tarigan, S. & Situmorang, M. (2018). Development of Innovative Learning Material with Multimedia to Increase Student Achievement and Motivation in Teaching Acid Base Titration. *Advances in Social Science, Education and Humanities Research*, 200, 422–425. <https://doi.org/10.2991/aisteel-18.2018.91>
- Sax, L. J., Kanny, M. A., Riggers-Piehl, T. A., Whang, H. & Paulson, L. N. (2015). “But I’m Not Good at Math”: The Changing Salience of Mathematical Self-Concept in Shaping Women’s and Men’s STEM Aspirations. *Research in Higher Education*, 56(8), 813–842. <https://doi.org/10.1007/s11162-015-9375-x>
- Schafer, J. L. (2021). *Analysis of Incomplete Multivariate Data under a Normal Model*.
- Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>

- Schenke, K., Lam, A. C., Conley, A. M. & Karabenick, S. A. (2015). Adolescents' help seeking in mathematics classrooms: Relations between achievement and perceived classroom environmental influences over one school year. *Contemporary Educational Psychology, 41*, 133–146. <https://doi.org/10.1016/j.cedpsych.2015.01.003>
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research, 8*(2), 23–74.
- Schmidt, M., Kafka, J. X., Kothgassner, O. D., Hlavacs, H., Beutl, L. & Felnhofer, A. (2013). Why Does It Always Rain on Me? Influence of Gender and Environmental Factors on Usability, Technology Related Anxiety and Immersion in Virtual Environments. In D. Reidsma, H. Katayose & A. Nijholt (Hrsg.), *Advances in Computer Entertainment* (S. 392–402). Springer International Publishing. [https://doi.org/10.1007/978-3-319-03161-3\\_29](https://doi.org/10.1007/978-3-319-03161-3_29)
- Schneider, S., Nebel, S., Beege, M. & Rey, G. D. (2018). The autonomy-enhancing effects of choice on cognitive load, motivation and learning with digital media. *Learning and Instruction, 58*, 161–172. <https://doi.org/10.1016/j.learninstruc.2018.06.006>
- Schnell, R., Hill, P. B. & Esser, E. (2018). Methoden der empirischen Sozialforschung. In *Methoden der empirischen Sozialforschung*. De Gruyter Oldenbourg.
- Schulz, S. (2020). *Selbstreguliertes Lernen mit mobil nutzbaren Technologien*. Springer Fachmedien.
- Schumacher, P. & Morahan-Martin, J. (2001). Gender, Internet and computer attitudes and experiences. *Computers in Human Behavior, 17*(1), 95–110. [https://doi.org/10.1016/S0747-5632\(00\)00032-7](https://doi.org/10.1016/S0747-5632(00)00032-7)
- Seibert, J., Schmoll, I., Kay, C. W. M. & Huwer, J. (2020). Promoting Education for Sustainable Development with an Interactive Digital Learning Companion Students Use to Perform Collaborative Phosphorus Recovery Experiments and Reporting. *Journal of Chemical Education, 97*(11), 3992–4000. <https://doi.org/10.1021/acs.jchemed.0c00408>

- Shen, S.-T., Woolley, M. & Prior, S. (2006). Towards culture-centred design. *Interacting with Computers*, 18(4), 820–852. <https://doi.org/10.1016/j.int-com.2005.11.014>
- Silvia, P. J. (2005). What Is Interesting? Exploring the Appraisal Structure of Interest. *Emotion*, 5(1), 89–102. <https://doi.org/10.1037/1528-3542.5.1.89>
- Simon, B., Treiblmaier, H. & Neumann, G. (2008). Elektronische Lernumgebungen in Bildungseinrichtungen: Eine Diskussion kritischer Erfolgsfaktoren. *Zeitschrift für Betriebswirtschaft*, 78(7), 715–733. <https://doi.org/10.1007/s11573-008-0042-y>
- Skitterphoto (2015, 17. April). *Flamme*. <https://pixabay.com/de/photos/flamme-feuer-inferno-orange-726268/> (letzter Zugriff: 22.09.2023)
- Smale-Jacobse, A. E., Meijer, A., Helms-Lorenz, M. & Maulana, R. (2019). Differentiated Instruction in Secondary Education: A Systematic Review of Research Evidence. *Frontiers in Psychology*, 10, 1–23.
- Sonderegger, A., Zbinden, G., Uebelbacher, A. & Sauer, J. (2012). The influence of product aesthetics and usability over the course of time: A longitudinal field experiment. *Ergonomics*, 55(7), 713–730. <https://doi.org/10.1080/00140139.2012.672658>
- Soriano, E. (2019). *Screen Flashlight*. [https://play.google.com/store/apps/details?id=com.eduardo\\_rsor.apps.linternapantalla&hl=de](https://play.google.com/store/apps/details?id=com.eduardo_rsor.apps.linternapantalla&hl=de) (letzter Zugriff: 22.09.2023)
- Spannagel, C. & Bescherer, C. (2009). Computerbezogene Selbstwirksamkeitserwartung in Lehrveranstaltungen mit Computernutzung. *Notes on Educational Informatics — Section A: Concepts and Techniques*, 5(1), 23–43.
- Spool, J. & Schroeder, W. (2001). Testing web sites: Five users is nowhere near enough. *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, 285–286. <https://doi.org/10.1145/634067.634236>
- Squires, D. & Preece, J. (1999). Predicting quality in educational software: Evaluating for learning, usability and the synergy between them. *Interacting with Computers*, 11(5), 467–483. [https://doi.org/10.1016/S0953-5438\(98\)00063-0](https://doi.org/10.1016/S0953-5438(98)00063-0)

- Ssemugabi, S. & de Villiers, R. (2007). A Comparative Study of Two Usability Evaluation Methods Using a Web-Based E-Learning Application. *Proceedings of the 2007 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries*, 132–142. <https://doi.org/10.1145/1292491.1292507>
- Stäudel, L., Franke-Braun, G. & Schmidt-Weigand, F. (2007). Komplexität erhalten - auch in heterogenen Lerngruppen: Aufgaben mit gestuften Lernhilfen. *CHEMKON*, 14(3), 115–122. <https://doi.org/10.1002/ckon.200710058>
- Stiller, C. & Wilde, M. (2021). Einfluss gestufter Lernhilfen als Unterstützungsmaßnahme beim Experimentieren auf den Lernerfolg im Biologieunterricht. *Zeitschrift für Erziehungswissenschaft*, 24(3), 743–763. <https://doi.org/10.1007/s11618-021-01017-4>
- Stoa, R. & Chu, A. (2023). An argument for implementing and testing novelty in the classroom. *Scholarship of Teaching and Learning in Psychology*, 9(1), 88–95. <https://doi.org/10.1037/stl0000223>
- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4)
- Sweller, J. (2016). Cognitive Load Theory, Evolutionary Educational Psychology, and Instructional Design. In D. C. Geary & D. B. Berch (Hrsg.), *Evolutionary Perspectives on Child Development and Education* (S. 291–306). Springer International Publishing. [https://doi.org/10.1007/978-3-319-29986-0\\_12](https://doi.org/10.1007/978-3-319-29986-0_12)
- Sweller, J., van Merriënboer, J. J. G. & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- The Apache Software Foundation (2022). *Welcome! - The Apache HTTP Server Project*. <https://httpd.apache.org/>
- The Linux Foundation (2022). *Linux Foundation*. <https://www.linuxfoundation.org> (letzter Zugriff: 22.09.2023)
- Thorpe, S., Fize, D. & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520–522. <https://doi.org/10.1038/381520a0>



- Tolle, H., Hafis, M., Afif, A. & Arai, K. (2020). Perceived Usability of Educational Chemistry Game Gathered via CSUQ Usability Testing in Indonesian High School Students. *International Journal of Advanced Computer Science and Applications*, 11(3), 715–724. <https://doi.org/10.14569/IJACSA.2020.0110389>
- Tossavainen, T. & Hirsto, L. (2017). Tablet computers and Finnish primary and lower secondary students' motivation in mathematics. *Nordic Research in Mathematics Education. Papers of NORMA 17*, 59–68.
- Tractinsky, N. (1997). Aesthetics and apparent usability: Empirically assessing cultural and methodological issues. *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 115–122.  
<https://doi.org/10.1145/258549.258626>
- Tractinsky, N., Katz, A. S. & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, 13(2), 127–145. [https://doi.org/10.1016/S0953-5438\(00\)00031-X](https://doi.org/10.1016/S0953-5438(00)00031-X)
- Trautmann, M. & Wischer, B. (2011). *Heterogenität in der Schule: Eine kritische Einführung* (1. Auflage). VS-Verlag für Sozialwissenschaften.
- Tricot, A. & Sweller, J. (2014). Domain-Specific Knowledge and Why Teaching Generic Skills Does Not Work. *Educational Psychology Review*, 26(2), 265–283.  
<https://doi.org/10.1007/s10648-013-9243-1>
- Trumper, R. (2006a). Factors Affecting Junior High School Students' Interest in Biology. *Journal of Science Education and Technology*, 17(1), 31–48.
- Trumper, R. (2006b). Factors Affecting Junior High School Students' Interest in Physics. *Journal of Science Education and Technology*, 15(1), 47–58.  
<https://doi.org/10.1007/s10956-006-0355-6>
- Tselios, N., Avouris, N., Dimitracopoulou, A. & Daskalaki, S. (2001). Evaluation of Distance-Learning Environments: Impact of Usability on Student Performance. *International Journal of Educational Telecommunications*, 7(4), 355–378.
- Tsigilis, N. & Theodosiou, A. (2003). Temporal Stability of the Intrinsic Motivation Inventory. *Perceptual and Motor Skills*, 97(1), 271–280.  
<https://doi.org/10.2466/pms.2003.97.1.271>

- Tullis, T. & Albert, B. (2013). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics* (2. Auflage). Elsevier/Morgan Kaufmann.
- Urhahne, D., Jeschke, J., Krombaß, A. & Harms, U. (2004). Die Validierung von Fragebogenerhebungen zum Interesse an Tieren und Pflanzen durch computergestützte Messdaten: The Validation of Questionnaire Data on Interest in Animals and Plants with Log Files. *Zeitschrift für Pädagogische Psychologie*, 18(3/4), 213–219. <https://doi.org/10.1024/1010-0652.18.34.213>
- Stevens Institute (2022). *WRS* (Version 40) [R]. University of Southern California. <https://dornsife.usc.edu/assets/sites/239/docs/Rallfun-v40.txt> (letzter Zugriff: 22.09.2023)
- Vaessen, B. E., Prins, F. J. & Jeurig, J. (2014). University students' achievement goals and help-seeking strategies in an intelligent tutoring system. *Computers & Education*, 72, 196–208. <https://doi.org/10.1016/j.compedu.2013.11.001>
- Vallerand, R. J. (1997). Toward A Hierarchical Model of Intrinsic and Extrinsic Motivation. In M. P. Zanna (Hrsg.), *Advances in Experimental Social Psychology* (Bd. 29, S. 271–360). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60019-2](https://doi.org/10.1016/S0065-2601(08)60019-2)
- Vallerand, R. J. & Ratelle, C. F. (2002). Intrinsic and extrinsic motivation: A hierarchical model. In E. L. Deci & R. M. Ryan (Hrsg.), *Handbook of self-determination research* (S. 37–63). University of Rochester Press.
- Van Loon, A.-M., Ros, A. & Martens, R. (2012). Motivated learning with digital learning tasks: What about autonomy and structure? *Educational Technology Research and Development*, 60(6), 1015–1032. <https://doi.org/10.1007/s11423-012-9267-0>
- Van Merriënboer, J. J. G. & Kester, L. (2014). The four-component instructional design model: Multimedia principles in environments for complex learning. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (2. Auflage, S. 104–148). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.007>
- Van Nuland, S. E., Eagleson, R. & Rogers, K. A. (2017). Educational software usability: Artifact or Design? *Anatomical Sciences Education*, 10(2), 190–199. <https://doi.org/10.1002/ase.1636>

- Venkatesh, V. (1999). Creation of Favorable User Perceptions: Exploring the Role of Intrinsic Motivation. *MIS Quarterly*, 23(2), 239–260.  
<https://doi.org/10.2307/249753>
- Venkatesh, V. (2000). Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. *Information Systems Research*, 11(4), 342–365.  
<https://doi.org/10.1287/isre.11.4.342.11872>
- Villacorta, P., J. (2017). The welchADF Package for Robust Hypothesis Testing in Unbalanced Multivariate Mixed Models with Heteroscedastic and Non-normal Data. *The R Journal*, 9(2), 309–328. <https://doi.org/10.32614/RJ-2017-049>
- Virzi, R. A. (1992). Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? *Human Factors*, 34(4), 457–468.  
<https://doi.org/10.1177/001872089203400407>
- Vock, M. & Gronostaj, A. (2017). *Umgang mit Heterogenität in Schule und Unterricht* (2. Auflage). Friedrich-Ebert-Stiftung.
- Wachsmuth, C. (2019). *Computerbasiertes Lernen mit Aufmerksamkeitsdefizit: Unterstützung des selbstregulierten Lernens durch metakognitive Prompts*.
- Wallace, S. & Yu, H.-C. (2009). The effect of culture on usability: Comparing the perceptions and performance of Taiwanese and North American MP3 player users. *Journal of Usability Studies*, 4(3), 136–146.
- Wang, Y. (2021). Gaps Between Continuous Measurement Methods: A Longitudinal Study of Perceived Usability. *Interacting with Computers*, 33(3), 223–237.  
<https://doi.org/10.1093/iwc/iwab023>
- Weichbroth, P. (2020). Usability of Mobile Applications: A Systematic Literature Study. *IEEE Access*, 8, 55563–55577. <https://doi.org/10.1109/ACCESS.2020.2981892>
- Wenning, N. (2007). Heterogenität in Schule und Unterricht. Handlungsansätze zum pädagogischen Umgang mit Vielfalt. In S. Boller, E. Rosowski & T. Stroot (Hrsg.), *Heterogenität in Schule und Unterricht. Handlungsansätze zum pädagogischen Umgang mit Vielfalt*. (S. 21–31). Beltz.

- Wheaton, B., Muthen, B., Alwin, D. F. & Summers, G. F. (1977). Assessing Reliability and Stability in Panel Models. *Sociological Methodology*, 8, 84–136.  
<https://doi.org/10.2307/270754>
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, 66(5), 297–333.
- Wikipedia Foundation Inc. (2008). *Global harmonisiertes System zur Einstufung und Kennzeichnung von Chemikalien*. [https://de.wikipedia.org/wiki/Global\\_harmonisiertes\\_System\\_zur\\_Einstufung\\_und\\_Kennzeichnung\\_von\\_Chemikalien](https://de.wikipedia.org/wiki/Global_harmonisiertes_System_zur_Einstufung_und_Kennzeichnung_von_Chemikalien) (letzter Zugriff: 22.09.2023)
- Wilcox, R. R. (2021). *Introduction to Robust Estimation and Hypothesis Testing* (5. Auflage). Elsevier.
- Wilde, M., Bätz, K., Kovaleva, A. & Urhahne, D. (2009). Überprüfung einer Kurzskala intrinsischer Motivation (KIM). *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 31–45.
- Wirtz, M. A. (2004). Über das Problem fehlender Werte: Wie der Einfluss fehlender Informationen auf Analyseergebnisse entdeckt und reduziert werden kann. *Die Rehabilitation*, 43(2), 109–115. <https://doi.org/10.1055/s-2003-814839>
- Withers, C (2021). *xlrd—Xlrd 2.0.1 documentation*. <https://xlrd.readthedocs.io/en/latest/> (letzter Zugriff: 22.09.2023)
- Wolff, G. & Wolff, K. (2016). Multimediale Hilfekarten. Einsatz von Smartphones im mathematisch-naturwissenschaftlichen Unterricht. In *MNU Journal* (Bd. 69, Nummer 6, S. 406–411).
- Wolters, C. A., Pintrich, P. R. & Karabenick, S. A. (2005). Assessing Academic Self-Regulated Learning. In K. A. Moore & L. H. Lippman (Hrsg.), *What Do Children Need to Flourish?* (Bd. 3, S. 251–270). Springer. [https://doi.org/10.1007/0-387-23823-9\\_16](https://doi.org/10.1007/0-387-23823-9_16)
- Wood, H. & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education*, 33(2), 153–169. [https://doi.org/10.1016/S0360-1315\(99\)00030-5](https://doi.org/10.1016/S0360-1315(99)00030-5)
- Woodworth, R. S. (1918). *Dynamic Psychology*. In *Dynamic Psychology*. Columbia University Press.

- Wößmann, L., Freundl, V., Grewenig, E., Lergetporer, P., Werner, K. & Zierow, L. (o. J.). Bildung in der Coronakrise: Wie haben die Schulkinder die Zeit der Schulschließungen verbracht, und welche Bildungsmaßnahmen befürworten die Deutschen? *ifo Schnelldienst*, 73(9), 25–39.
- Wouters, P., Paas, F. & van Merriënboer, J. J. G. (2009). Observational learning from animated models: Effects of modality and reflection on transfer. *Contemporary Educational Psychology*, 34, 1–8. <https://doi.org/10.1016/j.cedpsych.2008.03.001>
- Wrosch, C., Scheier, M. F., Miller, G. E., Schulz, R. & Carver, C. S. (2003). Adaptive Self-Regulation of Unattainable Goals: Goal Disengagement, Goal Reengagement, and Subjective Well-Being. *Personality and Social Psychology Bulletin*, 29(12), 1494–1508. <https://doi.org/10.1177/0146167203256921>
- Yang, Y. & Cao, L. (2013). Differential influences of achievement approach goals and intrinsic/extrinsic motivation on help-seeking in e-learning. *Knowledge Management & E-Learning: An International Journal*, 5(2), 153–169. <https://doi.org/10.34105/j.kmel.2013.05.011>
- Yuhui, W., Chunfu, L., Tian, L. & Qin, H. (2019). Longitudinal Study on Retrospective Assessment of Perceived Usability: A New Method and Perspectives. *Interacting with Computers*, 31(4), 393–412. <https://doi.org/10.1093/iwc/iwz026>
- Zaharias, P. & Poulymenakou, A. (2009). Developing a Usability Evaluation Method for e-Learning Applications: Beyond Functional Usability. *International Journal of Human-Computer Interaction*, 25(1), 75–98. <https://doi.org/10.1080/10447310802546716>
- Zeng, G., Hou, H. & Peng, K. (2016). Effect of Growth Mindset on School Engagement and Psychological Well-Being of Chinese Primary and Middle School Students: The Mediating Role of Resilience. *Frontiers in Psychology*, 7, 1–8. <https://doi.org/10.3389/fpsyg.2016.01873>

## 9. Anhang

### Abbildungsverzeichnis Anhang

<i>Abbildung 92: IMI Modell 1 - Alle fünf Faktoren Interesse / Spaß (INT), Druck (DRUCK), wahrgenommene Wahlfreiheit (WAHL), Kompetenzerleben (KOMP), Anstrengung (ANS).</i>	435
<i>Abbildung 93: IMI Modell 2 - Vier Faktoren Interesse / Spaß (INT), Druck (DRUCK), wahrgenommene Wahlfreiheit (WAHL), Kompetenzerleben (KOMP).</i>	436
<i>Abbildung 94: Usability Modell 1 - Alle vier Faktoren Nützlichkeit (NÜTZ), Zufriedenheit (ZUF), Einfachheit des Erlernens der Nutzung (EERL), Einfachheit der Nutzung (ENUTZ).</i>	436
<i>Abbildung 95: Usability Modell 2 - Drei Faktoren Nützlichkeit (NÜTZ), Zufriedenheit (ZUF), Einfachheit des Erlernens der Nutzung (EERL).</i>	437
<i>Abbildung 96: Usability Modell 3 - Drei Faktoren Nützlichkeit (NÜTZ), Zufriedenheit (ZUF), Einfachheit der Nutzung (ENUTZ).</i>	437
<i>Abbildung 97: Usability Modell 4 - Drei Faktoren Nützlichkeit (NÜTZ), Zufriedenheit (ZUF), Einfachheit der Nutzung und des Erlernens der Nutzung (EINF).</i>	438
<i>Abbildung 98: Usability Modell 5 - Alle vier Faktoren Nützlichkeit (NÜTZ), Zufriedenheit (ZUF), Einfachheit des Erlernens der Nutzung (EERL), Einfachheit der Nutzung (ENUTZ). Reduzierte Itemzahl.</i>	438
<i>Abbildung 99: Tabletbezogene Kompetenzen Modell 1 - Alle drei Faktoren Selbstwirksamkeit (WIRKSAMKEIT), Nützlichkeit beim Lernen (LERNEN), Angst vor Tablets (ANGST).</i>	439
<i>Abbildung 100: Tabletbezogene Kompetenzen Modell 2 - Zwei Faktoren Selbstwirksamkeit (WIRKSAMKEIT), Nützlichkeit beim Lernen (LERNEN).</i>	439
<i>Abbildung 101: Tabletbezogene Kompetenzen Modell 3 - Zwei Faktoren Selbstwirksamkeit (WIRKSAMKEIT), Nützlichkeit beim Lernen (LERNEN). Reduzierte Itemzahl.</i>	440
<i>Abbildung 102: Cognitive Load Modell 1 - Drei Faktoren Intrinsic Load, Extraneous Load, Germane Load.</i>	440
<i>Abbildung 103: Cognitive Load Modell 2 - Zwei Faktoren Intrinsic Load, Extraneous Load, Germane Load. Reduzierte Itemzahl.</i>	441
<i>Abbildung 104: Einführung in die Nutzung der Tablets.</i>	442
<i>Abbildung 105: Benutzeroberfläche von HyperDocSystems.</i>	443
<i>Abbildung 106: HyperDoc der 1. Stunde der Unterrichtsreihe (Mittelstufe). Bildquellen, die nicht mehr zur Verfügung stehen, wurden im Folgenden durch einen Platzhalter ersetzt. Textfelder verkleinert. Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008).</i>	444

<i>Abbildung 107: HyperDoc der 2. Stunde der Unterrichtsreihe (Mittelstufe). Verändert nach Czubatinski et al. (2020). Kleinkind nach OmarMedinaFilms (2016).....</i>	<i>445</i>
<i>Abbildung 108: HyperDoc der 3. Stunde der Unterrichtsreihe (Mittelstufe). Verändert nach Czubatinski et al. (2020) .....</i>	<i>446</i>
<i>Abbildung 109: HyperDoc der 4. Stunde der Unterrichtsreihe (Mittelstufe). Verändert nach Czubatinski et al. (2020) .....</i>	<i>447</i>
<i>Abbildung 110: HyperDoc der 1. Stunde der Unterrichtsreihe (Oberstufe). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008) ..</i>	<i>448</i>
<i>Abbildung 111: HyperDoc der 2. Stunde der Unterrichtsreihe (Oberstufe). Verändert nach Nieß et al. (o. D.).....</i>	<i>449</i>
<i>Abbildung 112: HyperDoc der 3. Stunde der Unterrichtsreihe (Oberstufe). Verändert nach Nieß et al. (o. D.).....</i>	<i>450</i>
<i>Abbildung 113: HyperDoc der 4. Stunde der Unterrichtsreihe (Oberstufe). Verändert nach Nieß et al. (o. D.).....</i>	<i>451</i>
<i>Abbildung 114: Arbeitsblatt (Teil 1) der 1. Stunde der Unterrichtsreihe in der Vergleichsgruppe (Mittel- und Oberstufe). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008). .....</i>	<i>452</i>
<i>Abbildung 115: Arbeitsblatt (Teil 2) der 1. Stunde der Unterrichtsreihe in der Vergleichsgruppe (Mittel- und Oberstufe). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008). .....</i>	<i>453</i>
<i>Abbildung 116: Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 1. Stunde der Unterrichtsreihe der Vergleichsgruppe (Mittelstufe). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008). Handschuhe nach pagefact (2014). Sprühdose: Eigene Darstellung nach roegger (2014) und Skitterphoto (2015).....</i>	<i>461</i>
<i>Abbildung 117: Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 2. Stunde der Unterrichtsreihe der Vergleichsgruppe (Mittelstufe). Kleinkind nach OmarMedinaFilms (2016).....</i>	<i>462</i>
<i>Abbildung 118: : Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 3. Stunde der Unterrichtsreihe der Vergleichsgruppe (Mittelstufe). .....</i>	<i>463</i>
<i>Abbildung 119: : Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 4. Stunde der Unterrichtsreihe der Vergleichsgruppe (Mittelstufe). Hilfe Aufgabe 2: Eigene Darstellung nach OpenClipart-Vectors (2016, 2017) und JJuni (2016), Teilchenmodell Hilfe 3: Eigene Darstellung nach Czubatinski et al. (2020) ..</i>	<i>464</i>
<i>Abbildung 120: Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 1. Stunde der Unterrichtsreihe der Vergleichsgruppe (Oberstufe). Gas-flaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008). Handschuhe nach pagefact (2014). Sprühdose: Eigene Darstellung nach roegger (2014) und Skitterphoto (2015).....</i>	<i>465</i>
<i>Abbildung 121: Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 2. Stunde der Unterrichtsreihe der Vergleichsgruppe (Oberstufe).....</i>	<i>466</i>

<i>Abbildung 122: Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 3. Stunde der Unterrichtsreihe der Vergleichsgruppe (Oberstufe). Aufbau Photometer verändert nach Nieß et al. (o. D.)</i> .....	467
<i>Abbildung 123: Lernhilfen des Arbeitsblattes der 4. Stunde der Unterrichtsreihe der Vergleichsgruppe (Oberstufe). Photometer verändert nach Nieß et al. (o. D.). Hilfe 4. Eigene Darstellung nach OpenClipart-Vectors (2016, 2017) und Jjuni (2016)</i> .....	468
<i>Abbildung 124: Verlaufsplan zur ursprünglichen Unterrichtsreihe von Nieß et al. (2020). Unveröffentlichtes Dokument (Nieß et al., o. D.)</i> .....	471
<i>Abbildung 125: Arbeitsblatt der 1. Stunde der Unterrichtsreihe von Nieß et al. (2020). Unveröffentlichtes Dokument (Nieß et al., o. D.)</i> .....	472
<i>Abbildung 126: Arbeitsblatt der 2. Stunde der Unterrichtsreihe von Nieß et al. (2020). Unveröffentlichtes Dokument (Nieß et al., o. D.)</i> .....	473
<i>Abbildung 127: Arbeitsblatt der 3. Stunde der Unterrichtsreihe von Nieß et al. (2020). Unveröffentlichtes Dokument (Nieß et al., o. D.)</i> .....	474
<i>Abbildung 128: HyperDoc der 1. Stunde der Unterrichtsreihe für die Mittelstufe (2. Pilotierung). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008)</i> .....	475
<i>Abbildung 129: HyperDoc der 2. Stunde der Unterrichtsreihe für die Mittelstufe (2. Pilotierung). Verändert nach Czubatinski et al. (2020)</i> .....	476
<i>Abbildung 130: HyperDoc der 3. Stunde der Unterrichtsreihe für die Mittelstufe (2. Pilotierung). Verändert nach Czubatinski et al. (2020)</i> .....	477
<i>Abbildung 131: HyperDoc der 4. Stunde der Unterrichtsreihe für die Mittelstufe (2. Pilotierung). Verändert nach Czubatinski et al. (2020)</i> .....	478
<i>Abbildung 132: HyperDoc der 1. Stunde der Unterrichtsreihe für die Oberstufe (2. Pilotierung). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008)</i> .....	479
<i>Abbildung 133: HyperDoc der 2. Stunde der Unterrichtsreihe für die Oberstufe (2. Pilotierung). Verändert nach Nieß et al. (o. D.)</i> .....	480
<i>Abbildung 134: HyperDoc der 3. Stunde der Unterrichtsreihe für die Oberstufe (2. Pilotierung). Verändert nach Nieß et al. (o. D.)</i> .....	481
<i>Abbildung 135: HyperDoc der 4. Stunde der Unterrichtsreihe für die Oberstufe (2. Pilotierung). Verändert nach Nieß et al. (o. D.)</i> .....	482
<i>Abbildung 136: HyperDoc der 1. Stunde der Unterrichtsreihe für die Mittelstufe (3. Pilotierung). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008)</i> .....	483
<i>Abbildung 137: HyperDoc der 2. Stunde der Unterrichtsreihe für die Mittelstufe (3. Pilotierung). Verändert nach Czubatinski et al. (2020). Kleinkind nach OmarMedinaFilms (2016)</i> .....	484



<i>Abbildung 138: HyperDoc der 3. Stunde der Unterrichtsreihe für die Mittelstufe (3. Pilotierung). Verändert nach Czubatinski et al. (2020).</i>	485
<i>Abbildung 139: HyperDoc der 4. Stunde der Unterrichtsreihe für die Mittelstufe (3. Pilotierung). Verändert nach Czubatinski et al. (2020).</i>	486
<i>Abbildung 140: HyperDoc der 1. Stunde der Unterrichtsreihe für die Oberstufe (3. Pilotierung). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008).</i>	487
<i>Abbildung 141: HyperDoc der 2. Stunde der Unterrichtsreihe für die Oberstufe (3. Pilotierung). Verändert nach Nieß et al. (o. D.).</i>	488
<i>Abbildung 142: HyperDoc der 3. Stunde der Unterrichtsreihe für die Oberstufe (3. Pilotierung). Verändert nach Nieß et al. (o. D.).</i>	489
<i>Abbildung 143: HyperDoc der 4. Stunde der Unterrichtsreihe für die Oberstufe (3. Pilotierung). Verändert nach Nieß et al. (o. D.).</i>	490
<i>Abbildung 144: Screeplot der Hauptkomponentenanalyse.</i>	492

## **Tabellenverzeichnis Anhang**

<i>Tabelle 164: Betrachtungsdauer für die 1. Stunde in der Mittel- und Oberstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt. Kursive Hilfen wurden bei der Nutzungshäufigkeit nicht berücksichtigt.</i>	428
<i>Tabelle 165: Betrachtungsdauer für die 2. Stunde in der Mittelstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt. Kursive Hilfen wurden bei der Nutzungshäufigkeit nicht berücksichtigt.</i>	429
<i>Tabelle 166: Betrachtungsdauer für die 3. Stunde in der Mittelstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt. Kursive Hilfen wurden bei der Nutzungshäufigkeit nicht berücksichtigt.</i>	430
<i>Tabelle 167: Betrachtungsdauer für die 4. Stunde in der Mittelstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt.</i>	431
<i>Tabelle 168: Betrachtungsdauer für die 2. Stunde in der Oberstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt. Kursive Hilfen wurden bei der Nutzungshäufigkeit nicht berücksichtigt.</i>	432
<i>Tabelle 169: Betrachtungsdauer für die 3. Stunde in der Oberstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt.</i>	433
<i>Tabelle 170: Betrachtungsdauer für die 4. Stunde in der Oberstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt. Kursive Hilfen wurden bei der Nutzungshäufigkeit nicht berücksichtigt.</i>	434
<i>Tabelle 171: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der ersten Stunde in der Mittel- und Oberstufe.</i>	454

<i>Tabelle 172: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der zweiten Stunde in der Mittelstufe.....</i>	<i>455</i>
<i>Tabelle 173: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der dritten Stunde in der Mittelstufe.....</i>	<i>456</i>
<i>Tabelle 174: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der vierten Stunde in der Mittelstufe. ....</i>	<i>457</i>
<i>Tabelle 175: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der zweiten Stunde in der Oberstufe.....</i>	<i>458</i>
<i>Tabelle 176: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der dritten Stunde in der Oberstufe. ....</i>	<i>459</i>
<i>Tabelle 177: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der vierten Stunde in der Oberstufe.....</i>	<i>460</i>
<i>Tabelle 178: Korrelationsmatrix der Hauptkomponentenanalyse. ....</i>	<i>491</i>
<i>Tabelle 179: Anti-Image-Matrizen der Hauptkomponentenanalyse.....</i>	<i>491</i>
<i>Tabelle 180: Erklärte Gesamtvarianz der Hauptkomponentenanalyse. ....</i>	<i>491</i>
<i>Tabelle 181: Lösungswahrscheinlichkeiten der Aufgaben in den jeweiligen Stunden der 3. Pilotierung in der Mittelstufe. ....</i>	<i>493</i>
<i>Tabelle 182: Lösungswahrscheinlichkeiten der Aufgaben in den jeweiligen Stunden der 3. Pilotierung in der Oberstufe.....</i>	<i>494</i>
<i>Tabelle 183: Nutzungsangebot digitaler Endgeräte im häuslichen Umfeld. n = 198 (Gymnasium), n = 108 (Gesamtschule).....</i>	<i>495</i>
<i>Tabelle 184: Nutzungsangebot digitaler Endgeräte im schulischen Umfeld. Mittelstufe: n = 198 (Gymnasium), n = 108 (Gesamtschule).....</i>	<i>495</i>
<i>Tabelle 185: Nutzungsangebot digitaler Endgeräte im häuslichen Umfeld. Oberstufe: n = 101 (Gymnasium), n = 8 (Gesamtschule).....</i>	<i>495</i>
<i>Tabelle 186: Nutzungsangebot digitaler Endgeräte im schulischen Umfeld. Oberstufe: n = 101 (Gymnasium), n = 8 (Gesamtschule).....</i>	<i>495</i>
<i>Tabelle 187: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Gruppe“ (hier: Interventionsgruppe). ....</i>	<i>496</i>
<i>Tabelle 188: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Gruppe“ (hier: Vergleichsgruppe).....</i>	<i>496</i>

<i>Tabelle 189: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Geschlecht“ (hier: weiblich).....</i>	<i>497</i>
<i>Tabelle 190: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Geschlecht“ (hier: männlich). ....</i>	<i>497</i>
<i>Tabelle 191: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Gruppe“ (hier: Interventionsgruppe). ....</i>	<i>497</i>
<i>Tabelle 192: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Gruppe“ (hier: Vergleichsgruppe). ....</i>	<i>498</i>
<i>Tabelle 193: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Geschlecht“ (hier: weiblich).....</i>	<i>498</i>
<i>Tabelle 194: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Geschlecht“ (hier: männlich). ....</i>	<i>499</i>
<i>Tabelle 195: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Gruppe“ (hier: Interventionsgruppe). ....</i>	<i>499</i>
<i>Tabelle 196: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Gruppe“ (hier: Vergleichsgruppe). ....</i>	<i>499</i>
<i>Tabelle 197: Post-hoc-Analyse des Haupteffekts „Messzeitpunkt“ (Interventionsgruppe – Vergleichsgruppe).....</i>	<i>500</i>
<i>Tabelle 198: Post-hoc-Analyse des Haupteffekts „Messzeitpunkt“ (Gymnasium - Gesamtschule) .....</i>	<i>500</i>
<i>Tabelle 199: Deskriptive Statistik der Variable "Nützlichkeit der Lernhilfen" in der Interventions- und Vergleichsgruppe der Mittelstufe.....</i>	<i>501</i>
<i>Tabelle 200: Deskriptive Statistik der bevorzugten Art der Hilfenutzung (digital / analog) in der Mittelstufe.....</i>	<i>501</i>
<i>Tabelle 201: Deskriptive Statistik für die Gründe der Nutzung der Lernhilfen in der Interventionsgruppe der Mittelstufe. ....</i>	<i>502</i>
<i>Tabelle 202: Deskriptive Statistik für die Gründe der Nutzung der Vertiefungsaufgabe in der Interventionsgruppe der Mittelstufe. ....</i>	<i>503</i>
<i>Tabelle 203: Deskriptive Statistik für die Gründe der Nutzung der Lernhilfen in Vergleichsgruppe der Mittelstufe.....</i>	<i>504</i>
<i>Tabelle 204: Deskriptive Statistik für die Gründe der Nutzung der Vertiefungsaufgabe in der Vergleichsgruppe der Mittelstufe.....</i>	<i>505</i>
<i>Tabelle 205: Deskriptive Statistik für die Variablen der Zeit in der Interventionsgruppe der Mittelstufe.....</i>	<i>506</i>
<i>Tabelle 206: Deskriptive Statistik für die Variablen der Zeit in der Vergleichsgruppe der Mittelstufe.....</i>	<i>506</i>

<i>Tabelle 207: Unkorrigierte Hilfenutzung pro Schüler:in pro Stunde am Gymnasium (n = 205).</i>	507
<i>Tabelle 208: Unkorrigierte Hilfenutzung pro Schüler:in pro Stunde an der Gesamtschule (n = 115).</i>	507
<i>Tabelle 209: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Geschlecht“ (hier: weiblich).</i>	507
<i>Tabelle 210: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Geschlecht“ (hier: männlich).</i>	507
<i>Tabelle 211: Unkorrigierte Hilfenutzung pro Schüler:in pro Stunde am Gymnasium (n = 101).</i>	508
<i>Tabelle 212: Deskriptive Statistik der Variable "Nützlichkeit der Lernhilfen" in der Interventions- und Vergleichsgruppe der Oberstufe.</i>	508
<i>Tabelle 213: Deskriptive Statistik der bevorzugten Art der Hilfenutzung (digital / analog) in der Oberstufe.</i>	508
<i>Tabelle 214: Deskriptive Statistik für die Gründe der Nutzung der Lernhilfen in der Interventionsgruppe der Oberstufe.</i>	509
<i>Tabelle 215: Deskriptive Statistik für die Gründe der Nutzung der Vertiefungsaufgabe in der Interventionsgruppe der Oberstufe.</i>	510
<i>Tabelle 216: Deskriptive Statistik für die Gründe der Nutzung der Lernhilfen in Vergleichsgruppe der Oberstufe.</i>	511
<i>Tabelle 217: Deskriptive Statistik für die Gründe der Nutzung der Vertiefungsaufgabe in der Vergleichsgruppe der Mittelstufe.</i>	512
<i>Tabelle 218: Deskriptive Statistik für die Variablen der Zeit in der Interventionsgruppe der Oberstufe.</i>	513
<i>Tabelle 219: Deskriptive Statistik für die Variablen der Zeit in der Vergleichsgruppe der Oberstufe.</i>	513
<i>Tabelle 220: Deskriptive Statistik für die Variablen TAB1 und TAB 2 in der Interventionsgruppe der Mittelstufe.</i>	514
<i>Tabelle 221: Deskriptive Statistik für die Variable TAB3 in der Interventionsgruppe der Mittelstufe.</i>	514
<i>Tabelle 222: Deskriptive Statistik für die Variablen TAB 1 und TAB 2 in der Interventionsgruppe der Oberstufe.</i>	515
<i>Tabelle 223: Deskriptive Statistik für die Variable TAB3 in der Interventionsgruppe der Oberstufe.</i>	515
<i>Tabelle 224: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Gruppe“ (hier: Interventionsgruppe).</i>	516

---

<i>Tabelle 225: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Gruppe“ (hier: Vergleichsgruppe).....</i>	<i>516</i>
<i>Tabelle 226: Post-hoc-Analyse Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Geschlecht“ (hier: weiblich).....</i>	<i>516</i>
<i>Tabelle 227: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Geschlecht“ (hier: männlich). ....</i>	<i>517</i>
<i>Tabelle 228: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Geschlecht“ (hier: weiblich).....</i>	<i>517</i>
<i>Tabelle 229: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Geschlecht“ (hier: männlich). ....</i>	<i>517</i>
<i>Tabelle 230: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Gruppe“ (hier: Interventionsgruppe). ....</i>	<i>518</i>
<i>Tabelle 231: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Gruppe“ (hier: Vergleichsgruppe).....</i>	<i>518</i>
<i>Tabelle 232: Post-hoc-Analyse des Haupteffekts „Messzeitpunkt“ (Interventionsgruppe - Vergleichsgruppe).....</i>	<i>518</i>
<i>Tabelle 233: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Gruppe“ (hier: Interventionsgruppe). ....</i>	<i>519</i>
<i>Tabelle 234: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Gruppe“ (hier: Vergleichsgruppe).....</i>	<i>519</i>
<i>Tabelle 235: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Schulform“ (hier: Gymnasium).....</i>	<i>520</i>
<i>Tabelle 236: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt*Schulform“ (hier: Gesamtschule).....</i>	<i>520</i>
<i>Tabelle 237: Post-hoc-Analyse des Haupteffekts „Messzeitpunkt“ (Gymnasium – Gesamtschule).....</i>	<i>520</i>

## 9.1 Methoden

### 9.1.1 Betrachtungsdauer Hilfen und Vertiefungsaufgaben

#### Mittelstufe

*Tabelle 164: Betrachtungsdauer für die 1. Stunde in der Mittel- und Oberstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt. Kursive Hilfen wurden bei der Nutzungshäufigkeit nicht berücksichtigt.*

<i>Hilfe / Vertiefung</i>	<i>Dauer</i>	<i>Art</i>
AB1A2HA	00:00:17	Audio
AB1A1HA	00:00:10	Audio
AB1A1HB	00:00:01	Bild
AB1A2HB	00:00:04	Bild
<i>AB1L1</i>	<i>00:00:04</i>	<i>Sonstige</i>
<i>AB1HV<sub>i</sub></i>	<i>00:01:24</i>	<i>Sonstige</i>
<i>AB1HT</i>	<i>00:00:02</i>	<i>Sonstige</i>
<i>AB1HA</i>	<i>00:00:09</i>	<i>Sonstige</i>
<i>AB1HB</i>	<i>00:00:08</i>	<i>Sonstige</i>
<i>AB1HV</i>	<i>00:00:03</i>	<i>Sonstige</i>
AB1A2HT	00:00:05	Text
AB1A1HT	00:00:03	Text
AB1V	00:00:01	Vertiefung
AB1A1HV	00:00:24	Video
AB1A2HV	00:00:43	Video

*Tabelle 165: Betrachtungsdauer für die 2. Stunde in der Mittelstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt. Kursive Hilfen wurden bei der Nutzungshäufigkeit nicht berücksichtigt.*

<i>Hilfe / Vertiefung</i>	<i>Dauer</i>	<i>Art</i>
AB2MHUA	00:00:14	Audio
AB2MA1HA	00:00:24	Audio
AB2MA3HA	00:00:19	Audio
AB2MA2HA	00:00:11	Audio
AB2MHUB	00:00:01	Bild
AB2MA3HB	00:00:07	Bild
AB1MA1HB	00:00:05	Bild
<i>AB2ML1</i>	<i>00:00:01</i>	<i>Sonstige</i>
<i>AB2ML2</i>	<i>00:00:06</i>	<i>Sonstige</i>
AB2MHUT	00:00:03	Text
AB2MA1HT	00:00:09	Text
AB2MA3HT	00:00:06	Text
AB2MA2HT	00:00:09	Text
A2MV	00:00:19	Vertiefung
AB2MA3HV	00:00:52	Video
AB2MA1HV	00:00:56	Video
AB2MHUV	00:00:29	Video

*Tabelle 166: Betrachtungsdauer für die 3. Stunde in der Mittelstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt. Kursive Hilfen wurden bei der Nutzungshäufigkeit nicht berücksichtigt.*

<i>Hilfe / Vertiefung</i>	<i>Dauer</i>	<i>Art</i>
AB3MA1HA	00:00:22	Audio
AB3MA3HA	00:00:21	Audio
<i>AB3MODA</i>	<i>00:01:32</i>	<i>Audio</i>
AB3MA2HA	00:00:32	Audio
AB3MOA1HB	00:00:02	Bild
<i>AB3MODB</i>	<i>00:00:16</i>	<i>Bild</i>
AB3MA3HB	00:00:12	Bild
AB3MA2HB	00:00:12	Bild
AB3MOA1HT	00:00:06	Text
AB3MA3HT	00:00:09	Text
<i>AB3MODT</i>	<i>00:00:34</i>	<i>Text</i>
AB3MA2HT	00:00:10	Text
AB3MV	00:00:03	Vertiefung
<i>AB3MODV</i>	<i>00:03:50</i>	<i>Video</i>
AB3MA3HV	00:01:32	Video
AB3MA2HV	00:02:01	Video
AB3MOA1HV	00:00:35	Video



*Tabelle 167: Betrachtungsdauer für die 4. Stunde in der Mittelstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt.*

<i>Hilfe / Vertiefung</i>	<i>Dauer</i>	<i>Art</i>
AB4MA1HA	00:00:18	Audio
AB4MA2HA	00:00:34	Audio
AB4MA3HA	00:00:56	Audio
AB4MA1HB	00:00:08	Bild
AB4MOA2HB	00:00:05	Bild
AB4MA3HB	00:00:08	Bild
AB4MA1HT	00:00:10	Text
AB4MA3HT	00:00:22	Text
AB4MA2HT	00:00:12	Text
AB4MV	00:00:05	Vertiefung
AB4MA1HV	00:01:01	Video
AB4MOA2HV	00:00:42	Video
AB4MA3HV	00:01:12	Video

**Oberstufe**

*Tabelle 168: Betrachtungsdauer für die 2. Stunde in der Oberstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt. Kursive Hilfen wurden bei der Nutzungshäufigkeit nicht berücksichtigt.*

<i>Hilfe / Vertiefung</i>	<i>Dauer</i>	<i>Art</i>
AB3MA1HA	00:00:22	Audio
AB2OA2HA	00:00:44	Audio
AB2OA3HA	00:00:15	Audio
<i>AB3MODA</i>	<i>00:01:32</i>	<i>Audio</i>
AB3MOA1HB	00:00:02	Bild
AB2OA3HB	00:00:03	Bild
<i>AB3MODB</i>	<i>00:00:16</i>	<i>Bild</i>
AB2OA2HB	00:00:19	Bild
AB3MOA1HT	00:00:06	Text
AB4OA2HT	00:00:10	Text
AB2OA3HT	00:00:05	Text
<i>AB3MODT</i>	<i>00:00:34</i>	<i>Text</i>
AB2OV	00:00:05	Vertiefung
AB3MOA1HV	00:00:35	Video
AB2OA2HV	00:03:44	Video
AB2OA3HV	00:00:49	Video
<i>AB3MODV</i>	<i>00:03:50</i>	<i>Video</i>

*Tabelle 169: Betrachtungsdauer für die 3. Stunde in der Oberstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt.*

<i>Hilfe / Vertiefung</i>	<i>Dauer</i>	<i>Art</i>
AB3OA1HA	00:00:17	Audio
AB3OA3HA	00:00:18	Audio
AB3OA4HA	00:00:10	Audio
AB3OA4HB	00:00:03	Bild
AB3OA3HB	00:00:01	Bild
AB3OI	11:59:59	Sonstige
AB3OA1HT	00:00:02	Text
AB3OA4HT	00:00:02	Text
AB3OA3HT	00:00:05	Text
AB3OV	00:00:14	Vertiefung
AB3OA1HV	00:00:40	Video
AB3OA4HV	00:01:06	Video
AB3OA3HV	00:01:11	Video

*Tabelle 170: Betrachtungsdauer für die 4. Stunde in der Oberstufe, ab der eine Hilfe oder Vertiefungsaufgabe als "gelesen" gilt. Kursive Hilfen wurden bei der Nutzungshäufigkeit nicht berücksichtigt.*

<i>Hilfe / Vertiefung</i>	<i>Dauer</i>	<i>Art</i>
AB2OA2HA	00:00:44	Audio
AB4OA1HA	00:00:33	Audio
AB4OA4HA	00:00:18	Audio
<i>AB4ODA</i>	<i>00:01:02</i>	<i>Audio</i>
<i>AB4OWA</i>	<i>00:01:03</i>	<i>Audio</i>
AB2OA2HB	00:00:19	Bild
AB4MOA2HB	00:00:05	Bild
<i>AB4ODB</i>	<i>00:00:15</i>	<i>Bild</i>
<i>AB4OWB</i>	<i>00:00:07</i>	<i>Bild</i>
AB4OA2HT	00:00:10	Text
AB4OA1HT	00:00:12	Text
<i>AB4ODT</i>	<i>00:00:21</i>	<i>Text</i>
<i>AB4OWT</i>	<i>00:00:18</i>	<i>Text</i>
AB4OA4HT	00:00:06	Text
AB4OV	00:00:03	Vertiefung
AB2OA2HV	00:03:43	Video
AB4MOA2HV	00:00:42	Video
<i>AB4ODV</i>	<i>00:01:07</i>	<i>Video</i>
<i>AB4OWV</i>	<i>00:01:32</i>	<i>Video</i>

## 9.1.2 Faktorenmodelle

### Intrinsische Motivation (*Intrinsic Motivation Inventory*)

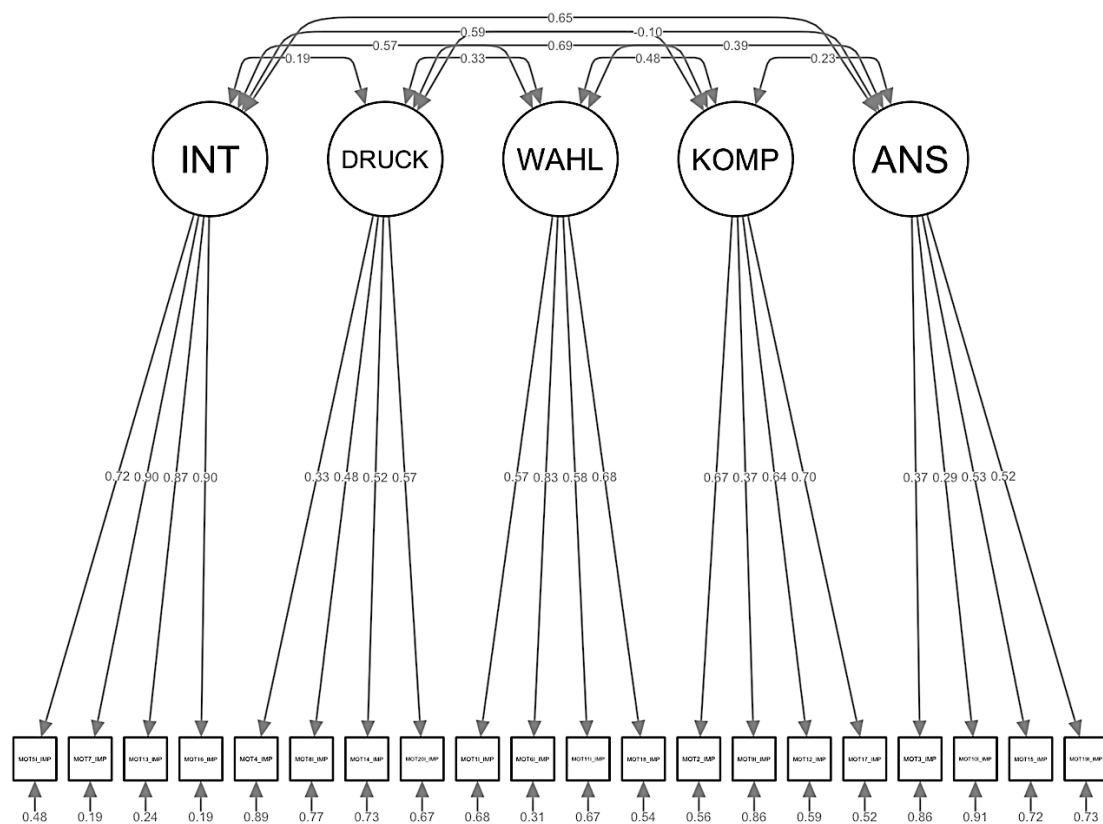


Abbildung 92: IMI Modell 1 - Alle fünf Faktoren Interesse / Spaß (INT), Druck (DRUCK), wahrgenommene Wahlfreiheit (WAHL), Kompetenzerleben (KOMP), Anstrengung (ANS).

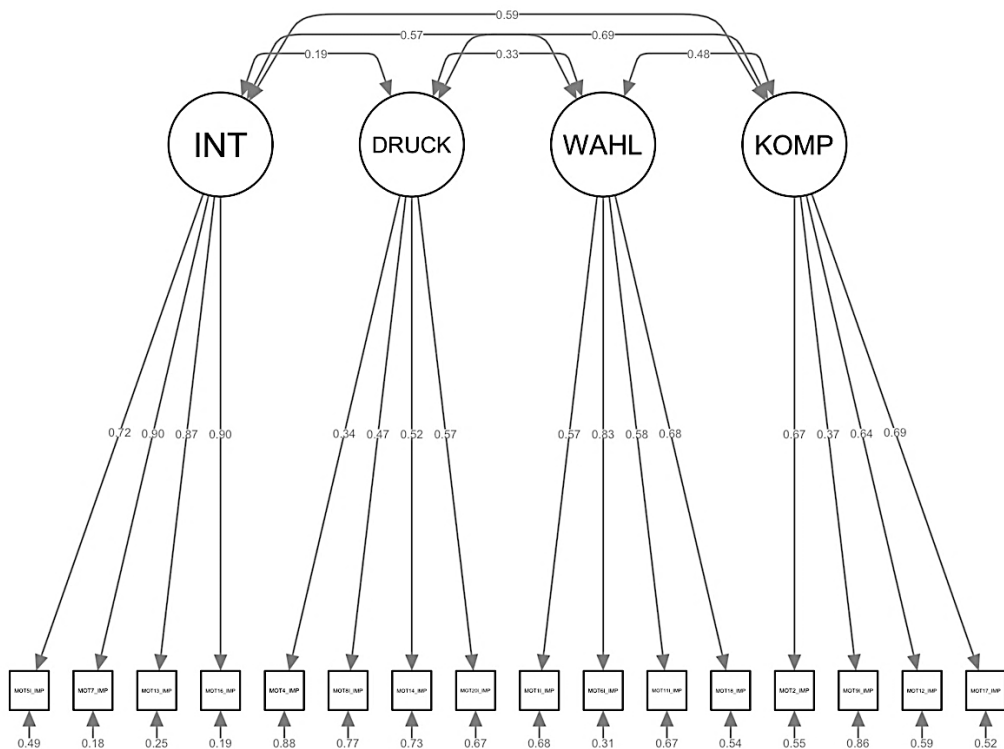


Abbildung 93: IMI Modell 2 - Vier Faktoren Interesse / Spaß (INT), Druck (DRUCK), wahrgenommene Wahlfreiheit (WAHL), Kompetenzerleben (KOMP).

**Usability**

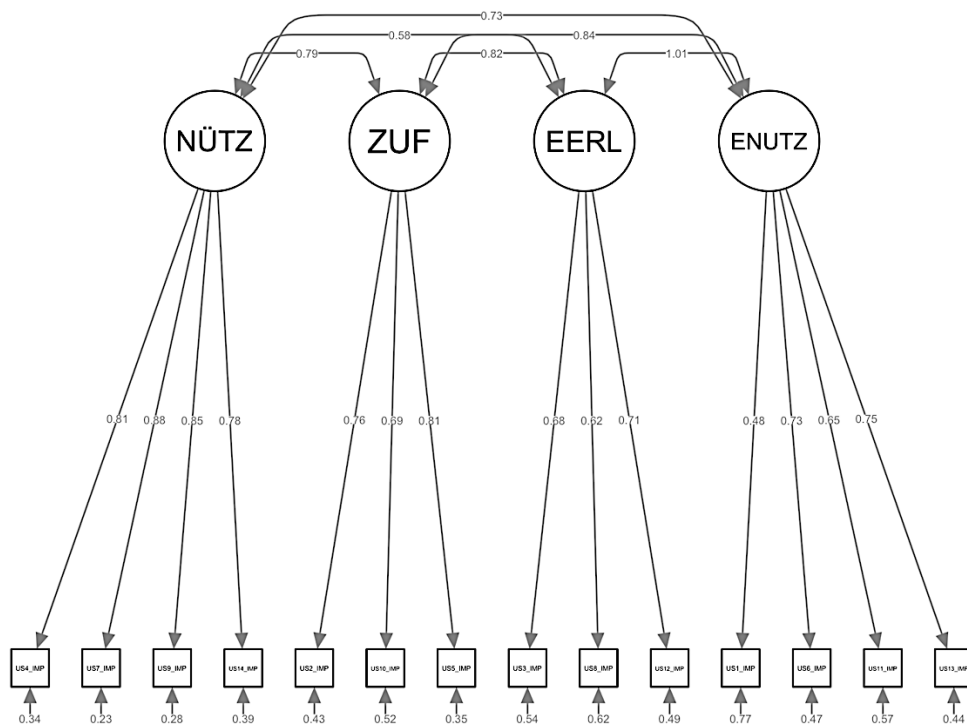


Abbildung 94: Usability Modell 1 - Alle vier Faktoren Nützlichkeit (NÜTZ), Zufriedenheit (ZUF), Einfachheit des Erlernens der Nutzung (EERL), Einfachheit der Nutzung (ENUTZ).

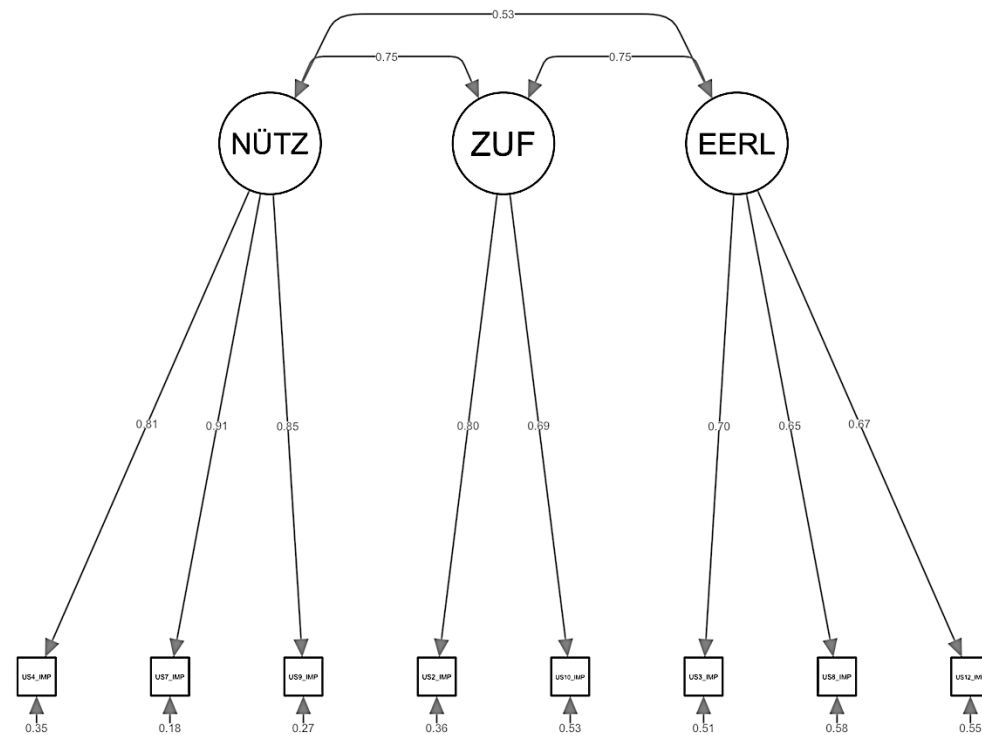


Abbildung 95: Usability Modell 2 - Drei Faktoren Nützlichkeit (NÜTZ), Zufriedenheit (ZUF), Einfachheit des Erlernens der Nutzung (EERL).

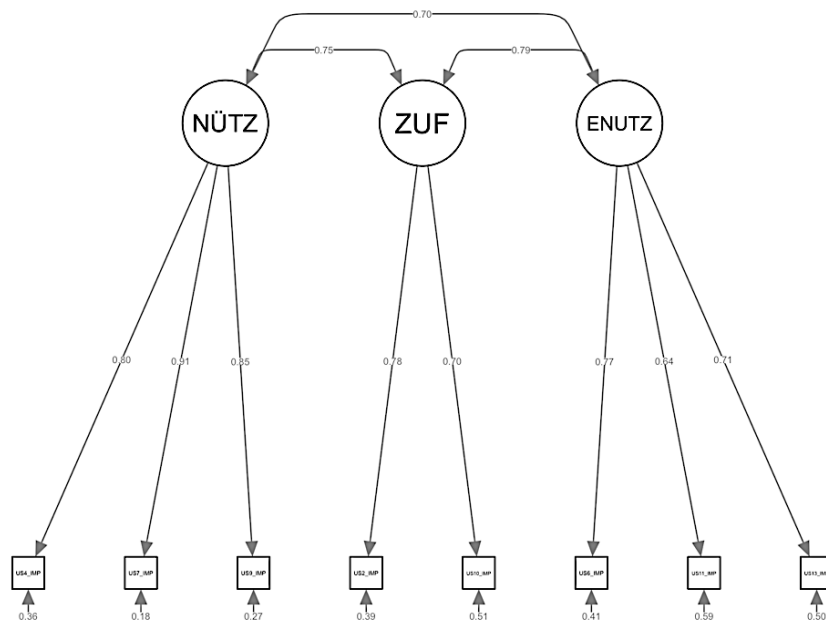


Abbildung 96: Usability Modell 3 - Drei Faktoren Nützlichkeit (NÜTZ), Zufriedenheit (ZUF), Einfachheit der Nutzung (ENUTZ).

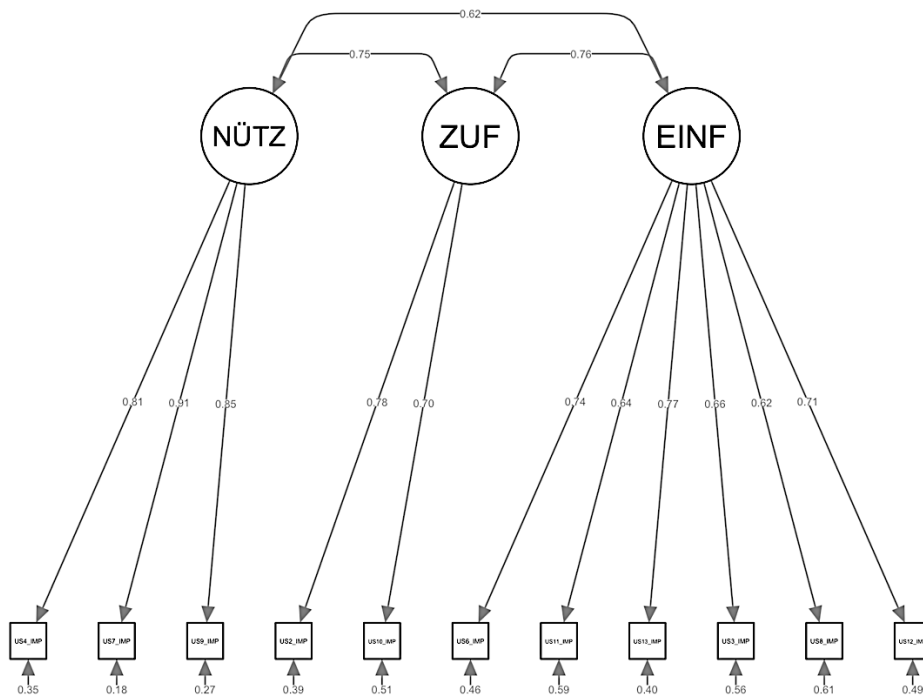


Abbildung 97: Usability Modell 4 - Drei Faktoren Nützlichkeit (NÜTZ), Zufriedenheit (ZUF), Einfachheit der Nutzung und des Erlernens der Nutzung (EINF).

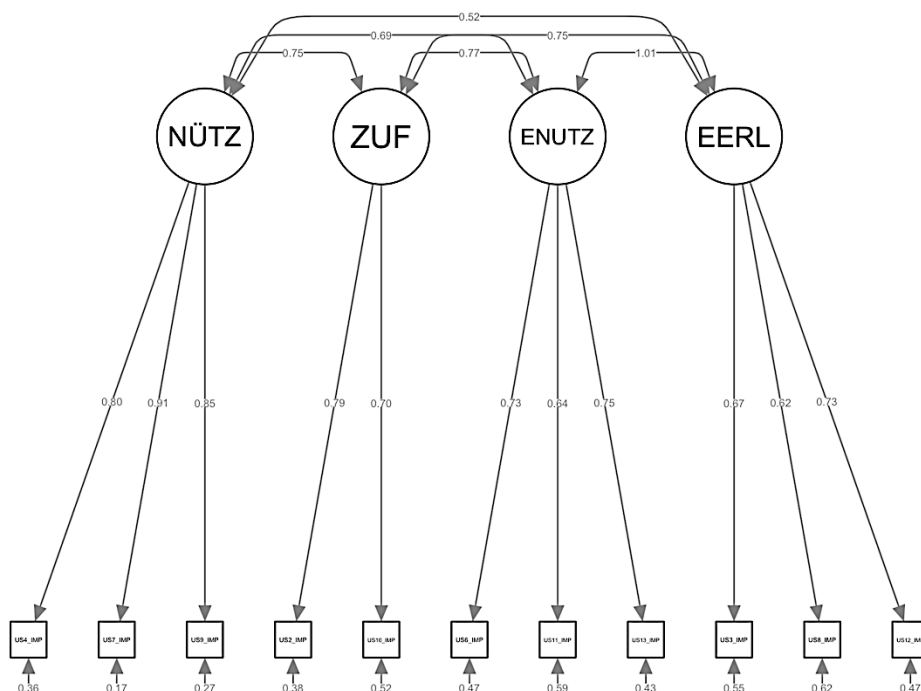


Abbildung 98: Usability Modell 5 - Alle vier Faktoren Nützlichkeit (NÜTZ), Zufriedenheit (ZUF), Einfachheit des Erlernens der Nutzung (EERL), Einfachheit der Nutzung (ENUTZ). Reduzierte Itemzahl.



**Tabletbezogene Kompetenzen**

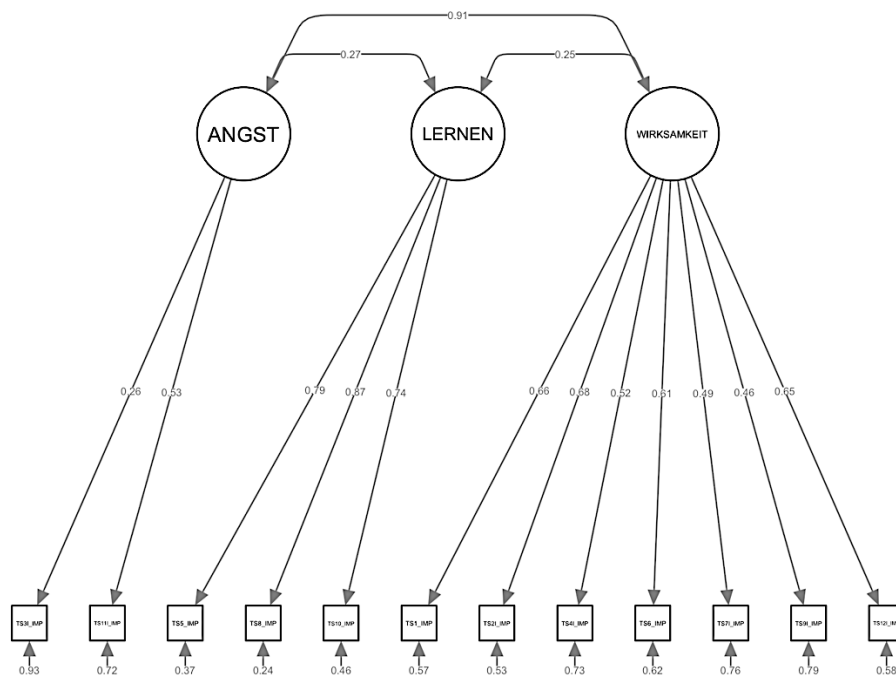


Abbildung 99: Tabletbezogene Kompetenzen Modell 1 - Alle drei Faktoren Selbstwirksamkeit (WIRKSAMKEIT), Nützlichkeit beim Lernen (LERNEN), Angst vor Tablets (ANGST).

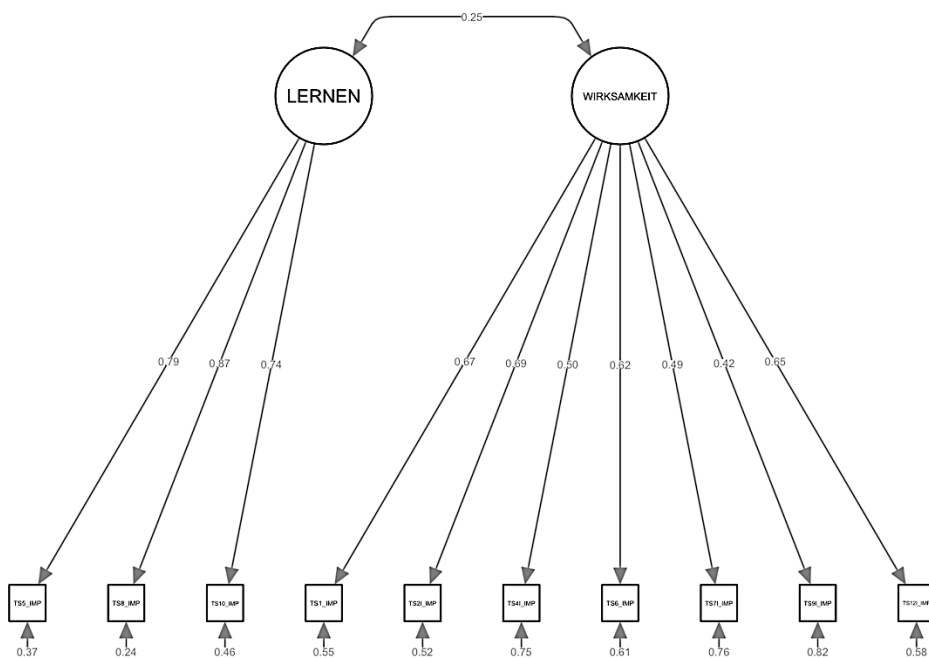


Abbildung 100: Tabletbezogene Kompetenzen Modell 2 - Zwei Faktoren Selbstwirksamkeit (WIRKSAMKEIT), Nützlichkeit beim Lernen (LERNEN).

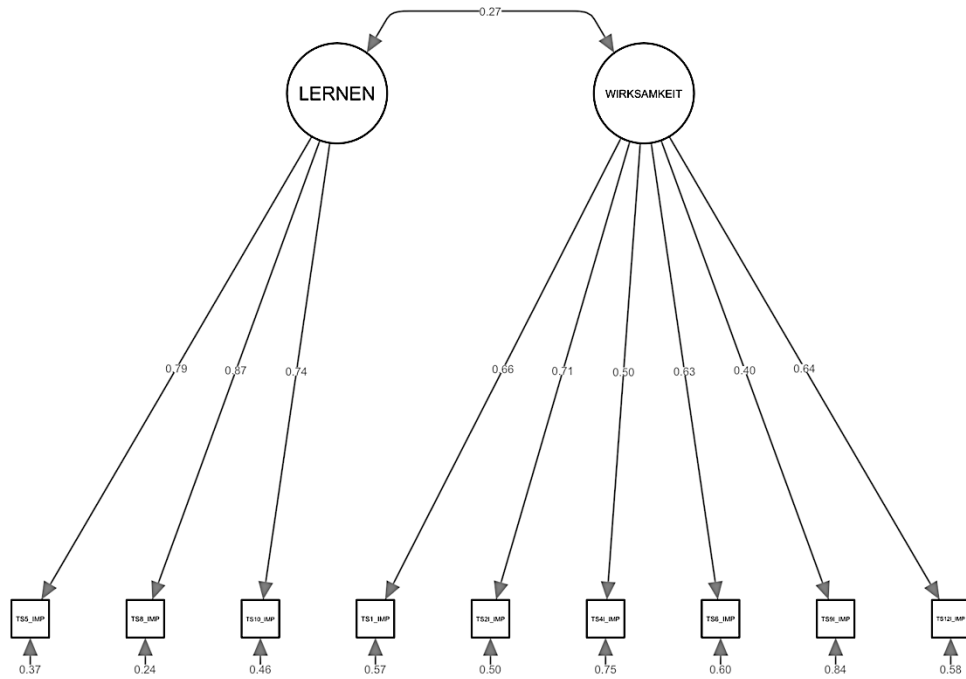


Abbildung 101: Tabletbezogene Kompetenzen Modell 3 - Zwei Faktoren Selbstwirksamkeit (WIRKSAMKEIT), Nützlichkeit beim Lernen (LERNEN). Reduzierte Itemzahl.

**Cognitive Load**

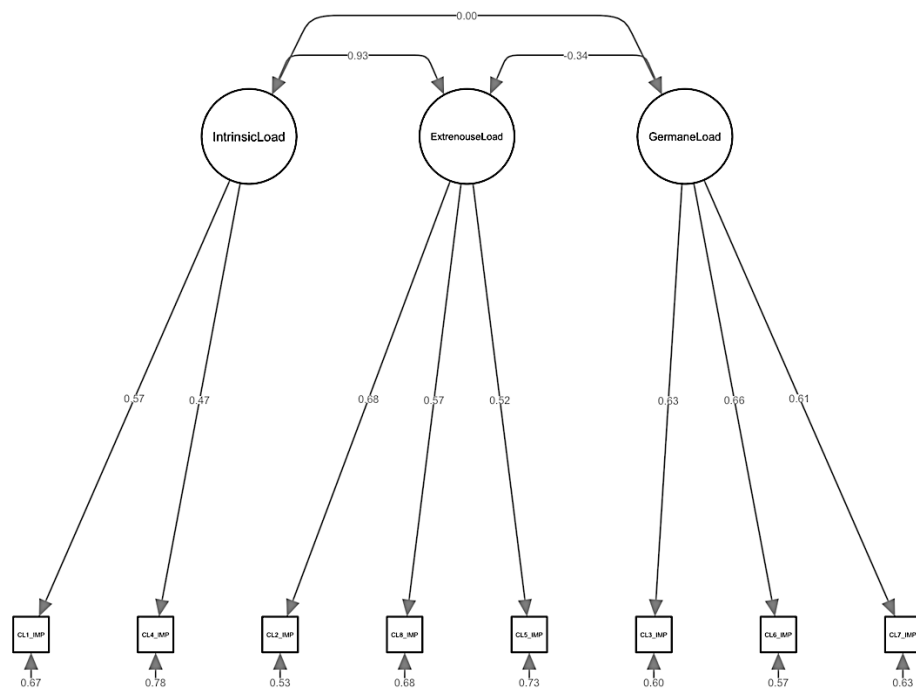


Abbildung 102: Cognitive Load Modell 1 - Drei Faktoren Intrinsic Load, Extraneous Load, Germane Load.

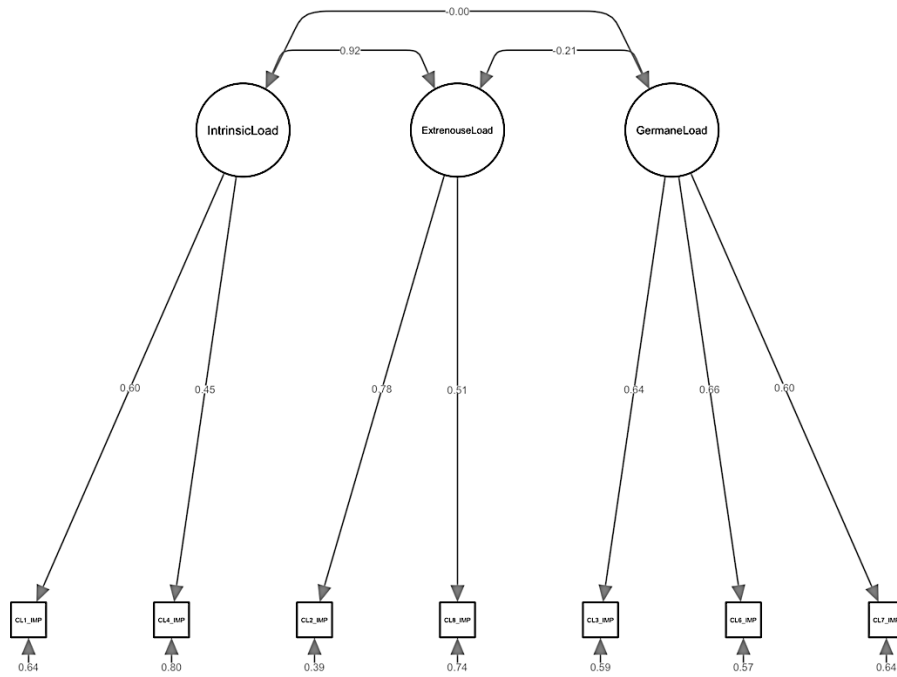





Abbildung 103: Cognitive Load Modell 2 – Drei Faktoren Intrinsic Load, Extraneous Load, Germane Load. Reduzierte Itemzahl.


### 9.1.3 Einführung und HyperDocs der Interventionsgruppe






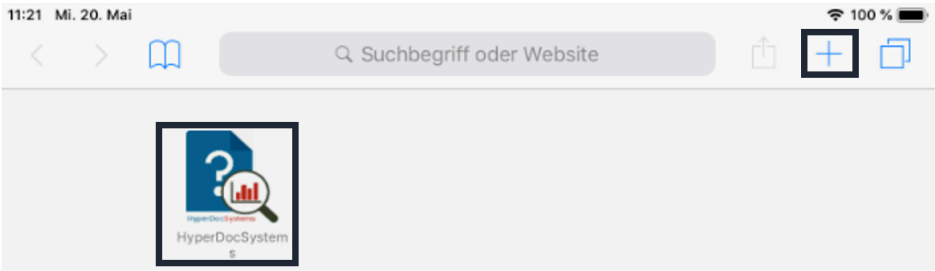



**Einführung:**

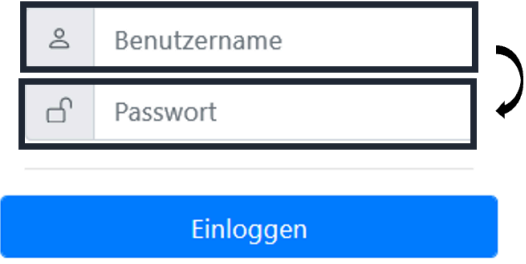
- Schalte das Tablet ein, indem du das Tablet öffnest und auf den runden Knopf am unteren Tabletrand drückst.
- Stelle sicher, dass du mit dem WLAN „FD Chemie“ verbunden bist. Gehe dazu in die Einstellungen → WLAN und wähle das Netzwerk „FD Chemie“ aus. 
- Öffne Safari (Kompass-Symbol):



- Öffne nun *HyperDocSystems*. Hierbei handelt es sich um das digitale Arbeitsblatt. Drücke gegebenenfalls auf das +, um die Favoritenseite zu öffnen:



- Nutze das Tablet in der Horizontalen (quer)  oder Vertikalen (aufrecht) 
- Logge dich mit deinem spezifischen Benutzernamen ein. **Das Passwort ist gleich dem Benutzernamen:**



- Am Ende der Stunde ist es wichtig, dass du dich wieder ausloggst (oben rechts):

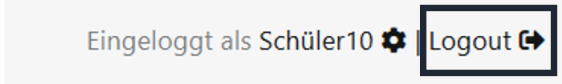


Abbildung 104: Einführung in die Nutzung der Tablets.

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN FACHDIDAKTIK CHEMIE

# HyperDocSystems

Digitale Arbeitsblätter erstellen, verwalten und bearbeiten.

Login

Benutzername

Passwort

Einloggen

Registrieren

Cookie-Disclaimer

Die Website verwendet Cookies, um eine bestmögliche Nutzung zu gewährleisten. Nähere Informationen und die Möglichkeit, die Verwendung von Cookies einzuschränken finden Sie [hier](#).

Aufzeichnung

Die Website zeichnet das Verhalten der Benutzer im Umgang mit digitalen Arbeitsblättern auf, um eine Lerndiagnose zu ermöglichen. Die Daten werden dauerhaft gespeichert und sind accountbezogen.

↓ Login

Arbeitsblätter

Eingelogg als test | Logout

Sonstige zugewiesene Arbeitsblätter:

Arbeitsblatt

Arbeitsblatt: "Safety First!"

Einführung: "Safety First!" (engl. Sicherheit zuerst) - so heißt es oftmals im Alltag, wenn Gefahren drohen. Aber auch in der Schule gilt der Spruch. Gerade beim Experimentieren in Naturwissenschaften können gefährliche Situationen entstehen. Sei es durch das Experimentieren mit Elektrizität im Physikunterricht oder durch Chemikalien im Chemieunterricht. Methode: Im Folgenden erhaltet ihr eine Einführung in die digitalen Arbeitsblätter. Gleichzeit lernt [...].

Arbeitsblatt öffnen

Zuletzt bearbeitet am 23.08.2020 von fitting

Arbeitsblatt

Arbeitsblatt: Farbanalyse mit dem Smartphone (O)

Erinnerung: Du solltest das Tablet beim Öffnen eines Arbeitsblattes bereits in der gewünschten Position verwenden, da das Drehen des Tablets nach dem Öffnen eines Arbeitsblattes zu Problemen beim Schreiben auf den Textfeldern führen kann. Drehe das Tablet ggf. in eine neue Position und öffne das Arbeitsblatt erneut. Einstieg Versuch Du kannst du hier nochmal die Versuchsdurchführung anschauen, anhören oder durch [...].

Arbeitsblatt öffnen

Zuletzt bearbeitet am 23.08.2020 von fitting

Abbildung 105: Benutzeroberfläche von HyperDocSystems.



Arbeitsblätter | Eingelogg als test | Logout

### Arbeitsblatt: Farbstoffe in Lebensmitteln

**Hinweisung:** Du schneid das Tablet beim Öffnen eines Arbeitsblattes bereits in der gewünschten Position verwenden, da das Drehen des Tablets nach dem Öffnen eines Arbeitsblattes zu Problemen beim Schreiben auf den Textfeldern führen kann. Drehe das Tablet ggf. in eine neue Position und öffne das Arbeitsblatt erneut.

**Einstieg: Sirup**

Sirupflasche

**Lebensmittelzusatzstoffe**

Die Europäische Union (EU) verfügt eine E-Nummer für jeden zugelassenen Stoff. Insgesamt gibt es in der EU derzeit 341 zugelassene Zusatzstoffe. Eine Zulassung wird nur erteilt, wenn keine Gesundheitsrisiken bestehen, wenn der Zusatzstoff technisch notwendig ist und wenn die Verwendung nicht zu einer Täuschung des Verbrauchers führt. Zusatzstoffe dürfen nur nach ausdrücklicher Zulassung verwendet werden. Zudem müssen Zusatzstoffe auf dem Produkt kenntlich gemacht werden.

**Rechtliche Situation**

Die EU-Kommission beauftragte am 25. März 2010 die EFSA (European Food Safety Authority), eine neue chemische Bewertung färglicher Azofarbstoffe vorzunehmen. Seit dem 20. Juli 2010 müssen Lebensmittel, die die Azofarbstoffe Tartrazin (E 102), Gelborange 5 (E 110), Azorubin (E 122), Allurarot (E 129) oder Cochenillerot A (E 124) enthalten, in der Europäischen Union mit dem gesonderten Warnhinweis „Kann Aktivität und Aufmerksamkeit bei Kindern beeinträchtigen“ gekennzeichnet werden.

**Azorubin**

Azorubin (auch Karmesin, E 122) ist ein als Lebensmittelfarbe zugelassener roter Azofarbstoff und das Dinatriumsalz einer Disulfonäure. Azorubin wird für Getränke, Süß- und Zuckwaren, Marzipan, Puddingpulver, Götterspeise, Fruchtkonserven, Fettgruppen, Fertigprodukte, braune Saucen, Paniermehl für Tabletten und anderes verwendet. Azorubin hat große chemische Ähnlichkeit mit Amarant (E 123) und steht wie dieser Stoff im Verdacht, als Auslöser von Pseudoallergien zu wirken.

**Tartrazin**

Tartrazin (E 102) ist ein synthetischer Azofarbstoff, der als **zitronengelber bis orangefarbener** Lebensmittelarbstoff eingesetzt wird. Der Stoff ist farbtreu.

**Erlaubte Tagesdosis**

Die erlaubte Tagesdosis (ETD, englisch: acceptable daily intake, ADI) bezeichnet die Dosis einer Substanz, wie etwa eines Lebensmittelzusatzstoffs, Pestizids oder eines Medikaments, die bei lebenslanger täglicher Einnahme als medizinisch unbedenklich betrachtet wird. Handelt es sich um ungewollte Verunreinigungen, spricht man von einer tolerierten Tagesdosis (englisch: tolerable daily intake, TDI).

Angaben wird der ETD-Wert (ADI-Wert) in Milligramm bzw. Mikrogramm pro Kilogramm Körpergewicht und Tag, Synergieeffekte (Zusammenwirken verschiedener Lebensmittelzusatzstoffe) werden jedoch nur teilweise berücksichtigt.

**ADI-Wert von Azorubin:** 4 mg / kg Körpergewicht  
**ADI-Wert von Tartrazin:** 7,5 mg / kg Körpergewicht

Hilfe für Umrechnung

**Aufgaben**

1. Begründe die Einheit des ADI-Wertes (mg/kg Körpergewicht).

**2. Berechne die Masse, die ein Kleinkind und ein Erwachsener täglich von beiden Farbstoffen aufnehmen dürfen.**

Person	Azorubin [mg]	Tartrazin [mg]
Kleinkind (10 kg)	<input type="text"/>	<input type="text"/>
Bodybuilder (100 kg)	<input type="text"/>	<input type="text"/>

**3. Du möchtest das Risiko bei der Verwendung des Grenadine-Sirups abschätzen. Begründe, für welchen Farbstoff du zunächst den Gehalt in der Getränkeprobe bestimmen würdest.**

**Wertungsaufgabe**

Quelle(n):

https://www.wikipedia.org/wiki/Lebensmittelzusatzstoff#/media/Datei:Lebensmittelzusatzstoff.svg  
 https://www.wikipedia.org/wiki/Lebensmittelzusatzstoff#/media/Datei:Lebensmittelzusatzstoff.svg  
 Die Texte stehen unter der Lizenz: Creative Commons Attribution-ShareAlike 4.0 International

Speichern & zurück

Abbildung 107: HyperDoc der 2. Stunde der Unterrichtsreihe (Mittelstufe). Verändert nach Czubatinski et al. (2020). Kleinkind nach OmarMedinaFilms (2016).

Arbeitsblätter

Eingelogg als test | Logout

**Arbeitsblatt: Farbanalyse mit dem Smartphone**

---

**Versuch**  
Du kannst dir hier nochmal die Versuchsdurchführung anschauen, anhören oder durchlesen:

**Tabelle RGB-Werte**  
1. Trage die mit dem Handy gemessenen Werte in die Tabelle ein.

Konzentration [mg/l]	Rot-Werte	Grün-Wert	Blau-Werte
20 mg/l	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>
30 mg/l	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>
40 mg/l	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>
50 mg/l	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>
60 mg/l	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>
verd. Sirup	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>
unb. Konz.	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>

**Diagramm RGB-Werte**  
2. Übertrage die Werte der bekannten Konzentrationen in das Diagramm. Verwende entsprechende Farben.

**Aufgaben:**  
3. Versuche die Konzentration von Azorubin in der Getränkeprobe abzuschätzen, indem du den R-/G-/B-Wert der Getränkeprobe mit den Messwerten (R-/G-/B-Wert) der Kalibrierlösungen (20 - 60 mg/l) vergleichst. Ziehe dazu eine Linie durch alle Punkte einer Wertreihe, sodass alle Punkte einer Reihe in etwa gleich weit von der Linie entfernt sind. Du erhältst so am Ende drei Linien: 1 rote, 1 grüne und 1 blaue Linie. Betrachte die Hilfen, falls du dir unsicher bist.

Schätzung anhand der	R-Werte	G-Werte	B-Werte
Konzentration in mg/l	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>	<input style="width: 100%; height: 20px;" type="text"/>

**Vertiefungsaufgabe 1**

Speichern & zurück
B

Abbildung 108: HyperDoc der 3. Stunde der Unterrichtsreihe (Mittelstufe). Verändert nach Czubatinski et al. (2020).

Seite | 446



Arbeitsblätter

Eingelogg als test | Logout

---

**Arbeitsblatt: Konzentrationen**

**Table Konzentrationbestimmung**

1. Stelle die Konzentration der Kalibrierlösungen und der Getränkeprobe analog zu dem vorgegebenen Beispiel (50 mg/l) in der Tabelle dar! **Runde** die Konzentration der Probe dazu auf 10 mg/l!

20 mg/l	30 mg/l	40 mg/l	50 mg/l	60 mg/l	mg/l
		Kalibrierlösungen			Probe

**Der Konzentrationsbegriff**

Die Konzentration ist eine Gehaltsangabe, die sich auf das Volumen eines Gemisches bezieht. Die Konzentration gibt an, wie viel eines Stoffes in einem bestimmten Volumen vorhanden ist. Man unterscheidet so zum Beispiel zwischen der Stoffmengenkonzentration  $c$  mit der Einheit mol/l oder der Massenkonzentration  $p$  mit der Einheit g/l. Die Massenkonzentration gibt also an, wie viel Gramm Stoff sich in einem Liter des Gemisches befinden.

2. Wie viele Gläser (0,2 Liter) kannst du laut des ADI-Werts (=4 mg/kg Körpergewicht) von einem zubereiteten Getränk des Sirups täglich trinken?

**T1**

3. Stelle die Abbildung "Mischverhältnis" wie oben dar. Zeichne zunächst die richtige Anzahl an Azorubinteilchen in den Teil, der für die Getränkeprobe benötigt wird ("1 Teil Sirup"). Berechne anschließend die Konzentration des **unverdünnten** Sirups mithilfe des Mischungsverhältnisses. Zeichne dann die richtige Anzahl an Teilchen in die Sirupflasche.

**Abbildung Mischverhältnis**

Konzentration im unverdünnten Sirup:

**T1**

Vertiefungsaufgabe:

**T1**

Speichern & zurück

Abbildung 109: HyperDoc der 4. Stunde der Unterrichtsreihe (Mittelstufe). Verändert nach Czubatinski et al. (2020).

**Oberstufe**

Arbeitsblätter

Eingelogggt als: test | Logout

---

**Arbeitsblatt "Safety First!"**



**Einführung:**  
 "Safety First" (engl. Sicherheit zuerst) - so heißt es oftmals im Alltag, wenn Gefahren drohen. Aber auch in der Schule gilt der Spruch. Gerade beim Experimentieren in Naturwissenschaften können gefährliche Situationen entstehen. Sei es durch das Experimentieren mit Elektrizität im Physikunterricht oder durch Chemikalien im Chemieunterricht.

**Methode:**  
 Im Folgenden erhaltet ihr eine Einführung in die digitalen Arbeitsblätter. Gleichzeit lernt ihr etwas über gefährliche Stoffe (Gefahrenstoffe) im Chemieunterricht.  
 Digitale Arbeitsblätter sind zunächst "fest wie 'normale' Arbeitsblätter. Das bedeutet, du erhältst verschiedene Informationen und musst dann Aufgaben bearbeiten. Das besondere an den digitalen Arbeitsblättern: An verschiedenen Stellen sind Hilfen, Verlinkungsaufgaben oder weiterführende Informationen hinterlegt. Diese zusätzlichen Inhalte öffnest du über **Verlinkungen** (blau unterstrichen) oder Symbole. Je nach Symbol unterscheidet sich die Darstellung des Inhalts. Die Hilfen, Verlinkungsaufgaben oder zusätzlichen Informationen erscheinen am oberen Bildschirmrand. Probleme es aus:  
 - Hilfe als Video: - Hilfe als Text: - Hilfe als Tonspur: **Hilfe als Bild:** Dabei stehen nicht immer alle Darstellungsformen für jede Hilfe zur Verfügung. Es gibt gegenüber aber auch Verlinkungsaufgaben, die durch das Symbol gekennzeichnet sind. Längere Hilfen sind außerdem "scrollbar", das heißt du klickst auf der Hilfe hoch und runter wischen, um die gesamte Hilfe zu sehen.

**Hinweis!** Du solltest das Tablet beim Öffnen eines Arbeitsblattes bereits in der gewünschten Position verwenden, da das Drehen des Tablets nach dem Öffnen eines Arbeitsblattes zu Problemen beim Schreiben auf den Textfeldern führen kann.

---

**Gefahrenstoffe im Alltag:**  
 In unserem Alltag befinden sich eine Vielzahl von verschiedenen Gefahrenstoffen, auch wenn es auf den ersten Blick nicht so scheint. Hier sehest du zum Beispiel die Rückseite eines Geschirrspülmittels. Dort steht geschrieben: "Verursacht schwere Augenreizungen. Darf nicht in die Hände von Kindern gelangen [...]". Gleichzeitig ist ein großes Symbol mit einem Ausrufezeichen zu sehen.

Das war nur ein Beispiel. Du findest viele weitere Gefahrenstoffhinweise auf allen möglichen alltäglichen Verbrauchsgüterständen. Insgesamt gibt es neun Kategorisierungen für Gefahrenstoffe, jede Kategorie wird dabei durch ein spezielles Symbol gekennzeichnet:

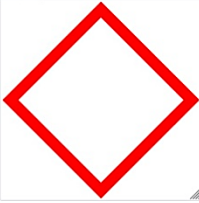
Symbol	Beschreibung	Beispiele für gefährliche Eigenschaften
	Explosierende Bombe	instabile explosive Stoffe
	Flammable	entzündbare Stoffe
	Flammable über 100°C	entzündende Stoffe
	Gasflasche	Gas unter Druck
	Ätzwirkung	schwere Augenschädigungen, hautreizend
	Tierkopffopf mit gekreuzten Knochen	akute Toxizität
	dickes Ausrufezeichen	hautreizend, augenreizend
	Gesundheitgefährlich	diverse Gesundheitsgefahren
	Umwelt	gesundheitgefährlich

**Hinweis!** Die folgenden Textfelder kannst du entweder mit der virtuellen Tastatur ausfüllen (Standardanstellung) oder mit dem Apple Pencil. Klicke dazu auf das Symbol. Mit der Werkzeugleiste kannst du zwischen dem Radierer und Pinsel wechseln, die Farbe einstellen und die Pinselgröße ändern. Achtung: Über den Besen weisst du das gesamte Textfeld! Manchmal ist das Tastatur- und Pinselsymbol auch über einem Bild zu sehen. Du kannst dann das Bild beschriften (s. Aufgabe 1).


**Wichtig!** Du musst deine Lösungen speichern, bevor du das Tablet schließt oder die Seite neu lädt. Drücke dazu den Button "Lösung speichern" am Ende der Seite.

**Aufgaben:**

1. "Vervollständige das Gefahrensymbol! "Dickes Ausrufezeichen".



2. Folgende Stoffe sind gegeben:



a)

b)

**Natronlauge**

c)

**Deodorant**

Daher jedem Stoff ein Gefahrensymbol zu. Schreibe dazu die Bezeichnung des Gefahrensymbols auf.

**T1**

**Verlinkungsaufgabe:**


**T1**

**Beim:**

Copyright © 1999-2018. Alle Rechte vorbehalten. Dieses Dokument ist ein Produkt der Firma Bru-nO. Alle Rechte vorbehalten.

Speichern & zurück

Abbildung 110: HyperDoc der 1. Stunde der Unterrichtsreihe (Oberstufe). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008).



Arbeitsblätter

Eingelogg als test | Logout


**Arbeitsblatt: Farbanalyse mit dem Smartphone (0)**

**Erinnerung:** Du solltest das Tablet beim Öffnen eines Arbeitsblattes bereits in der gewünschten Position verwenden, da das Drehen des Tablets nach dem Öffnen eines Arbeitsblattes zu Problemen beim Schreiben auf den Textfeldern führen kann. Drehe das Tablet ggf. in eine neue Position und öffne das Arbeitsblatt erneut.

Sirupflasche


**Versuch**

Du kannst du hier nochmal die Versuchsdurchführung anschauen, anhören oder durchlesen.

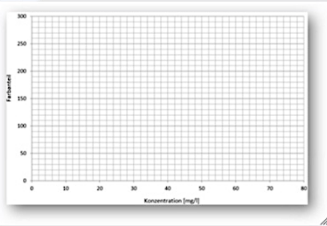


**Tabelle RGB-Werte**

1. Trage die mit dem Smartphone gemessenen Werte in die Tabelle ein.




Konzentration (mg / l)	Rot-Werte	Grün-Wert	Blau-Werte
20 mg/l			
30 mg/l			
40 mg/l			
50 mg/l			
60 mg/l			
Getränkprobe - unbekannte Konz.			



**Aufgaben**

3. Bestimme den Acorubin-Gehalt in der Getränkeprobe möglichst genau - begründe, welche Wertreihe sich hierfür am besten eignet.



Schätzung anhand der ...

R-Werte	G-Werte	B-Werte
Konzentration		

Begründung

**II**

Vertiefungsaufgabe:

**II**

Speichern & zurück

B

Abbildung 111: HyperDoc der 2. Stunde der Unterrichtsreihe (Oberstufe). Verändert nach Nieß et al. (o. D.).

Arbeitsblätter

Eingeloggt als: test | Logout

---

🏠 Arbeitsblatt: Das Spektralphotometer

**Aufbau eines Photometers**

1. Lies dir die beiden Informationsseiten durch und beschrifte die **Abbildung 1**.

**Information**

2. Lies den folgenden Text und vergleiche mit der **Abbildung 2**.

Im Labor der Universität wurde das Spektralphotometer zunächst dazu benutzt, das Absorptionsverhalten der beiden Azofarbstoffe zu untersuchen. Dazu wurden die Lösungen beider Stoffe jeweils im Photometer bzgl. ihrer Extinktion bei allen Wellenlängen im sichtbaren Bereich untersucht. Die Darstellung der Extinktionswerte über den betreffenden Wellenlängenbereich nennt man Absorptionsspektrum (s. **Abbildung 2**). Aus ihm wurde abgeleitet, dass die Konzentrationsbestimmung bei Tartrazin bei einer Wellenlänge von 425 nm gemessen werden muss. Durch die Verwendung spezifischer Wellenlängen, die dem Absorptionsmaximum des Stoffes entsprechen, können auch verschiedene Stoffe in einer Mischung nebeneinander analysiert werden. Die Konzentrationsbestimmung mit dem Spektralphotometer erfolgt analog der Konzentrationsbestimmung, die du mit dem Smartphotometer schon durchgeführt hast. Es werden auch hier zunächst die Werte für Lösungen bekannter Konzentrationen bestimmt ("Eichlösungen" oder "Kalibrierlösungen"). Aus diesen ergibt sich dann eine Eichgerade, mit deren Hilfe die Konzentration der unbekannt Probe ermittelt werden kann. Die Messung ergab im trinkfertigen Getränk für Tartrazin eine Konzentration von 19 mg/l.

**Aufgaben**

3. Erkläre den Begriff „Extinktion“.

Ti

4. Erläutere am Beispiel der Untersuchung von Tartrazin, wie man mit dem Photometer ein Absorptionsspektrum (s. **Abbildung 2**) aufnimmt und wie man eine Konzentrationsbestimmung durchführt.

Ti

5. Vergleiche das Spektralphotometer mit dem „Smartphotometer“. Beschreibe die Unterschiede.

Ti


Vertiefungsaufgabe

Ti

Speichern & zurück

🔒

Abbildung 112: HyperDoc der 3. Stunde der Unterrichtsreihe (Oberstufe). Verändert nach Nieß et al. (o. D.).




Arbeitsblätter


Eingeloggt als test | Logout

---

**Arbeitsblatt: Optimierte Farbanalyse mit dem Smartphone**

**Einstieg**

Wiederholung Vergleich Smartphotometer - Spektralphotometer: 

1. Nenne konkrete Möglichkeiten, wie die Bestimmung der Azorubin-Konzentration mit dem ‚Smartphotometer‘ noch weiter optimiert werden könnte. Berachte dir dazu die zuvor erarbeiteten Unterschiede zwischen dem Smartphotometer und dem Spektralphotometer. 


Hinweis: Die App, die die RGB-Werte misst, kann auch parallel eine Referenz vermessen. Die App, die das Durchlicht erzeugt, bietet noch weitere Farben.

**T!**

---

**Versuch**

Du kannst dir hier nochmal die Versuchsdurchführung anschauen, anhören oder durchlesen:



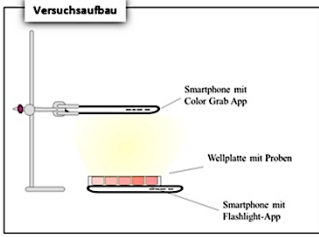
---

**Tabelle Grün-Werte**

2. Übertrage die Grünwerte in die Tabelle.

Konzentration [mg/l]	Grün-Werte
20	<input style="width: 90%;" type="text"/>
30	<input style="width: 90%;" type="text"/>
40	<input style="width: 90%;" type="text"/>
50	<input style="width: 90%;" type="text"/>
60	<input style="width: 90%;" type="text"/>
verd. Sirup (Probe)	<input style="width: 90%;" type="text"/>


Versuchsaufbau

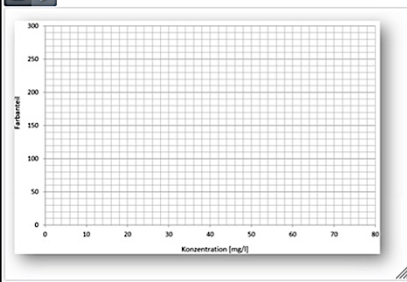


---

**Diagramm Grün-Werte**

3. Übertrage die Grünwerte in den Graphen und bestimme die Konzentration der Probe.






**Konzentration Probe:**

**T!**


---

**Aufgaben**

4. Berechne die Menge des fertigen Getränks und die Anzahl an Gläsern (200 ml), die entsprechend des ADI-Wertes (4,0 mg/kg Körpergewicht und Tag) für Azorubin täglich von dir getrunken werden dürfe.



**T!**

Vorfahrungsaufgabe 

**T!**

Speichern & zurück




Abbildung 113: HyperDoc der 4. Stunde der Unterrichtsreihe (Oberstufe). Verändert nach Nieß et al. (o. D.).

## 9.1.4 Arbeitsblatt 1. Stunde Vergleichsgruppe






Erster und zweiter Buchstabe des Vornamens deiner Mutter	Tag	Zweiter	Letzter
Erster		Zweiter	Letzter

**"Safety First!"**

**Einführung:**

"Safety First!" (engl. *Sicherheit zuerst*) - so heißt es oftmals im Alltag, wenn Gefahren drohen. Aber auch in der Schule gilt der Spruch. Gerade beim Experimentieren in Naturwissenschaften können gefährliche Situationen entstehen. Sei es durch das Experimentieren mit Elektrizität im Physikunterricht oder durch Chemikalien im Chemieunterricht.

Im Gegensatz zu „normalen“ Arbeitsblättern, findet Ihr bei diesen Arbeitsblättern zu jeder Aufgabe in der Regel eine Hilfe in zwei verschiedenen Darstellungen:

Eine Hilfe in Form von Text  oder in Form von Bild . Ein farblicher Kasten wie z.B.  oder  oder  markiert zudem die passende Hilfe zu der jeweiligen Aufgabe.


Außerdem wird es auf jedem Arbeitsblatt eine Vertiefungsaufgabe geben, die Ihr bearbeiten könnt. Markiert bitte jedes Mal, wenn Ihr euch die Hilfen oder Vertiefungsaufgabe betrachtet. Dazu macht Ihr einfach ein Kreuz in das vorhandene Feld - Beispiel:


Betrachtet:

**Gefahrenstoffe im Alltag:**

In unserem Alltag befinden sich eine Vielzahl von verschiedenen Gefahrenstoffen, auch wenn es auf den ersten Blick nicht so scheint.

Hier siehst du zum Beispiel die Rückseite eines Geschirrspülmittels. Dort steht geschrieben: "Verursacht schwere Augenreizungen. Darf nicht in die Hände von Kindern gelangen [...]". Gleichzeitig ist ein großes Symbol mit einem Ausrufezeichen zu sehen:










	<p><b>Achtung</b></p> <p>Geschir-Reiniger-Tablets. Verursacht schwere Augenreizung. Darf nicht in die Hände von Kindern gelangen. Ist ärztlicher Beistand erforderlich, Verpackung oder Kennzeichnung ablesen. BEI KONTAKT MIT DEN AUGEN: Einige Minuten lang behutsam mit Wasser spülen. Nach Möglichkeit: Kontaktlinsen entfernen, wenn sie einsetzbar sind. Bei anhaltender Augenreizung: Weiterhin ärztliche Hilfe herbeiführen. ENTHÄLT SODIUMHYDROXYD. Bei Unwohlsein GIFTINFORMATIONEN. Bei allergischen Reaktionen hervorheben.</p>
---	---

	<p><b>Gefahr</b></p> <p>Verursacht schwere Augenreizung. Darf nicht in die Hände von Kindern gelangen. Ist ärztlicher Beistand erforderlich, Verpackung oder Kennzeichnung ablesen. BEI KONTAKT MIT DEN AUGEN: Einige Minuten lang behutsam mit Wasser spülen. Nach Möglichkeit: Kontaktlinsen entfernen, wenn sie einsetzbar sind. Bei anhaltender Augenreizung: Weiterhin ärztliche Hilfe herbeiführen. ENTHÄLT SODIUMHYDROXYD. Bei Unwohlsein GIFTINFORMATIONEN. Kann allergische Reaktionen hervorrufen.</p>
---	--

Das war nur ein Beispiel. Du findest viele weitere Gefahrenstoffhinweise auf allen möglichen alltäglichen Verbrauchsgegenständen. Insgesamt gibt es neun Kategorisierungen für Gefahrenstoffe. Jede Kategorie wird dabei durch ein spezielles Symbol gekennzeichnet (s. Tabelle: nächste Seite).

[https://de.wikipedia.org/wiki/Global\\_harmonisiertes\\_System\\_zur\\_Einstufung\\_und\\_Kennzeichnung\\_von\\_Chemikalien](https://de.wikipedia.org/wiki/Global_harmonisiertes_System_zur_Einstufung_und_Kennzeichnung_von_Chemikalien). Die Texte stehen unter der Lizenz: Creative Commons Attribution-ShareAlike 3.0 Unported

Tabelle 1: Gefahrsymbole und ihre Bedeutung

Symbol	Beschreibung	Beispiele für gefährliche Eigenschaften
	Explodierende Bombe	instabile explosive Stoffe
	Flamme	entzündbare Stoffe
	Flamme über Kreis	entzündende Stoffe
	Gasflasche	Gas unter Druck
	Ätzwirkung	schwere Augenschädigungen, hautätzend
	Totenkopf mit gekreuzten Knochen	akute Toxizität
	dickes Ausrufezeichen	hautreizend, augenreizend
	Gesundheitsgefahr	diverse Gesundheitsgefahren
	Umwelt	gewässergefährdend

**Aufgaben:**

1. Vervollständige das Gefahrensymbol "Dickes Ausrufezeichen".




Abbildung 114: Arbeitsblatt (Teil 1) der 1. Stunde der Unterrichtsreihe in der Vergleichsgruppe (Mittel- und Oberstufe). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008).

The worksheet is divided into two main sections. The top section, labeled '2.', contains the instruction '2. Folgende Stoffe sind gegeben. Ordne jedem Stoff ein Gefahrensymbol zu.' (2. The following substances are given. Assign a hazard symbol to each substance). Below this instruction are three columns labeled 'a)', 'b)', and 'c)'. Column 'a)' contains a photograph of two gas cylinders. Column 'b)' contains a box labeled 'Natronlauge' (Sodium hydroxide). Column 'c)' contains a box labeled 'Deodorant' (Deodorant). To the right of these columns is a large empty rectangular area for drawing or writing. The bottom section, labeled '3.', contains the instruction '3. Vertiefungsaufgabe' (3. Deepening task) and a red square icon with a white symbol inside.

Abbildung 115: Arbeitsblatt (Teil 2) der 1. Stunde der Unterrichtsreihe in der Vergleichsgruppe (Mittel- und Oberstufe). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008).

### 9.1.5 Inhalte der Lernhilfen

#### Mittelstufe

*Tabelle 171: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der ersten Stunde in der Mittel- und Oberstufe.*

<i>Bezug</i>	<i>Typ</i>	<i>Inhalt</i>	<i>Didaktische Intention</i>
Aufgabe 1	Text	Verweis auf Tabelle mit Gefahrensymbolen. Beschreibung dieser beachten.	Fokussierung auf Tabelle mit Gefahrensymbolen als relevante Informationsquelle zur Bearbeitung der Aufgabe.
	Bild	Tabelle mit Gefahrensymbolen (Beschreibung und Beispiel).	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung.	
	Audio	s. Text	
Aufgabe 2	Text	Nähere Beschreibung der abgebildeten Stoffe und mögliche Gefahrenquelle.	Nähere Informationen über die Stoffe, um das Gefahrenpotential einschätzen und Gefahrensymbole wählen zu können.
	Bild	s. Text – zusätzliche Bilder zur Veranschaulichung	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung.	
	Audio	s. Text	



*Tabelle 172: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der zweiten Stunde in der Mittelstufe.*

<b>Bezug</b>	<b>Typ</b>	<b>Inhalt</b>	<b>Didaktische Intention</b>
Umrechnung	Text	Bezug zwischen Gramm und Milligramm.	Umrechnung von Gramm zu Milligramm vereinfachen.
	Bild	Visualisierung des Bezugs anhand eines Säulendiagramms.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung.	
	Audio	s. Text	
Aufgabe 1	Text	Beschreibung der Einheit „Milligramm pro Kilogramm Körpergewicht“ mit Bezug auf unterschiedliche Personen.	Verdeutlichung des Problems anhand des Vergleichs zweier unterschiedlicher Personen.
	Bild	Menschen mit verschiedenen Körpergewichten gezeigt. Bezug zur Einheit.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung.	
	Audio	s. Text	
Aufgabe 2	Text	Hinführung zur Berechnung der maximalen täglichen Verzehrmenge.	Heranführung an die Berechnung der maximalen täglichen Verzehrmenge.
	Bild	/	
	Video	/	
	Audio	s. Text	
Aufgabe 3	Text	Verweis auf ADI-Wert und Menge im Getränk in Bezug auf die beiden Farbstoffe.	Fokus auf ADI-Wert und enthaltener Menge (optisch) der Farbstoffe im Sirup, um Schädigungspotential zu bewerten
	Bild	s. Text – zusätzliches Bild des Sirups und der gelösten Farbstoffe.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung.	
	Audio	s. Text	

*Tabelle 173: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der dritten Stunde in der Mittelstufe.*

<b>Bezug</b>	<b>Typ</b>	<b>Inhalt</b>	<b>Didaktische Intention</b>
Versuchsdurchführung	Text	Schilderung der Versuchsdurchführung.	Versuch besser nachvollziehen und erneute Betrachtung ermöglichen.
	Bild	Versuchsdurchführung mit Bildern.	
	Video	Versuchsdurchführung als Video.	
	Audio	s. Text	
Aufgabe 1	Text	Beschreibung des RGB-Wertes mit Beispiel.	Zustandekommen des RGB -Wertes (R, G, B) verdeutlichen.
	Bild	s. Text – mit Bild.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung.	
	Audio	s. Text	
Aufgabe 2	Text	Schrittweise Heranführung an das Einzeichnen der Messwerte.	Selbstständiges Einzeichnen der Messwerte ermöglichen.
	Bild	s. Text – mit Bildern.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung).	
	Audio	s. Text	
Aufgabe 3	Text	Erklärung für das Einzeichnen der Kalibriergeraden.	Selbstständiges Einzeichnen der Kalibriergeraden ermöglichen.
	Bild	s. Text – mit Bildern.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung.	
	Audio	s. Text	

*Tabelle 174: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der vierten Stunde in der Mittelstufe.*

<i>Bezug</i>	<i>Typ</i>	<i>Inhalt</i>	<i>Didaktische Intention</i>
Aufgabe 1	Text	Beschreibung des Teilchenmodells.	Verständnis zum Teilchenmodell und Konzentrationsbegriff fördern.
	Bild	s. Text – mit Bild (Beschriftung)	
	Video	s. Bild – mit verbaler (gesprochener Beschreibung)	
	Audio	s. Text	
Aufgabe 2	Text	Vorgehen zur Berechnung der Verzehrmenge des Mischgetränks.	Hinführung zur selbstständigen Berechnung der maximalen Verzehrmenge des Mischgetränks.
	Bild	s. Text – mit Bildern, die die Schritte visualisieren.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung.	
	Audio	s. Text	
Aufgabe 3	Text	Beschreibung des Teilchenmodells und Verdeutlichung der Aufgabe.	Verständnis zum gezeigten Teilchenmodell fördern und Arbeitsauftrag nochmal in anderen Worten wiedergeben.
	Bild	s. Text – mit Bildern, die das Teilchenmodell erklären.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung)	
	Audio	s. Text	

**Oberstufe**

*Tabelle 175: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der zweiten Stunde in der Oberstufe.*

<i>Bezug</i>	<i>Typ</i>	<i>Inhalt</i>	<i>Didaktische Intention</i>
Versuchsdurchführung	Text	Schilderung der Versuchsdurchführung.	Versuch besser nachvollziehen und erneute Betrachtung ermöglichen.
	Bild	Versuchsdurchführung mit Bildern.	
	Video	Versuchsdurchführung als Video.	
	Audio	s. Text	
Aufgabe 1	Text	Beschreibung des RGB-Wertes mit Beispiel.	Zustandekommen des RGB -Wertes (R, G, B) verdeutlichen.
	Bild	s. Text – mit Bild.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung.	
	Audio	s. Text	
Aufgabe 2	Text	Schrittweise Heranführung an das Einzeichnen der Messwerte und das Erstellen der Kalibriergerade.	Selbstständiges Einzeichnen der Messwerte und Kalibriergeraden ermöglichen.
	Bild	s. Text – mit Bildern.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung).	
	Audio	s. Text	
Aufgabe 3	Text	Hinweis auf den Verlauf der Wertereihen und Ablesen des Y-Wertes.	Verdeutlichung des Problems, dass lediglich Reihe der Grünwerte zur Auswertung herangezogen werden kann.
	Bild	s. Text – mit Bildern.	
	Video	/	
	Audio	s. Text	

*Tabelle 176: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der dritten Stunde in der Oberstufe.*

<b>Bezug</b>	<b>Typ</b>	<b>Inhalt</b>	<b>Didaktische Intention</b>
Aufgabe 1	Text	Begriffe, die zugeordnet werden müssen	Begriffe vorgeben, damit sie nicht aus Text extrahiert werden müssen.
	Bild	/	
	Video	/	
	Audio	s. Text	
Aufgabe 3	Text	Beschreibung des Begriffs der Extinktion.	Verbesserung des Verständnisses für den Extinktionsbegriff.
	Bild	Verdeutlichung des Begriffs Extinktion anhand einer Abbildung.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung.	
	Audio	s. Text	
Aufgabe 5	Text	Beispiele für Bauteile, die verglichen werden können.	Fokus auf einzelne Bauteile, die sinnvoll miteinander verglichen werden können.
	Bild	Markierung der Bereiche beider Messvorrichtungen, die verglichen werden können.	
	Video	s. Bild – mit verbaler (gesprochener Beschreibung).	
	Audio	s. Text	

*Tabelle 177: Inhalt und didaktische Intention der verwendeten Lernhilfen auf dem HyperDoc (Interventionsgruppe: Text, Bild, Video, Audio) beziehungsweise Arbeitsblatt (Vergleichsgruppe: Text, Bild) der vierten Stunde in der Oberstufe.*

<b>Bezug</b>	<b>Typ</b>	<b>Inhalt</b>	<b>Didaktische Intention</b>
Einstieg / Wiederholung	Text	Vergleich im Aufbau zwischen Smartphotometer und Spektralphotometer.	Wiederholung des Vergleichs der beiden Aufbauten für Hinführung Aufgabe 1.
	Bild	s. Text – mit Bildern.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung.	
	Audio	s. Text	
Aufgabe 1	Text	Hinweise auf mögliche Anpassungen im Aufbau und in der App.	Fokussierung des Problems auf Farbe des Durchlichts und Weißabgleich („Blank“).
	Bild	/	
	Video	/	
	Audio	s. Text	
Versuchsdurchführung	Text	Schilderung der Versuchsdurchführung.	Versuch besser nachvollziehen und erneute Betrachtung ermöglichen.
	Bild	Versuchsdurchführung mit Bildern.	
	Video	Versuchsdurchführung als Video.	
	Audio	s. Text	
Aufgabe 3	Text	Schrittweise Heranführung an das Einzeichnen der Messwerte und das Erstellen der Kalibriergerade.	Selbstständiges Einzeichnen der Messwerte und Kalibriergeraden ermöglichen.
	Bild	s. Text – mit Bildern.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung).	
	Audio	s. Text	
Aufgabe 4	Text	Vorgehen zur Berechnung der Verzehrmenge des Mischgetränks.	Hinführung zur selbstständigen Berechnung der maximalen Verzehrmenge des Mischgetränks
	Bild	s. Text – mit Bildern, die die Schritte visualisieren.	
	Video	s. Bild – mit verbaler (gesprochener) Beschreibung.	
	Audio	s. Text	


## 9.1.6 Lernhilfen Vergleichsgruppe

### Mittelstufe

**Hilfe zu Aufgabe 1:** Betrachtet:

Du findest das Gefahrensymbol in der Tabelle. Die Beschreibung des Gefahrensymbols gibt dir den Hinweis.

**Hilfe zu Aufgabe 2:** Betrachtet:



→

Gasflaschen stehen unter Druck.

→


Manometer

Natronlauge

→

Laugen sind ätzend.

→




Deodorant

→

In Deos wird ein "Treibgas" verwendet. Dieses ist leicht entzündlich.










→



**Vertiefungsaufgabe:** Betrachtet:

Nenne weitere Gefahrenstoffe aus deinem Alltag.

**Hilfe zu Aufgabe 1:** Betrachtet:

Symbol	Beschreibung	Beispiele für gefährliche Eigenschaften
	Explosierende Bombe	instabile explosive Stoffe
	Flamme	entzündbare Stoffe
	Flamme über Kreis	entzündende Stoffe
	Gasflasche	Gas unter Druck
	Ätzwirkung	schwere Augenschädigungen, hautätzend
	Totenkopf mit gekreuzten Knochen	akute Toxizität
	dickes Ausrufezeichen	hautreizend, augenreizend
	Gesundheitsgefahr	diverse Gesundheitsgefahren
	Umwelt	gewässergefährdend

**Hilfe zu Aufgabe 2:** Betrachtet:

a) Dabei handelt es sich Druckbehälter mit Gas.  
b) Laugen sind ätzend.  
c) In Deos wird häufig ein brennbares "Treibgas" verwendet, um den Duftstoff zu versprühen.

Abbildung 116: Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 1. Stunde der Unterrichtsreihe der Vergleichsgruppe (Mittelstufe). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008). Handschuhe nach pagefact (2014). Sprühdose: Eigene Darstellung nach roegger (2014) und Skitterphoto (2015).

**Hilfe zu Aufgabe 1:** Betrachtet:

Durch die gewählte Einheit wird die täglich erlaubte Menge nicht auf eine ganze Person bezogen, sondern auf jedes Kilogramm des Körpers eines Menschen. Warum ist das sinnvoll? Bedenke, dass die Zusatzstoffe ja von unterschiedlichen Menschen aufgenommen werden (z.B. ein Kleinkind mit 10 kg und ein Bodybuilder mit 100 kg).

**Hilfe zu Aufgabe 2:** Betrachtet:

Der ADI-Wert bezieht sich immer auf das Körpergewicht einer Person. Du musst den ADI-Wert also mit dem Gewicht der Person verrechnen.

**Hilfe zu Aufgabe 3:** Betrachtet:

Vergleiche die ADI-Werte von Trarazin und Azorubin. Welcher Stoff erscheint dir schädlicher? Betrachte nun die Farbe des Grenadine-Sirups und der beiden gelösten Farbstoffe. Von welchem Farbstoff ist deutlich mehr im Sirup?

**Vertiefungsaufgabe:** Betrachtet:

Information:

Nicht nur Azofarbstoffe können eine schädigende Wirkung auf den Körper haben. Während Gewürze in der Regel in geringen Mengen gesundheitsfördernd sind, haben manche Gewürze in hohen Mengen negative Folgen für den Körper. So auch Cumarin, ein Aromastoff in Zimt. Der ETD/ADI-Wert von Cumarin liegt bei **0,1 mg/kg** Körpergewicht. Man unterscheidet generell zwischen zwei Arten von Zimt: Ceylon-Zimt und Cassia-Zimt. Cassia-Zimt enthält deutlich mehr Cumarin (bis zu 1 Prozent).


Aufgabe:

Für einen Lebkuchen werden 20 mg Cassia-Zimt verwendet. Im Lebkuchen befinden sich ca. 0,2 mg Cumarin.

**Wie viele dieser Lebkuchen dürfte ein 40 kg-schweres Kind ca. essen?**

**Hilfe zu Aufgabe 1:** Betrachtet:

Menschen mit verschiedenen Gewichten:




$\neq$   
ADI: mg/kg pro Tag

**Bodybuilder**

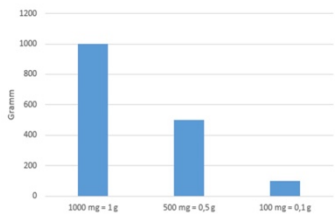
Der ADI-Wert bezieht sich auf jedes Kilogramm des Körpers eines Menschen. Warum ist das sinnvoll?

**Hilfe zu Aufgabe 3:** Betrachtet:

Betrachte dir die Farbstoffe Azorubin (links) und Trarazin (rechts). In der Mitte ist der Sirup abgebildet. Von welchem Farbstoff ist mehr im Sirup? Beachte außerdem den ADI-Wert der Farbstoffe.



**Hilfe zu Umrechnung:** Betrachtet:



**Hilfe zu Umrechnung:** Betrachtet:

1 g entspricht 1000 mg, das bedeutet 0,5 g entspricht 500 mg und 0,1 g entspricht 100 mg.

Abbildung 117: Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 2. Stunde der Unterrichtsreihe der Vergleichsgruppe (Mittelstufe). Kleinkind nach OmarMedinaFilms (2016).



**Versuchsdurchführung:**
Betrachtet:

Zunächst müssen die Kalibrierlösungen und die Getränkeprobe in eine Zellkulturplatte pipettiert werden:

- Dazu stehen Kalibrierlösungen in den Konzentrationen von 20 - 60 mg/l bereit, sowie unsere Getränkeprobe in einer 1:7 Verdünnung, so wie auf dem Etikett angegeben. Zur Erinnerung: Das bedeutet 1 Teil Sirup und 6 Teile Wasser.
- Die Lösungen werden mit einer Automatikpipette pipettiert. Diese hat ein fest eingestelltes Volumen von 1 ml.

Nachdem die fünf Kalibrierlösungen und unsere Probe in die Vertiefungen der Zellkulturplatte pipettiert wurden, muss ein Stativ und zwei Tablets aufgebaut werden:


- Ein Tablet dient als Lichtquelle, das andere als eigentliches Messinstrument.
- Auf dem unteren Tablet wird eine App geöffnet mit deren Hilfe das Display des Tablets auf eine bestimmte Farbe eingestellt werden kann. Wir wählen die Farbe „Weiß“.
- Danach wird die Zellkulturplatte auf dem Display positioniert.
- Auf dem oberen Tablet öffnen wir die App zur Messung.
- Wir müssen nun die jeweilige Lösung unter dem Messfenster der App positionieren. Mit einem Klick auf das Display wird der Messwert gespeichert.
- Wir vermessen so alle Lösungen. Am Ende können wir uns eine Übersicht der Messwerte anzeigen lassen.

**Hilfe zu Aufgabe 1:**
Betrachtet:

Mit dem RGB-Wert lassen sich Farben ausdrücken. Ein RGB-Wert von z.B. (100, 55, 125) bedeutet, dass der Farbton sich aus 100 Rot, zu 55 aus Grün und zu 125 aus Blau zusammensetzt.

**Hilfe zu Aufgabe 1:**
Betrachtet:

Beispiel für den RGB-Wert (100, 55, 125):



**Hilfe zu Aufgabe 2:**
Betrachtet:


- Der Graph zeigt die Konzentration in Abhängigkeit des Farbanteils.
- Markiere die Werte einer Wertreihe (Rot, Grün, Blau) mit einem kleinen Kreuz oder Punkt der entsprechenden Farbe. Du musst dazu den Y-Wert (Farbanteil) auf Höhe des X-Wertes (Konzentration) setzen.
- Du erhältst so im Diagramm 5 rote, 5 grüne und 5 blaue Punkte.

**Hilfe zu Aufgabe 3:**
Betrachtet:

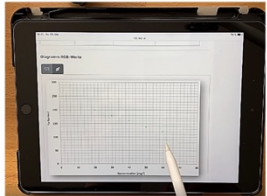
- Die Messpunkte folgen in etwa einer Geraden.
- Du kannst nun versuchen, durch alle Messpunkte eine Linie zu ziehen.
- Markiere den R-, G- und B-Wert in einer entsprechenden Farbe und vergleiche mit der Linie.
- Mithilfe der Geraden kannst du auch unbekannte X-Werte zu einem gegebenen Y-Wert ablesen.

**Hilfe zu Aufgabe 2:**
Betrachtet:

- Der Graph zeigt die Konzentration in Abhängigkeit des Farbanteils.
- Der Farbanteil ist auf der Y-Achse aufgetragen, die Konzentration auf der X-Achse. Hier wurde z.B. für eine Konzentration von 20 mg/l ein Grünwert von 200 gemessen.



- Du erhältst so einen Graphen mit 5 roten, 5 grünen und 5 blauen Messpunkten.



**Hilfe zu Aufgabe 3:**
Betrachtet:

Markiere jeweils die R-, G- und B-Werte der Getränkeprobe mit einem kleinen Punkt in der betreffenden Farbe. Vergleiche die Lage des Punktes mit den Werten der bekannten Lösungen. Die Gerade hilft dir dabei, eine Konzentration abzulesen, wenn der Wert der Probe zwischen zwei bekannten Konzentrationen liegt.

**Vertiefungsaufgabe:**
Betrachtet:


Erkläre, welche Wertreihe (Rot, Grün, Blau) du am ehesten für die Konzentrationsbestimmung verwenden würdest.

*Abbildung 118: : Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 3. Stunde der Unterrichtsreihe der Vergleichsgruppe (Mittelstufe).*

**Hilfe zu Aufgabe 1:** Betrachtet:

Die farblosen weißen Punkte stellen die Wasserteilchen dar.  
Die schwarzen Punkte stellen die Teilchen des Azorubins dar.  
Die Rechtecke stehen jeweils für ein Volumen. Es kommt bei Konzentrationen auf die Menge an Farbstoff an, die in einem bestimmten Volumen gelöst ist. Zur Angabe bezieht man sich immer auf einen Liter Wasser als Bezugsgröße.

**Hilfe zu Aufgabe 1:** Betrachtet:



Die Rechtecke stehen jeweils für ein Volumen. Es kommt bei Konzentrationen auf die Menge an Farbstoff an, die in einem bestimmten Volumen gelöst ist. Zur Angabe bezieht man sich immer auf einen Liter Wasser als Bezugsgröße.

**Hilfe zu Aufgabe 2:** Betrachtet:

Um die maximale Anzahl an Gläsern zu berechnen, gehe wie folgt vor:


1. Berechne zunächst den ADI-Wert für dein eigenes Körpergewicht.
2. Vergleiche diesen Wert mit der Konzentration von Azorubin im Getränk. Zur Erinnerung: Die Konzentration von Azorubin in der Probe/Getränk haben wir im Experiment bestimmt.
3. Berechne nun, wie viele Gläser du von dem Getränk maximal trinken dürftest, bis der ADI-Wert überschritten ist.

**Hilfe zu Aufgabe 2:** Betrachtet:

1. ADI-Wert für eigenes Körpergewicht berechnen

2. Mit Konzentration von Azorubin im Getränk vergleichen

3. Maximale Anzahl an Gläsern (0,2 Liter) unter Beachtung des eigenen ADI-Wertes berechnen

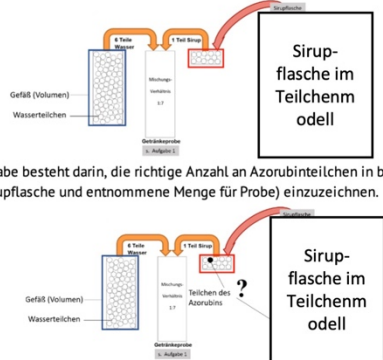


**Hilfe zu Aufgabe 3:** Betrachtet:

Du siehst auf der linken Seite den Anteil des Wassers, auf der rechten Seite den Teil des Sirups, der für die Getränkeprobe benötigt wird und die gesamte Sirupflasche.  
In der Mitte befindet sich unsere Probe: Die Angabe der Abbildung *Mischungsverhältnis 1:7* ist wie folgt zu verstehen: Setzt sich die Mischung aus insgesamt 7 Teilen zusammen, sind es 6 Teile Wasser und 1 Teil Sirup.  
Die Rechtecke stellen erneut das Gefäß bzw. das Volumen dar. Die weißen bzw. farblosen Kreise sind Wasserteilchen.  
Deine Aufgabe besteht nun darin, die richtige Anzahl an Teilchen des Azorubins in der Sirupflasche und im entnommenen Teil einzuzichnen. Die Konzentration von Azorubin in der Probe kannst du aus Aufgabe 1 entnehmen.

**Hilfe zu Aufgabe 3:** Betrachtet:

Erklärung der Abbildung: Aus der Sirupflasche wird ein Teil Sirup entnommen, um die Getränkeprobe herzustellen.



Deine Aufgabe besteht darin, die richtige Anzahl an Azorubinteilchen in beide Gefäße (Sirupflasche und entnommene Menge für Probe) einzuzichnen.

**Vertiefungsaufgabe:** Betrachtet:

Ein Kleinkind (10 kg) trinkt in einem unbeobachteten Moment 0,1 Liter (ein kleines Glas) des unverdünnten Sirups.  
**Berechne, ob das Kind den ADI-Wert überschritten hat.**


Abbildung 119: : Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 4. Stunde der Unterrichtsreihe der Vergleichsgruppe (Mittelstufe). Hilfe Aufgabe 2: Eigene Darstellung nach Open-Clipart-Vectors (2016, 2017) und JJuni (2016), Teilchenmodell Hilfe 3: Eigene Darstellung nach Czubatinski et al. (2020).

## Oberstufe

**Hilfe zu Aufgabe 1:** Betrachtet:

Du findest das Gefahrensymbol in der Tabelle. Die Beschreibung des Gefahrensymbols gibt dir den Hinweis.

**Hilfe zu Aufgabe 2:** Betrachtet:



Gasflaschen stehen unter Druck.


→

Manometer

Natronlauge

→


Laugen sind ätzend.












Deodorant

→

In Deos wird ein „Treibgas“ verwendet. Dieses ist leicht entzündlich.



**Hilfe zu Aufgabe 1:** Betrachtet:

Symbol	Beschreibung	Beispiele für gefährliche Eigenschaften
	Explosierende Bombe	instabile explosive Stoffe
	Flamme	entzündbare Stoffe
	Flamme über Kreis	entzündende Stoffe
	Gasflasche	Gas unter Druck
	Ätzwirkung	schwere Augenschädigungen, hautätzend
	Totenkopf mit gekreuzten Knochen	akute Toxizität
	dickes Ausrufezeichen	hautreizend, augenreizend
	Gesundheitsgefahr	diverse Gesundheitsgefahren
	Umwelt	gewässergefährdend

**Hilfe zu Aufgabe 2:** Betrachtet:

a) Dabei handelt es sich Druckbehälter mit Gas.  
b) Laugen sind ätzend.  
c) In Deos wird häufig ein brennbares "Treibgas" verwendet, um den Duftstoff zu versprühen.

**Vertiefungsaufgabe:** Betrachtet:

Nenne weitere Gefahrenstoffe aus deinem Alltag.

Abbildung 120: Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 1. Stunde der Unterrichtsreihe der Vergleichsgruppe (Oberstufe). Gas-flaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008). Handschuhe nach pagefact (2014). Sprühdose: Eigene Darstellung nach roegger (2014) und Skitterphoto (2015).

**Versuchsdurchführung:**
Betrachtet:

Zunächst müssen die Kalibrierlösungen und die Getränkeprobe in eine Zellkulturplatte pipettiert werden:

- Dazu stehen Kalibrierlösungen in den Konzentrationen von 20 - 60 mg/l bereit, sowie unsere Getränkeprobe in einer 1:7 Verdünnung, so wie auf dem Etikett angegeben. Zur Erinnerung: Das bedeutet 1 Teil Sirup und 6 Teile Wasser.
- Die Lösungen werden mit einer Automatikpipette pipettiert. Diese hat ein fest eingestelltes Volumen von 1 ml.

Nachdem die fünf Kalibrierlösungen und unsere Probe in die Vertiefungen der Zellkulturplatte pipettiert wurden, muss ein Stativ und zwei Tablets aufgebaut werden:


- Ein Tablet dient als Lichtquelle, das andere als eigentliches Messinstrument.
- Auf dem unteren Tablet wird eine App geöffnet mit deren Hilfe das Display des Tablets auf eine bestimmte Farbe eingestellt werden kann. Wir wählen die Farbe „Weiß“.
- Danach wird die Zellkulturplatte auf dem Display positioniert.
- Auf dem oberen Tablet öffnen wir die App zur Messung.
- Wir müssen nun die jeweilige Lösung unter dem Messfenster der App positionieren. Mit einem Klick auf das Display wird der Messwert gespeichert.
- Wir vermessen so alle Lösungen. Am Ende können wir uns eine Übersicht der Messwerte anzeigen lassen.

**Hilfe zu Aufgabe 1:**
Betrachtet:

Mit dem RGB-Wert lassen sich Farben ausdrücken. Ein RGB-Wert von z.B. (100, 55, 125) bedeutet, dass der Farbton sich aus 100 Rot, zu 55 aus Grün und zu 125 aus Blau zusammensetzt.

**Hilfe zu Aufgabe 1:**
Betrachtet:

Beispiel für den RGB-Wert (100, 55, 125):

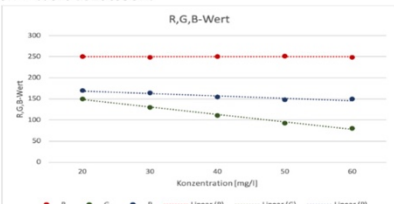


**Hilfe zu Aufgabe 2:**
Betrachtet:

- Markiere die Werte mit einem kleinen Kreuz der entsprechenden Farbe. Du musst dazu den Y-Wert (Farbanteil) auf Höhe des X-Wertes (Konzentration) setzen.
- Du erhältst so im Diagramm 5 grüne Punkte.
- Ziehe je eine Gerade durch alle Punkte.
- Dadurch kannst du einfacher zu einem Y-Wert einen (unbekannten) X-Wert ablesen.

**Hilfe zu Aufgabe 3:**
Betrachtet:

Bei welcher Wertereihe fällt es dir leichter, eine Konzentration zu einem gegebenen Y-Wert abzulesen?



**Versuchsdurchführung:**
Betrachtet:

Hier siehst du eine Bildfolge des Versuchs:

- Versuchsmaterialien: Kalibrierlösungen, Zellkulturplatte, Automatikpipette.
- Pipettiervorgang
- Pipettiervorgang beendet.
- Aufbau von Stativ und Tablet zur Messung.
- Positionierung der Zellkulturplatte über weißem Display des unteren Tablets.
- Messung der Lösungen durch App auf oberem Tablet.
- Messwerte im Überblick.

**Hilfe zu Aufgabe 2:**
Betrachtet:

- Du musst zunächst die Messwerte in den Graphen eintragen:
- Der Farbanteil ist auf der Y-Achse aufgetragen, die Konzentration auf der X-Achse. Hier würde z.B. für eine Konzentration von 20 mg/l ein Grünwert von 200 gemessen.
- Am Ende folgend alle Punkte in etwa einer Geraden.
- Danach kannst du mit deinem Lineal eine Linie durch alle Punkte einer Farbe ziehen.
- Dadurch kannst du auch unbekannte X-Werte zu einem gegebenen Y-Wert ablesen.

**Hilfe zu Aufgabe 3:**
Betrachtet:

Bei welcher Wertereihe fällt es dir leichter, zu einem Punkt der Messwerte der Getränkeprobe auf der x-Achse eine Konzentration abzulesen? Betrachte dir auch den Verlauf der Messwerte.

**Vertiefungsaufgabe:**
Betrachtet:

Erkläre auf Teilchenebene, wieso die gewählte Wertereihe aus Aufgabe 3 sich hierfür am besten eignet.

*Hinweis: Überlege dir, warum ein Stoff oder Gegenstand in einer bestimmten Farbe erscheint.*

Abbildung 121: Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 2. Stunde der Unterrichtsreihe der Vergleichsgruppe (Oberstufe).

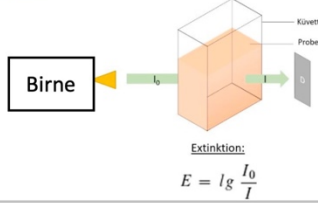
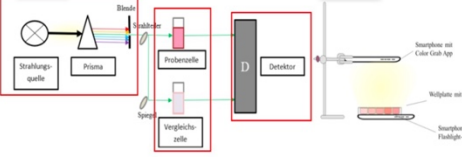
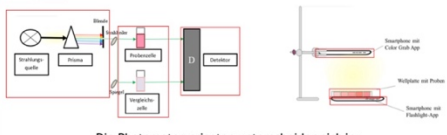
<p><b>Hilfe zu Aufgabe 1:</b> <span style="float: right;">Betrachtet: <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></span></p> <p>Folgende Begriffe müssen zugeordnet werden: Strahlungsquelle, Detektor/Photozelle, Prisma, Probenzelle, Vergleichszelle.</p>	<p><b>Hilfe zu Aufgabe 3:</b> <span style="float: right;">Betrachtet: <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></span></p>  <p style="text-align: center;">Extinktion: <math>E = \lg \frac{I_0}{I}</math></p>
<p><b>Hilfe zu Aufgabe 3:</b> <span style="float: right;">Betrachtet: <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></span></p> <p>Die Extinktion ist das logarithmierte Verhältnis zwischen der eingehenden Intensität <math>I_0</math> eines Lichtstrahls und der ausgehenden Intensität <math>I</math> (nach Durchstrahlung der Probe) des gleichen Lichtstrahls.</p>	<p><b>Hilfe zu Aufgabe 5:</b> <span style="float: right;">Betrachtet: <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></span></p> <p>Vergleiche die rot markierten Bereiche des Spektralphotometers mit dem Smartphotometer.</p> 
<p><b>Hilfe zu Aufgabe 5:</b> <span style="float: right;">Betrachtet: <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></span></p> <p>Vergleiche zum Beispiel die Strahlungsquelle beider Messmethoden, die verschiedenen Messzellen oder den Detektor.</p>	<p><b>Vertiefungsaufgabe:</b> <span style="float: right;">Betrachtet: <input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/><input type="checkbox"/></span></p> <p><b>Folgende Situation ist gegeben:</b></p> <p>Es wurde eine photometrische Konzentrationsbestimmung von Tartrazin durchgeführt. Dazu wurde für eine Kalibrierlösung mit einer bekannten Konzentration von 20 mg/l eine Extinktion von 0,234 gemessen.</p> <p>Danach wurde eine Probe aus einem Getränkekonzentrat wie folgt hergestellt:</p> <p>20 ml des Konzentrats wurden in Wasser in einem 100 ml Messkolben gelöst. Für die Probe wurde anschließend eine Extinktion von 0,129 gemessen.</p> <p><b>Berechne die Konzentration von Tartrazin in dem unverdünnten Getränkekonzentrat.</b></p>

Abbildung 122: Lernhilfen und Vertiefungsaufgabe des Arbeitsblattes der 3. Stunde der Unterrichtsreihe der Vergleichsgruppe (Oberstufe). Aufbau Photometer verändert nach Nieß et al. (o. D.).



**Einstieg:** Betrachtet:



Die Photometervarianten unterscheiden sich in:

Strahlungsquelle: Monochromatisches Licht beim Spektrophotometer und polychromatisches aus mehreren Farbtönen beim Smartphotometer.

Vergleichszelle: Beim Smartphotometer nicht vorhanden.

Detektor: Spezialisierter Detektor beim Spektrophotometer. Kamera beim Smartphotometer (zweckentfremdet)

**Einstieg:** Betrachtet:

Die Photometervarianten unterscheiden sich in folgenden Dingen: Strahlungsquelle, Vergleichszelle und Detektor.

- Strahlungsquelle: Beim Spektrophotometer wird monochromatisches Licht, also Licht einer spezifischen Wellenlänge auf die Probe projiziert. Beim Smartphotometer ist das nicht der Fall. Dort setzt sich das weiße Licht des Displays aus mehreren Farbtönen zusammen.
- Vergleichszelle: Die Vergleichszelle entfällt beim Smartphotometer. Sie dient normalerweise zur Messung einer Referenzlösung, wie z.B. dem Lösungsmittel.
- Detektor: Der Detektor dient beim Spektrophotometer ausschließlich zur Messung der Lichtintensität und ist dahingehend optimiert. Beim Smartphotometer verwenden wir die Kamera, welche in der Regel zum Erstellen von Bildern und Videos dient.

**Versuchsdurchführung:** Betrachtet:

Für die Getränkeprobe dient eine Tartrazinlösung als Referenzlösung bzw. Vergleichslösung:

- Die Tartrazinlösung befindet sich in der Vertiefung neben der Getränkeprobe.
- Die Tartrazinlösung besitzt die gleiche Konzentration wie in der Getränkeprobe.

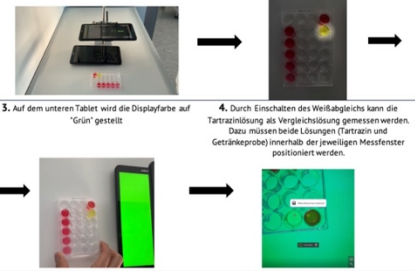
Die Messung erfolgt ähnlich zum ersten Versuch. Die Änderungen sind kursiv geschrieben:

- Auf dem unteren Tablet wird die Displayfarbe auf "Grün" eingestellt.
- Die Zellkulturplatte wird auf dem Display positioniert.
- Mit dem oberen Tablet werden erneut der RGB-Wert gemessen.
- Für die Kalibrierlösungen erfolgt die Messung analog zum ersten Versuch.
- Für die Getränkeprobe wird in der App der sogenannte "Weißabgleich" aktiviert. Dadurch wird ein zusätzliches Messfenster aktiviert.
- Beide Messfenster der App müssen sich nun über der Getränkeprobe und der Tartrazinlösung (Referenz-/Vergleichszelle) befinden

**Versuchsdurchführung:** Betrachtet:

Hier siehst du eine Bildfolge des Versuchs:

- Zellkulturplatte mit ergäuter Tartrazinlösung
- Nahaufnahme der Zellkulturplatte. Hervorgehoben die gleiche Konzentration wie in der Getränkeprobe), Vergleichszelle mit Tartrazin.



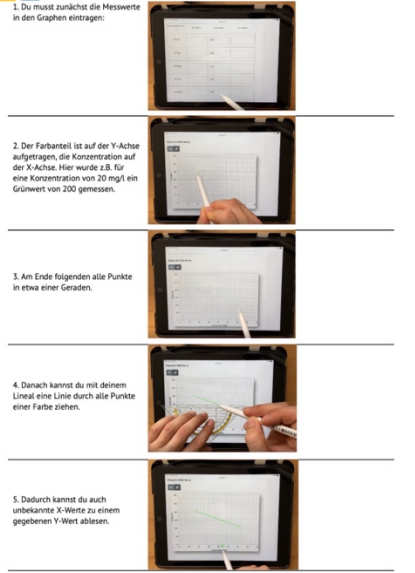
- Auf dem unteren Tablet wird die Displayfarbe auf "Grün" gestellt
- Durch Einschalten des Weißabgleichs kann die Tartrazinlösung als Vergleichslösung gemessen werden. Dazu müssen beide Lösungen (Tartrazin und Getränkeprobe) innerhalb der jeweiligen Messfenster positioniert werden.

**Hilfe zu Aufgabe 1:** Betrachtet:

- Die App Color Grap (mit der die Lösungen mit dem Smartphone vermessen werden) bietet die Möglichkeit einen sog. "Weißabgleich" als Referenzwert zu messen. Welche Lösung bietet sich hierfür an?
- Die App Flashscreen (mit der das Durchlicht erzeugt wird) kann auch noch weitere Farben erzeugen. Welche Farbe bietet sich hierfür an? Es sollte die Farbe verwendet werden, die am stärksten absorbiert wird.

**Hilfe zu Aufgabe 3:** Betrachtet:


- Markiere die Werte mit einem kleinen Kreuz der entsprechenden Farbe. Du musst dazu den Y-Wert (Farbanteil) auf Höhe des X-Wertes (Konzentration) setzen.
- Du erhältst so im Diagramm 5 grüne Punkte.
- Ziehe je eine Gerade durch alle Punkte.
- Dadurch kannst du einfacher zu einem Y-Wert einen (unbekannten) X-Wert ablesen.



- Du musst zunächst die Messwerte in den Graphen eintragen:
- Der Farbanteil liegt auf der Y-Achse aufgetragen, die Konzentration auf der X-Achse. Hier wurde z.B. für eine Konzentration von 20 mg/l ein Grünwert von 200 gemessen.
- Am Ende folgenden alle Punkte in etwa einer Geraden.
- Danach kannst du mit deinem Lineal eine Linie durch alle Punkte einer Farbe ziehen.
- Dadurch kannst du auch unbekannte X-Werte zu einem gegebenen Y-Wert ablesen.

**Hilfe zu Aufgabe 4:** Betrachtet:

- ADI-Wert für eigenes Körpergewicht berechnen
- Mit Konzentration von Azorubin im Getränk vergleichen
- Maximale Anzahl an Gläsern (0,2 Liter) unter Beachtung des eigenen ADI-Wertes berechnen



Berechne zunächst, wie viel Azorubin du pro Tag zu dir nehmen darfst. Berechne anschließend, wie viel Liter des Getränks dies entspricht. Rechne zuletzt die Anzahl der Liter in Gläser um.

**Vertiefungsaufgabe:** Betrachtet:

Berechne die Konzentration von Azorubin im unverdünnten Sirup. Beachte dazu das Etikett:

Sirupflasche „Mischungsverhältnis 1:7“

Abbildung 123: Lernhilfen des Arbeitsblattes der 4. Stunde der Unterrichtsreihe der Vergleichsgruppe (Oberstufe). Photometer verändert nach Nieß et al. (o. D.). Hilfe 4. Eigene Darstellung nach OpenClipart-Vectors (2016, 2017) und JJuni (2016).

Seite | 468

### 9.1.7 Quellen der Unterrichtsreihe

#### Arbeitsmaterial

<b>HyperDocs</b>	<b>Quelle</b>
Stunde 1 Mittelstufe	eigene Konzeption
Stunde 2 Mittelstufe	Verändert nach Czubatinski et al. (2020): Veränderung des Designs, Ergänzung Vertiefung u. Lernhilfen
Stunde 3 Mittelstufe	Verändert nach Czubatinski et al. (2020): Veränderung des Designs, Umformulierung Aufgabe 3, Ergänzung Vertiefung u. Lernhilfen, Ergänzung Versuchsdurchführung
Stunde 4 Mittelstufe	Verändert nach Czubatinski et al. (2020): Veränderung des Designs, Ergänzung Definition „Konzentration“, Umformulierung Aufgabe 3, Aufgabe 4 als Vertiefung, Ergänzung Lernhilfen
Stunde 1 Oberstufe	eigene Konzeption
Stunde 2 Oberstufe	Verändert nach Nieß et al. (o. D.): Veränderung des Designs, Ergänzung Vertiefung u. Lernhilfen, Ergänzung Versuchsdurchführung
Stunde 3 Oberstufe	Verändert nach Nieß et al. (o. D.): Veränderung des Designs, Entfernen von Aufgabe 5, Ergänzung Vertiefung u. Lernhilfen
Stunde 4 Oberstufe	Verändert nach Nieß et al. (o. D.): Veränderung des Designs, Ergänzung Aufgabe 1 aus Stunde 3, Ergänzung Vertiefung u. Lernhilfen, Ergänzung Versuchsdurchführung

**Stundenverlauf**

<b>Lernhilfen</b>	<b>Quelle</b>
Stunde 1 Mittelstufe	eigene Konzeption
Stunde 2 Mittelstufe	Verändert nach Czubatinski et al. (2020): Aufteilung auf zwei Stunden (2 + 3), Einzelarbeit anstelle von Partnerarbeit
Stunde 3 Mittelstufe	Verändert nach Czubatinski et al. (2020): Aufteilung auf zwei Stunden (2 + 3), Lehrer:innenexperiment anstelle von Schüler:innenexperiment
Stunde 4 Mittelstufe	Verändert nach Czubatinski et al. (2020): Definition Konzentrationsbegriff, Einzelarbeit anstelle von Partnerarbeit
Stunde 1 Oberstufe	eigene Konzeption
Stunde 2 Oberstufe	Verändert nach Nieß et al. (o. D.): Lehrer:innenexperiment anstelle von Schüler:innenexperiment, Einzelarbeit anstelle von Partnerarbeit, Einzelarbeit anstelle von Gruppenarbeit
Stunde 3 Oberstufe	Verändert nach Nieß et al. (o. D.): Aufteilung auf zwei Stunden (3 + 4), Einzelarbeit anstelle von Partnerarbeit
Stunde 4 Oberstufe	Verändert nach Nieß et al. (o. D.): Aufteilung auf zwei Stunden (3 + 4), Lehrer:innenexperiment anstelle von Schüler:innenexperiment, Einzelarbeit anstelle von Gruppenarbeit



### 9.1.8 Verlaufsplan und Material der ursprünglichen Unterrichtsreihe für die Oberstufe

Thema			
„Die photometrische Untersuchung von Azofarbstoffen in Erfrischungsgetränken“			
Phase	Lerninhalt	U-Form	Medien
<b>Einstieg</b>	Grenadine-Sirup ist zur Herstellung zahlreicher Mischgetränke sehr beliebt. Seine intensive Färbung erhält er durch Azofarbstoffe, die für Kinder allerdings problematisch sein können! (Problematik Azofarbstoffe, Tartrazin, Azorubin)	LV	Realobjekt Grenadine-Sirup (Bois) Etikett OHF „Azofarbstoffe“
<b>Problem</b>	Wie können wir mit Hilfe von Farbanalysen die Konzentration der Farbstoffe in dem Sirup bestimmen?	LV UG	Handy, Color Grab, Tafel
<b>Erarb. 1</b>	<b>Bestimmung von Azorubin mittels Smartphotometer</b>  <ul style="list-style-type: none"> <li>- Funktion der App</li> <li>- Verdünnungsreihe und Eichgerade</li> <li>- Planung der Versuchsreihe</li> <li>- Bestimmung der RGB Werte für Verdünnungsreihe und Probe</li> <li>- Grafische Auswertung</li> <li>- Konzentrationsbestimmung anhand der G Werte</li> </ul>	LV  PA S.-Exp. GA	Color Grab  AB 1 Smartphotometer Vers.-Material Heft
<b>Sich. 1</b>		SV/UG	Tafel
<b>Optionales Stundenende der 1. Stunde</b>			
<b>Erarb. 2</b>	<b>Das Spektralphotometer in der Analytik</b>  <ul style="list-style-type: none"> <li>- Gesetz von Lambert-Beer</li> <li>- Aufbau des Spektralphotometers</li> <li>- Absorptionsmaxima</li> <li>- Anwendung des Spektralphotometers</li> <li>- Eichgeraden</li> <li>- „Leerwert“</li> <li>- Optional: Demo des Spektralphotometres zur Bestimmung der Tartrazin Konzentration</li> </ul>	PA	AB2 Buch S. 265
<b>Sich. 2</b>		SV	OHF AB2
<b>Erarb. 3</b>	<b>Optimierung der Analyse mit dem Smartphotometer</b>  <ul style="list-style-type: none"> <li>- Einsatz monochromatischen Lichts</li> <li>- Tartrazin-Probe als Leerwert</li> <li>- Fehlerbetrachtung</li> <li>- Berechnung der erlaubten Tagesmenge</li> <li>- Risikobewertung</li> </ul>	S.-Exp. GA	Vers.-Material Heft AB 3
<b>Sich. 3</b>		SV	OHF Tafel

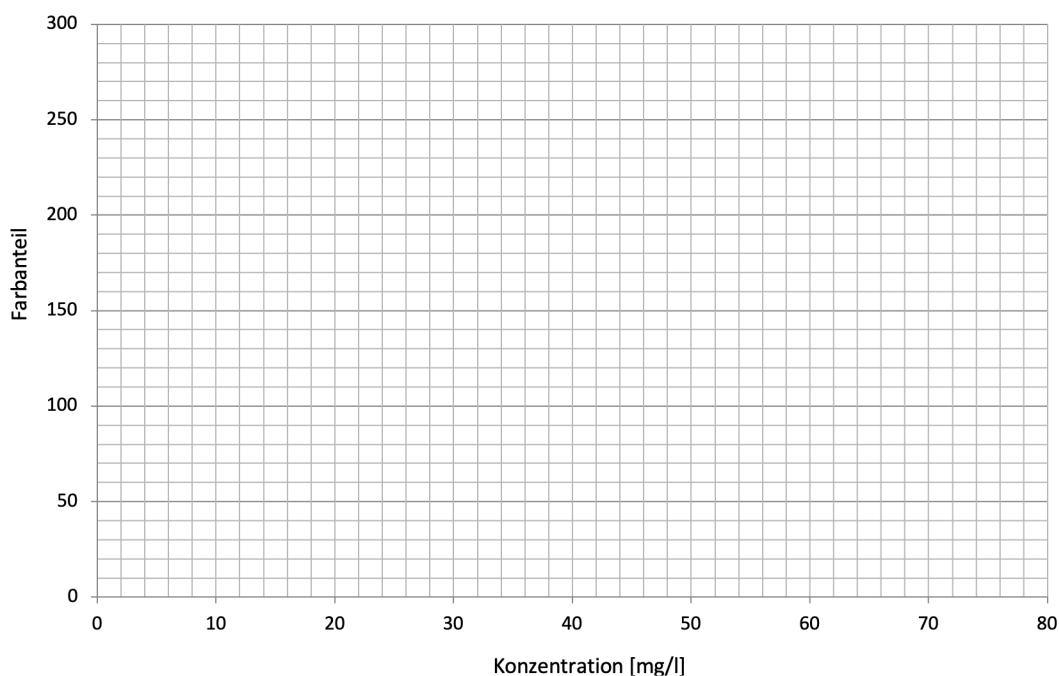
**Legende:**

PA: Partnerarbeit; UG: Unterrichtsgespräch; OHF: Overheadfolie; AB: Arbeitsblatt;  
LD: Lehrerdemo; SV: Schülervortrag; S.-Vers.: Schülerexperiment; GA:  
Gruppenarbeit; **Schulbuch:** Chemie heute SII(2016), Gesamtband, Schroedel-Verlag

Abbildung 124: Verlaufsplan zur ursprünglichen Unterrichtsreihe von Nieß et al. (2020). Unveröffentlichtes Dokument (Nieß et al., o. D.).

## Arbeitsblatt: Farbanalyse mit dem Handy

Konzentration [mg / l]	Rot-Werte	Grün Werte	Blau-Werte
20 mg/l			
30 mg/l			
40 mg/l			
50 mg/l			
60 mg/l			
<b>verd. Sirup unb. Konz.</b>			



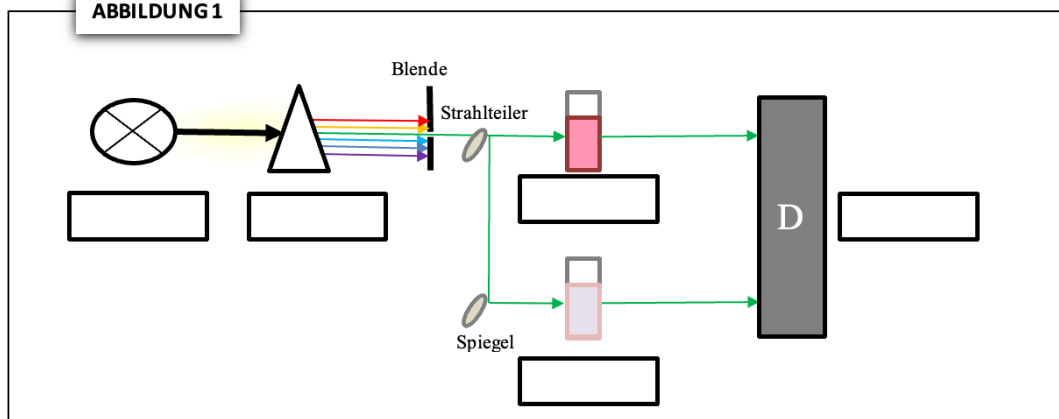
### Aufgaben:

1. Erstelle Eichgeraden für die drei RGB-Werte!
2. Bestimme den Azorubin-Gehalt möglichst genau! Welche Wertereihe bietet sich an?

Bestimmung anhand der ...	R-Werte	G-Werte	B-Werte
Konzentration:			

Abbildung 125: Arbeitsblatt der 1. Stunde der Unterrichtsreihe von Nieß et al. (2020). Unveröffentlichtes Dokument (Nieß et al., o. D.).

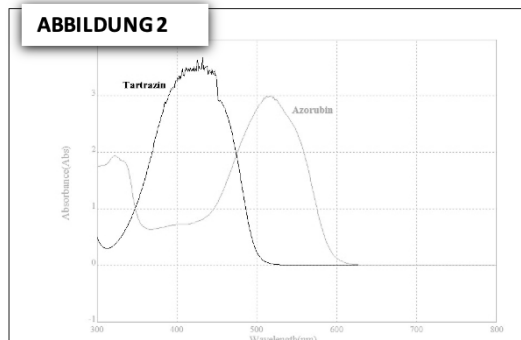
## Arbeitsblatt: Das Spektralphotometer

**ABBILDUNG 1**

**INFO**

Im Labor der Universität wurde das Spektralphotometer zunächst dazu benutzt, das Absorptionsverhalten der beiden Azofarbstoffe zu untersuchen. Dazu wurden die Lösungen beider Stoffe jeweils im Photometer bzgl. ihrer Extinktion bei allen Wellenlängen im sichtbaren Bereich untersucht. Die Darstellung der Extinktionswerte über den betreffenden Wellenlängenbereich nennt man Absorptionsspektrum (s. Abbildung 2). Aus ihm wurde abgeleitet, dass die Konzentrationsbestimmung bei Tartrazin bei einer Wellenlänge von 425 nm gemessen werden muss. Durch die Verwendung spezifischer Wellenlängen, die dem Absorptionsmaximum des Stoffes entsprechen, können auch verschiedene Stoffe in einer Mischung nebeneinander analysiert werden. Die Konzentrationsbestimmung mit dem Spektralphotometer erfolgt analog der Konzentrationsbestimmung, die du mit dem Smartphotometer schon durchgeführt hast. Es werden auch hier zunächst die Werte für Lösungen bekannter Konzentrationen bestimmt („Eichlösungen“). Aus diesen ergibt sich dann eine Eichgerade, mit deren Hilfe die Konzentration der unbekannt Probe ermittelt werden kann. Die Messung ergab im trinkfertigen Sirup für Tartrazin eine Konzentration von 19 mg/l.

**Aufgaben:**

1. Lies den Text im Buch auf S. 265 und beschrifte die Abb. 1!
2. Lies den Text im Info-Kasten und vergleiche mit der Abbildung 2!
3. Erkläre den Begriff „Extinktion“ und erläutere am Beispiel der Untersuchung von Tartrazin, wie man mit dem Photometer ein Absorptionsspektrum (S. Abbildung 2) aufnimmt und wie man eine Konzentrationsbestimmung durchführt!
4. Vergleiche das Spektralphotometer mit dem „Smartphotometer“! Beschreibe die Unterschiede!
5. Wie könnte die Bestimmung der Azorubin-Konzentration mit dem „Smartphotometer“ noch weiter optimiert werden!

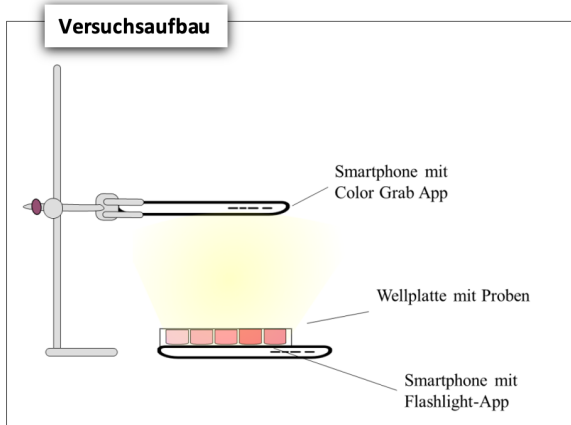
**ABBILDUNG 2**

**Tipps und Hilfen**

1. Die App Color Grab bietet die Möglichkeit einen sog. „Weißabgleich“ als Referenzwert zu messen. Welche Lösung bietet sich hier als Referenzwert an?
2. Mit „Flashscreen“ lassen sich beliebige RGB-Werte für das Display als Beleuchtungsfarbe einstellen. Welche Farbe bietet sich für die Bestimmung von Azorubin an? Betrachte dazu nochmal deine erste Messung auf dem Arbeitsblatt „Farbanalyse mit dem Handy“!
3. Die Analyse sollte mit der Wellenlänge („Farbe“) durchgeführt werden, die vom Farbstoff möglichst stark absorbiert wird.

Abbildung 126: Arbeitsblatt der 2. Stunde der Unterrichtsreihe von Nieß et al. (2020). Unveröffentlichtes Dokument (Nieß et al., o. D.).

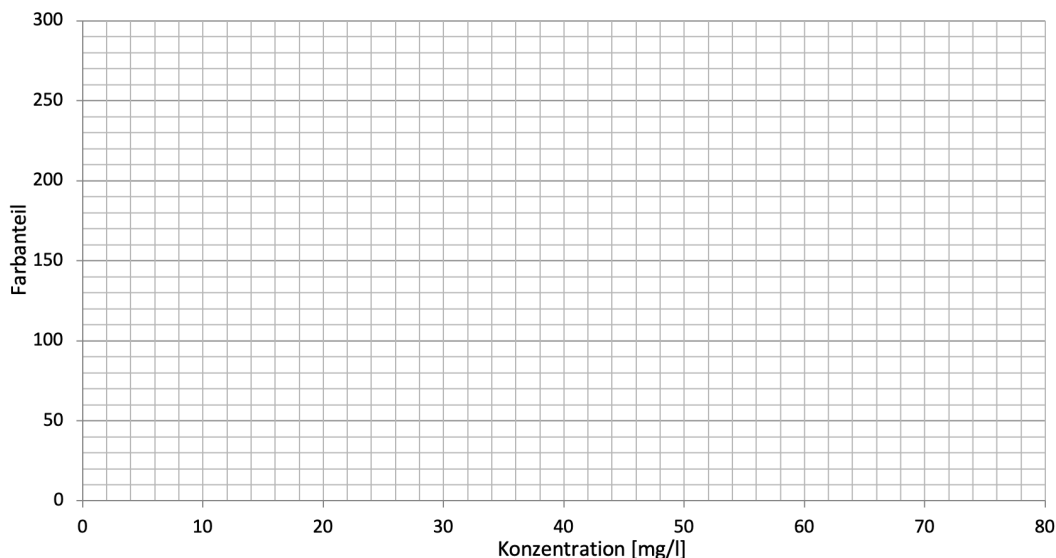
## Arbeitsblatt: Optimierte Farbanalyse mit dem Handy

Konzentration [mg / l]	Grün-Werte
20 mg/l	
30 mg/l	
40 mg/l	
50 mg/l	
60 mg/l	
<b>verd. Sirup unb. Konz.</b>	



**Aufgaben:**

1. Führe die Azorubin-Bestimmung nochmals durch (s. Versuchsaufbau)! Verwende als Referenzwert die bereitgestellte Tartrazin-Lösung ( $c = 19 \text{ mg/l}$ ). Zur Beleuchtung wird grünes Licht verwendet. Erzeuge dazu in Flashscreen eine Beleuchtung, bei der die R- bzw. B-Werte auf Null gesetzt werden und der Grün-Wert maximal eingestellt wird!
2. Berechne nun die Konzentration von Azorubin im trinkfertigen Sirup!
3. Berechne das Volumen des fertigen Getränks, das entsprechend des ADI-Wertes für Azorubin täglich von dir getrunken werden dürfte!



**Tipps und Hilfen**

1. Laut Hersteller-Empfehlung soll der Sirup im Verhältnis 1:7 mit Saft oder Wasser gemischt werden. Dies bedeutet, dass man 1 Teil Sirup zu 6 Teilen Flüssigkeit gibt. Die Mischung setzt sich also aus sieben gleichen Volumanteilen zusammen.
2. Der ADI-Wert für Azorubin beträgt  $4,5 \text{ mg/kg}$  Körpergewicht und Tag. Er gibt an, wieviel von dem Farbstoff bezogen auf jedes Kilo Körpergewicht jeden Tag maximal aufgenommen werden kann. Wechselwirkungen mit anderen Stoffen sind nur teilweise berücksichtigt!

Abbildung 127: Arbeitsblatt der 3. Stunde der Unterrichtsreihe von Nieß et al. (2020). Unveröffentlichtes Dokument (Nieß et al., o. D.).

## 9.2 Pilotierung

### 9.2.1 Arbeitsblätter der 2. Pilotierung

#### Mittelstufe

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN HyperDocSystems Arbeitsblätter Hilfen Benutzerverwaltung Monitoring FAQ Eingeloggt als fitting

### "Safety First!"

**Einführung:**  
 "Safety First!" (engl. Sicherheit zuerst) - so heißt es oftmals im Alltag, wenn Gefahren drohen. Aber auch in der Schule gilt der Spruch. Gerade beim Experimentieren in Naturwissenschaften können gefährliche Situationen entstehen. Sei es durch das Experimentieren mit Elektrizität im Physikunterricht oder durch Chemikalien im Chemieunterricht.

**Methode:**  
 Im Folgenden erhaltet ihr eine Einführung in die digitalen Arbeitsblätter. Gleichzeitig lernt ihr etwas über gefährliche Stoffe (Gefahrenstoffe) im Chemieunterricht.

Digitale Arbeitsblätter sind zunächst fast wie "normale" Arbeitsblätter. Das bedeutet, ihr erhaltet verschiedene Informationen und/oder müsst dann Aufgaben bearbeiten. Das besondere an den digitalen Arbeitsblättern: An verschiedenen Stellen sind Hilfen oder weiterführende Informationen hinterlegt. Diese zusätzlichen Inhalte öffnet ihr über Verlinkungen (blau unterstrichen) oder Symbole. Je nach Symbol unterscheidet sich die Darstellung der Hilfe:

 (Hilfe als Video),  (Hilfe als Text),  (Hilfe als Tonspur),  (Hilfe als Bild). Dabei stehen nicht immer alle Darstellungsformen für jede Hilfe zur Verfügung. Vertiefungsaufgaben sind durch das Symbol  gekennzeichnet.

**Gefahrenstoffe im Alltag:**  
 In unserem Alltag befinden sich eine Vielzahl von verschiedenen Gefahrenstoffen, auch wenn es auf den ersten Blick nicht so scheint. Hier siehst du zum Beispiel die Rückseite eines Geschirrspülmittels. Dort steht geschrieben: "Verursacht schwere Augenreizungen. Darf nicht in die Hände von Kindern gelangen [...]". Gleichzeitig ist ein großes Symbol mit einem Ausrufezeichen zu sehen:



Das war nur ein Beispiel. Du findest viele weitere Gefahrenstoffhinweise auf allen möglichen alltäglichen Verbrauchsgegenständen. Insgesamt gibt es neun Kategorisierungen für Gefahrenstoffe. Jede Kategorie wird dabei durch ein spezielles Symbol gekennzeichnet:

Symbol	Beschreibung	Beispiele für gefährliche Eigenschaften
	Explosierende Bombe	instabile explosive Stoffe
	Flamme	entzündbare Stoffe
	Flamme über Kreis	entzündende Stoffe
	Gasflasche	Gas unter Druck
	Ätzwirkung	schwere Augenschädigungen, hautätzend
	Totenkopf mit gekreuzten Knochen	akute Toxizität
	dickes Ausrufezeichen	hautreizend, augenreizend
	Gesundheitsgefahr	diverse Gesundheitsgefahren
	Umwelt	gewässergefährdend

Du kannst die Aufgaben direkt auf dem Tablet bearbeiten. Entweder du tippst die Lösung mit der virtuellen Tastatur ein oder du schreibst mit dem Apple Pencil. Drücke dafür auf das Symbol neben der Tastatur über dem Textfeld. Dir stehen außerdem verschiedene Werkzeuge zum Schreiben zur Verfügung. Deine Lösung speicherst du über den Button "Lösung speichern" am Ende des Arbeitsblattes.

**Aufgaben:**

Aufgabe	Hilfe
1) Nenne weitere Gefahrenstoffe aus deinem Alltag. Falls du nicht direkt einen Gefahrenstoff kennst, beschreibe mögliche Gefahrenquellen.	   
2) Im Folgenden siehst du verschiedene Chemikalien/Stoffe. Ordne jedem Stoff ein Gefahrensymbol zu:	   

a)  Natron-lauge Deodorant

Lösungen speichern
Arbeitsblatt drucken

Abbildung 128: HyperDoc der 1. Stunde der Unterrichtsreihe für die Mittelstufe (2. Pilotierung). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008).

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN HyperDocSystems
Arbeitsblätter Hilfen Benutzerverwaltung Monitoring
FAQ Eingelogggt als fitting

## Farbstoffe in Lebensmitteln

Auszug aus Wikipedia zu „Azofarbstoffen“ und ADI-Werten

**Lebensmittelzusatzstoffe**

Die Europäische Union (EU) vergibt eine E-Nummer für jeden zugelassenen Stoff. Insgesamt gibt es in der EU zurzeit 341 zugelassene Zusatzstoffe. Eine Zulassung wird nur erteilt, wenn keine Gesundheitsrisiken bestehen, wenn der Zusatzstoff technisch notwendig ist und wenn die Verwendung nicht zu einer Täuschung des Verbrauchers führt. Zusatzstoffe dürfen nur nach ausdrücklicher Zulassung verwendet werden. Zudem müssen Zusatzstoffe auf dem Produkt kenntlich gemacht werden.

**Rechtliche Situation**

Die EU-Kommission beauftragte am 25. März 2010 die EFSA (European Food Safety Authority), eine neuerliche gesundheitliche Bewertung fraglicher Azofarbstoffe vorzunehmen. Seit dem 20. Juli 2010 müssen Lebensmittel, die die Azofarbstoffe Tartrazin (E 102), Gelborange S (E 110), Azorubin (E 122), Allurarot (E 129) oder Cochenillerot A (E 124) enthalten, in der Europäischen Union mit dem gesonderten Warnhinweis „Kann Aktivität und Aufmerksamkeit bei Kindern beeinträchtigen“ gekennzeichnet werden.

**Azorubin**

Azorubin (auch Karmesin, E 122) ist ein als Lebensmittelfarbe zugelassener roter Azofarbstoff und das Dinatriumsalz einer Disulfonsäure. Azorubin wird für Getränke, Süß- und Zuckerwaren, Marzipan, Puddingpulver, Götterspeise, Fruchtkonserven, Fertigsuppen, Fertigprodukte, braune Sauce, Paniermehl, Färbemittel für Tabletten und anderes verwendet. Azorubin hat große chemische Ähnlichkeit mit Amaranth (E 123) und steht wie dieser Stoff im Verdacht, als Auslöser von Pseudoallergien zu wirken.

**Tartrazin**

Tartrazin (E 102) ist ein synthetischer Azofarbstoff, der als zitronengelber bis orangefarbener Lebensmittelfarbstoff eingesetzt wird. Der Stoff ist farbecht.

**Erlaubte Tagesdosis**

Die erlaubte Tagesdosis (ETD; englisch acceptable daily intake, ADI) bezeichnet die Dosis einer Substanz, wie etwa eines Lebensmittelzusatzstoffs, Pestizids oder eines Medikaments, die bei lebenslanger täglicher Einnahme als medizinisch unbedenklich betrachtet wird. Handelt es sich um ungewollte Verunreinigungen, spricht man von einer tolerierten Tagesdosis (engl. tolerable daily intake, TDI)

Angegeben wird der ETD-Wert in Milligramm bzw. Mikrogramm pro Kilogramm Körpergewicht und Tag. Synergieeffekte (Zusammenwirken verschiedener Lebensmittelzusatzstoffe) werden jedoch nur teilweise berücksichtigt.

**ADI-Wert von Azorubin:** 4 mg / kg Körpergewicht  
**ADI-Wert von Tartrazin:** 7,5 mg / kg Körpergewicht  
**Hinweis:** 1 Gramm (g) sind 1000 Milligramm (mg)

**Aufgaben:**

1. Sichte die oben zusammengestellten Informationen!
2. Begründe die Einheit des ADI-Wertes (mg/kg Körpergewicht)!
3. Berechne die Masse, die ein Kleinkind und ein Erwachsener täglich von beiden Farbstoffen aufnehmen dürfen!

Person	Azorubin [mg]	Tartrazin [mg]
Kleinkind (10 kg)		
Bodybuilder (100 kg)		

4. Du möchtest das Risiko bei der Verwendung des Grenadine-Sirups abschätzen. Der Gehalt welches Farbstoffs würdest du zuerst untersuchen? Begründe!

Lösungen speichern
 Arbeitsblatt drucken

Abbildung 129: HyperDoc der 2. Stunde der Unterrichtsreihe für die Mittelstufe (2. Pilotierung). Verändert nach Czubatinski et al. (2020).

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN HyperDocSystems    Arbeitsblätter    Hilfen    Benutzerverwaltung    Monitoring    FAQ    Eingeloggt als fitting

## Farbanalyse mit dem Handy

Konzentration [mg / l]	Rot-Werte	Grün-Wert	Blau-Werte
20 mg/l			
30 mg/l			
40 mg/l			
50 mg/l			
60 mg/l			
verd. Sirup			
unb. Konz.			

**Aufgaben:**

- Trage die mit dem Handy gemessenen Werte in die Tabelle ein!
- Übertrage die Werte der bekannten Konzentrationen in das Diagramm! Verwende entsprechende Farben!
- Versuche die Konzentration von Azorubin im Sirup abzuschätzen! Betrachte R-, G- und B-Werte!

Schätzung anhand der ...	R-Werte	G-Werte	B-Werte
Konzentration:			

Weiterer Hinweis:

Lösungen speichern    Arbeitsblatt drucken

Abbildung 130: HyperDoc der 3. Stunde der Unterrichtsreihe für die Mittelstufe (2. Pilotierung). Verändert nach Czubatinski et al. (2020).

HyperDocSystems    Arbeitsblätter    Hilfen    Benutzerverwaltung    Monitoring    FAQ, Eingeloggt als fitting

### Konzentrationen

20 mg/l	30 mg/l	40 mg/l	50 mg/l	60 mg/l	mg/l
Eichlösungen					Probe

Mischung:  
Verhältnis:  
1,7

6 Teile Wasser    1 Teil Sirup

Tip: Die Abbildung ist blau umrandet. Mit einem Klick darauf erhältst du eine Hilfe.

**Aufgaben:**

1. Stelle die Konzentration der Eichlösungen und der Probe analog dem vorgegebenen Beispiel (50 mg/l) in der Tabelle dar! Runde die Konzentration der Probe dazu auf 10 mg/l!
2. Wie viele Gläser (0,2 Liter) kannst du laut des ADI-Werts von einem Getränk mit dem Sirup täglich trinken?

Vertiefungsaufgabe 1:

Vertiefungsaufgabe 2:

Lösungen speichern    Arbeitsblatt drucken

Abbildung 131: HyperDoc der 4. Stunde der Unterrichtsreihe für die Mittelstufe (2. Pilotierung). Verändert nach Czubatinski et al. (2020).



## Oberstufe

HyperDocSystems
Arbeitsblätter
Hilfen
Benutzerverwaltung
Monitoring
FAQ Eingeloggt als fitting

### "Safety First!"

**Einführung:**  
 "Safety First!" (engl. Sicherheit zuerst) - so heißt es oftmals im Alltag, wenn Gefahren drohen. Aber auch in der Schule gilt der Spruch. Gerade beim Experimentieren in Naturwissenschaften können gefährliche Situationen entstehen. Sei es durch das Experimentieren mit Elektrizität im Physikunterricht oder durch Chemikalien im Chemieunterricht.

**Methode:**  
 Im Folgenden erhaltet ihr eine Einführung in die digitalen Arbeitsblätter. Gleichzeitig lernt ihr etwas über gefährliche Stoffe (Gefahrenstoffe) im Chemieunterricht.

Digitale Arbeitsblätter sind zunächst fast wie "normale" Arbeitsblätter. Das bedeutet, ihr erhaltet verschiedene Informationen und/oder müsst dann Aufgaben bearbeiten. Das besondere an den digitalen Arbeitsblättern: An verschiedenen Stellen sind Hilfen oder weiterführende Informationen hinterlegt. Diese zusätzlichen Inhalte öffnet ihr über Verlinkungen (blau unterstrichen) oder Symbole. Je nach Symbol unterscheidet sich die Darstellung der Hilfe:

(Hilfe als Video), 
 (Hilfe als Text), 
 (Hilfe als Tonspur), 
 (Hilfe als Bild). 
 Dabei stehen nicht immer alle Darstellungsformen für jede Hilfe zur Verfügung. Vertiefungsaufgaben sind durch das Symbol gekennzeichnet.

**Gefahrenstoffe im Alltag:**  
 In unserem Alltag befinden sich eine Vielzahl von verschiedenen **Gefahrenstoffen**, auch wenn es auf den ersten Blick nicht so scheint. Hier siehst du zum Beispiel die Rückseite eines Geschirrspülmittels. Dort steht geschrieben: "Verursacht schwere Augenreizungen. Darf nicht in die Hände von Kindern gelangen [...]". Gleichzeitig ist ein großes Symbol mit einem Ausrufezeichen zu sehen:



Das war nur ein Beispiel. Du findest viele weitere Gefahrenstoffhinweise auf allen möglichen alltäglichen Verbrauchsgegenständen. Insgesamt gibt es neun Kategorisierungen für Gefahrenstoffe. Jede Kategorie wird dabei durch ein spezielles Symbol gekennzeichnet:

Symbol	Beschreibung	Beispiele für gefährliche Eigenschaften
	Explosierende Bombe	instabile explosive Stoffe
	Flamme	entzündbare Stoffe
	Flamme über Kreis	entzündende Stoffe
	Gasflasche	Gas unter Druck
	Ätzwirkung	schwere Augenschädigungen, hautätzend
	Totenkopf mit gekreuzten Knochen	akute Toxizität
	dickes Ausrufezeichen	hautreizend, augenreizend
	Gesundheitsgefahr	diverse Gesundheitsgefahren
	Umwelt	gewässergefährdend

Du kannst die Aufgaben direkt auf dem Tablet bearbeiten. Entweder du tippst die Lösung mit der virtuellen Tastatur ein oder du schreibst mit dem Apple Pencil. Drücke dafür auf das Symbol neben der Tastatur über dem Textfeld. Dir stehen außerdem verschiedene Werkzeuge zum Schreiben zur Verfügung. Deine Lösung speicherst du über den Button "Lösung speichern" am Ende des Arbeitsblattes.

**Aufgaben:**

Aufgabe	Hilfe
1) Nenne weitere Gefahrenstoffe aus deinem Alltag. Falls du nicht direkt einen Gefahrenstoff kennst, beschreibe mögliche Gefahrenquellen.	  
2) Im Folgenden siehst du verschiedene Chemikalien/Stoffe. Ordne jedem Stoff ein Gefahrensymbol zu:	  



a)

Natron-  
lauge

Deodorant

Lösungen speichern
Arbeitsblatt drucken

Abbildung 132: HyperDoc der 1. Stunde der Unterrichtsreihe für die Oberstufe (2. Pilotierung). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008).

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN HyperDocSystems
Arbeitsblätter Hilfen Benutzerverwaltung
Monitoring
FAQ Eingeloggt als fitting

### Farbanalyse mit dem Handy Oberstufe

Konzentration [mg / l]	Rot-Werte	Grün-Wert	Blau-Werte
20 mg/l			
30 mg/l			
40 mg/l			
50 mg/l			
60 mg/l			
verd. Sirup			
unb. Konz.			

**Aufgaben:**

1. Trage die mit dem Handy gemessenen Werte in die Tabelle ein.
2. Erstelle die Eichgeraden für die drei RGB-Werte.  
Wähle dazu im Diagramm oben den "Pinset". Damit kannst du im Diagramm schreiben/malen.
3. Bestimme den Azorubin-Gehalt möglichst genau. Welche Wertreihe bietet sich an?

Schätzung anhand der ...	R-Werte	G-Werte	B-Werte
Konzentration:			

Begründung:

Lösungen speichern
 Arbeitsblatt drucken

Abbildung 133: HyperDoc der 2. Stunde der Unterrichtsreihe für die Oberstufe (2. Pilotierung).  
Verändert nach Nieß et al. (o. D.).

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN HyperDocSystems Arbeitsblätter Hilen Benutzerverwaltung Monitoring FAQ Eingeloggt als fitting

## Das Spektralphotometer

**ABBILDUNG 1**

**Information:**

Im Labor der Universität wurde das Spektralphotometer zunächst dazu benutzt, das Absorptionsverhalten der beiden Azofarbstoffe zu untersuchen. Dazu wurden die Lösungen beider Stoffe jeweils im Photometer bzgl. ihrer Extinktion bei allen Wellenlängen im sichtbaren Bereich untersucht. Die Darstellung der Extinktionswerte über den betreffenden Wellenlängenbereich nennt man Absorptionsspektrum (s. **Abbildung 2**). Aus ihm wurde abgeleitet, dass die Konzentrationsbestimmung bei Tartrazin bei einer Wellenlänge von 415 nm gemessen werden muss. Durch die Verwendung spezifischer Wellenlängen, die dem Absorptionsmaximum des Stoffes entsprechen, können auch verschiedene Stoffe in einer Mischung nebeneinander analysiert werden. Die Konzentrationsbestimmung mit dem Spektralphotometer erfolgt analog der Konzentrationsbestimmung, die du mit dem Smartphotometer schon durchgeführt hast. Es werden auch hier zunächst die Werte für Lösungen bekannter Konzentrationen bestimmt („Echtlösungen“). Aus diesen ergibt sich dann eine Eichgerade, mit deren Hilfe die Konzentration der unbekannteren Probe ermittelt werden kann. Die Messung ergab im stinkigen Sirup für Tartrazin eine Konzentration von 19 mg/l.

**ABBILDUNG 2**

**Aufgaben:**

- Lies dir die beiden **Buchseiten** durch und beschrifte die **Abbildung 1**.
- Lies den Text im Info-Kasten und vergleiche mit der **Abbildung 2**.
- Erkläre den Begriff „Extinktion“ und erläutere am Beispiel der Untersuchung von Tartrazin, wie man mit dem Photometer ein Absorptionsspektrum (s. **Abbildung 2**) aufnimmt und wie man eine Konzentrationsbestimmung durchführt.
- Vergleiche das Spektralphotometer mit dem „Smartphotometer“. Beschreibe die Unterschiede.
- Nenne Möglichkeiten, wie die Bestimmung der Azorubin-Konzentration mit dem „Smartphotometer“ noch weiter optimiert werden könnte.

Lösungen speichern Arbeitsblatt drucken

Abbildung 134: HyperDoc der 3. Stunde der Unterrichtsreihe für die Oberstufe (2. Pilotierung).  
Verändert nach Nieß et al. (o. D.).

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN HyperDocSystems    Arbeitsblätter    Hilfen    Benutzerverwaltung    Monitoring    FAQ Eingeloggt als fitting

## Optimierte Farbanalyse mit dem Handy von fitting

**Konzentration [mg/l]**

Grün-Werte

20	<input type="text"/>
30	<input type="text"/>
40	<input type="text"/>
50	<input type="text"/>
60	<input type="text"/>
verd. Sirup (Probe)	<input type="text"/>

**Versuchsaufbau**

Smartphone mit Color Grab App

Wellplatte mit Proben

Smartphone mit Flashlight-App

**Aufgaben:**

- Führe die Azorubin-Bestimmung nochmals durch (s. Versuchsaufbau). Verwende als Referenzwert die bereitgestellte Tartrazin-Lösung ( $c = 19 \text{ mg/l}$ ). Zur Beleuchtung wird grünes Licht verwendet. Erzeuge dazu in Flashscreen eine Beleuchtung, bei der die R- bzw. B-Werte auf Null gesetzt werden und der Grün-Wert maximal eingestellt wird.
- Berechne das Volumen des fertigen Getränks und die Anzahl an Gläsern (250 ml), das entsprechend des ADI-Wertes (4,5 mg/kg Körpergewicht und Tag) für Azorubin täglich von dir getrunken werden dürfte.
- Vertiefungsaufgabe

Lösungen speichern
Arbeitsblatt drucken

Abbildung 135: HyperDoc der 4. Stunde der Unterrichtsreihe für die Oberstufe (2. Pilotierung).  
Verändert nach Nieß et al. (o. D.).

## 9.2.2 Arbeitsblätter der 3. Pilotierung<sup>15</sup>

### Mittelstufe

HyperDocSystems Arbeitsblätter Hilfe Benutzerverwaltung Monitoring
F&G Eingelogg als fitting



### Arbeitsblatt: "Safety First!"

**Einführung:**  
 "Safety First" (engl. Schemet zuerst) - so heißt es oftmals im Alltag, wenn Gefahren drohen. Aber auch in der Schule. Gerade beim Experimentieren in Naturwissenschaften können gefährliche Situationen entstehen. Sei es durch das Experimentieren mit Biotopzoo in Physikunterricht oder durch Chemikalien im Chemieunterricht.

**Methode:**  
 Im Folgenden erarbeitest du eine Einführung in die digitalen Arbeitsblätter. Gehe dazu wie folgt vor:  
 Digitale Arbeitsblätter sind zunächst fast wie "normale" Arbeitsblätter. Das bedeutet, du erhältst verschiedene Informationen und musst dann Aufgaben bearbeiten. Das Besondere an den digitalen Arbeitsblättern: An verschiedenen Stellen sind Hilfen, Verlinkungsaufgaben oder interaktive Informationen hinterlegt. Diese zusätzlichen Inhalte öffnen du über Verlinkungen (blau unterstrichen) oder Symbole. Je nach Symbol unterscheidet sich die Darstellung des Inhalts. Die Hilfen, Verlinkungsaufgaben oder zusätzlichen Informationen erscheinen am rechten Bildschirmrand. Probiere es aus:  
 Hilfe als Video, Hilfe als Text, Hilfe als Symbol, Hilfe als Bild. Dabei stehen nicht immer alle Darstellungsoptionen für jede Hilfe zur Verfügung. Es gilt gegenseitig aber auch Verlinkungsaufgaben, die durch das Symbol gekennzeichnet sind.

**Hinweis:** Du solltest das Tablet beim Öffnen eines Arbeitsblattes bereits in der gewünschten Position verwenden, da das Drehen des Tablets nach dem Öffnen eines Arbeitsblattes zu Problemen beim Schreiben auf den Textfeldern führen kann.

**Gefahrenstoffe im Alltag:**  
 In unserem Alltag begegnest du einer Vielzahl von verschiedenen Gefahrenstoffen, auch wenn es auf den ersten Blick nicht so scheint. Hier sieht du zum Beispiel die Rückseite eines Geschirrspülmittels. Dort steht geschrieben: "Verursacht schwere Augenreizungen. Darf nicht in die Hände von Kindern gelangen." Gehe nachfolgend mit einem Aufzählensymbol zu sehen.

Das war nur ein Beispiel. Du findest viele weitere Gefahrenstoffhinweise auf allen möglichen alltäglichen Verbrauchsgüter. Insgesamt gibt es neun Kategorien für Gefahrenstoffe, jede Kategorie wird dabei durch ein spezielles Symbol gekennzeichnet.

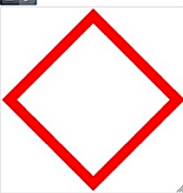
Symbol	Beschreibung	Beispiele für gefährliche Eigenschaften
	Explosionsfähige Bombe	instabile explosive Stoffe
	Flammable	entzündbare Stoffe
	Flammable über 100°C	entzündende Stoffe
	Gefäß unter Druck	Gas unter Druck
	Ätzend	schwere Augenschädigungen, Hautschädigung
	Toxisch	akute Toxizität
	Reizend	Reizend, Augenreizend
	Gesundheitsschädlich	diverse Gesundheitsschädigungen
	Umwelt	gewässerspezifisch

**Hinweis:** Die folgenden Textfelder kannst du entweder mit der virtuellen Tastatur aufrufen (Standardanstellung), oder mit dem Apple Pencil. Klicke dazu auf das Symbol. Mit der Maus kannst du Textfelder aufrufen. Du kannst auch zwischen dem Radier- und Pinsel-Modus wechseln, die Farbe einstellen und die Pinselgröße ändern. Achtung: Über den Bleistift weist du die gesamte Textfelder. Manchmal ist das Tastatur- und Pinsel-Symbol auch über einen Bild zu sehen. Du kannst dann das Bild beschriften (s. Aufgabe 1).


**Wichtig:** Du musst deine Lösungen speichern, bevor du das Tablet schließt oder die Seite neu lädst. Drücke dazu den Button "Lösung speichern" am Ende der Seite.

**Aufgaben:**

1. Vervollständige das Gefahrensymbol "Dickscheibenaufzählensymbol".



2. Folgende Stoffe sind gegeben:



**Natronlauge**

**Deodorant**

Drehe jedem Stoff ein Gefahrensymbol zu. Schreibe dazu die Bezeichnung des Gefahrensymbols auf.

3. Verlinkungsaufgabe:

**Beachte:**  
 \* Eingetragene Marken sind urheberrechtlich geschützt. Alle Rechte vorbehalten. © 2018 HyperDocSystems. Alle Rechte vorbehalten.

Lösungen speichern
Arbeitsblatt drucken

Abbildung 136: HyperDoc der 1. Stunde der Unterrichtsreihe für die Mittelstufe (3. Pilotierung). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008).

<sup>15</sup> Textfelder verkleinert

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN HyperDocSystems Arbeitsblätter Hilfen Benutzerverwaltung Monitoring
FAQ Eingeloggt als fitting

## Farbstoffe in Lebensmitteln

**Erinnerung!** Du solltest das Tablet beim Öffnen eines Arbeitsblattes bereits in der gewünschten Position verwenden, da das Drehen des Tablets nach dem Öffnen eines Arbeitsblattes zu Problemen beim Schreiben auf den Textfeldern führen kann. Drehe das Tablet ggf. in eine neue Position und öffne das Arbeitsblatt erneut.

Auszug aus Wikipedia zu „Azofarbstoffen“ und ADI-Werten (verändert)<sup>1</sup>

**Lebensmittelzusatzstoffe**

Die Europäische Union (EU) vergibt eine E-Nummer für jeden zugelassenen Stoff. Insgesamt gibt es in der EU zurzeit 341 zugelassene Zusatzstoffe. Eine Zulassung wird nur erteilt, wenn keine Gesundheitsrisiken bestehen, wenn der Zusatzstoff technisch notwendig ist und wenn die Verwendung nicht zu einer Täuschung des Verbrauchers führt. Zusatzstoffe dürfen nur nach ausdrücklicher Zulassung verwendet werden. Zudem müssen Zusatzstoffe auf dem Produkt kenntlich gemacht werden.

**Rechtliche Situation**

Die EU-Kommission beauftragte am 25. März 2010 die EFSA (European Food Safety Authority), eine neuere gesundheitliche Bewertung fraglicher Azofarbstoffe vorzunehmen. Seit dem 20. Juli 2010 müssen Lebensmittel, die die Azofarbstoffe Tartrazin (E 102), Gelborange S (E 110), Azorubin (E 122), Allurarot (E 129) oder Cochenillrot A (E 124) enthalten, in der Europäischen Union mit dem gesonderten Warnhinweis „Kann Aktivität und Aufmerksamkeit bei Kindern beeinträchtigen“ gekennzeichnet werden.

**Azorubin**

Azorubin (auch Karmesin, E 122) ist ein als Lebensmittelfarbe zugelassener roter Azofarbstoff und das Dinatriumsalz einer Disulfonsäure. Azorubin wird für Getränke, Süß- und Zuckerkwaren, Marzipan, Puddingpulver, Götterspeise, Fruchtkonserven, Fertigsuppen, Fertigprodukte, braune Sauce, Paniermehl, Färbemittel für Tabletten und anderes verwendet. Azorubin hat große chemische Ähnlichkeit mit Amaranth (E 123) und steht wie dieser Stoff im Verdacht als Auslöser von Pseudoallergien zu wirken.

**Tartrazin**

Tartrazin (E 102) ist ein synthetischer Azofarbstoff, der als zitronengelber bis orangefarbener Lebensmittelfarbstoff eingesetzt wird. Der Stoff ist farbecht.

**Erlaubte Tagesdosis**

Die erlaubte Tagesdosis (ETD; englisch: acceptable daily intake, ADI) bezeichnet die Dosis einer Substanz, wie etwa eines Lebensmittelzusatzstoffs, Pestizids oder eines Medikaments, die bei lebenslanger täglicher Einnahme als medizinisch unbedenklich betrachtet wird. Handelt es sich um ungewollte Verunreinigungen, spricht man von einer toxischen Tagesdosis (englisch: toxicologic daily intake, TDI). Angegeben wird der ETD-Wert (ADI-Wert) in Milligramm bzw. Mikrogramm pro Kilogramm Körpergewicht und Tag. Synergieeffekte (Zusammenwirken verschiedener Lebensmittelzusatzstoffe) werden jedoch nur teilweise berücksichtigt.

**ADI-Wert von Azorubin:** 4 mg / kg Körpergewicht  
**ADI-Wert von Tartrazin:** 7,5 mg / kg Körpergewicht

**Aufgaben**

- Begründe die Einheit des ADI-Wertes (mg/kg Körpergewicht).
- Berechne die Masse, die ein Kleinkind und ein Erwachsener täglich von beiden Farbstoffen aufnehmen dürfen.
- Du möchtest das Risiko bei der Verwendung des Granadine-Sirups abschätzen. Der Gehalt welches Farbstoffs würdest du zuerst untersuchen? Begründe!
- Vertiefungsaufgabe

<sup>1</sup> Quelle(n):  
<https://de.wikipedia.org/wiki/Azofarbstoff> (Seite der Autoren hier verfügbar)  
[https://de.wikipedia.org/wiki/Erlaubte\\_Tagesdosis](https://de.wikipedia.org/wiki/Erlaubte_Tagesdosis) (Seite der Autoren hier verfügbar)  
 Die Texte stehen unter der Lizenz Creative Commons Attribution-ShareAlike 3.0 Unported

Sirupflasche

Lösungen speichern
Arbeitsblatt drucken

Abbildung 137: HyperDoc der 2. Stunde der Unterrichtsreihe für die Mittelstufe (3. Pilotierung). Verändert nach Czubatinski et al. (2020). Kleinkind nach OmarMedinaFilms (2016).

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN HyperDocSystems Arbeitsblätter Hilfen Benutzerverwaltung Monitoring FAQ Eingelogg als fitting

## Farbanalyse mit dem Handy

**Versuch**  
Du kannst dir hier nochmal die Versuchsdurchführung anschauen, anhören oder durchlesen:

**Tabelle RGB-Werte**

Konzentration [mg/l]	Rot-Werte	Grün-Wert	Bau-Werte
20 mg/l			
30 mg/l			
40 mg/l			
50 mg/l			
60 mg/l			
verd. Sirup			
unb. Konz.			

**Diagramm RGB-Werte**

**Aufgaben:**

1. Trage die mit dem Handy gemessenen Werte in die Tabelle ein.
2. Übertrage die Werte der bekannten Konzentrationen in das Diagramm. Verwende entsprechende Farben.  
 *Tipp: Nutze gegebenenfalls dein Lineal, um eine gerade Linie zu ziehen.*
3. Versuche die Konzentration von Azorubin im Sirup abzuschätzen! Betrachte R-, G- und B-Werte.

Schätzung anhand der ... R-Werte G-Werte B-Werte

Konzentration

4. Vertiefungsaufgabe

[Lösungen speichern](#) [Arbeitsblatt drucken](#)

Abbildung 138: HyperDoc der 3. Stunde der Unterrichtsreihe für die Mittelstufe (3. Pilotierung). Verändert nach Czubatinski et al. (2020).

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN HyperDocSystems Arbeitsblätter Hilfen Benutzerverwaltung Monitoring FAQ Eingeloggt als fitting

### Konzentrationen (neu)

Tabelle Konzentrationsbestimmung

20 mg/l	30 mg/l	40 mg/l	50 mg/l	60 mg/l	mg/l
Eichlösungen					
Probe					

Abbildung Mischverhältnis

Mischungsverhältnis: 1:7

**Aufgaben**

- Stelle die Konzentration der Eichlösungen und der Probe analog dem vorgegebenen Beispiel (50 mg/l) in der Tabelle dar! **Runde** die Konzentration der Probe dazu auf 10 mg/l!
- Wie viele Gläser (0,2 Liter) kannst du laut des ADI-Werts von einem Getränk mit dem Sirup täglich trinken?
- Berechne die Konzentration des unverdünnten Sirups und stelle sie in der Abbildung wie oben dar! Begründe anhand der Abbildung.

Vorfahrungsaufgabe

Abbildung 139: HyperDoc der 4. Stunde der Unterrichtsreihe für die Mittelstufe (3. Pilotierung). Verändert nach Czubatinski et al. (2020).



**Oberstufe**

**HyperDocSystems Arbeitsblätter** Hilfen Benutzerverwaltung Monitoring FAQ Eingelogg als fitting



## Arbeitsblatt: "Safety First!"

**Einführung:**  
 "Safety First" (engl.) Sicherheit zuerst - so heißt es oftmals im Alltag, wenn Gefahren drohen. Aber auch in der Schule gilt der Spruch. Gerade beim Experimentieren in Naturwissenschaften können gefährliche Situationen entstehen. Sei es durch das Experimentieren mit Elektrostatik im Physikunterricht oder durch Chemikalien im Chemieunterricht.

**Methode:**  
 Im Folgenden erhaltet ihr eine Einführung in die digitalen Arbeitsblätter. Gleichzeit lernt ihr etwas über gefährliche Stoffe (Gefahrenstoffe) im Chemieunterricht.  
 Digitale Arbeitsblätter sind zunächst fast wie "normale" Arbeitsblätter. Das bedeutet, du erhältst verschiedene Informationen und musst dann Aufgaben bearbeiten. Das besondere an den digitalen Arbeitsblättern: An verschiedenen Stellen sind Hilfen, Vertiefungsaufgaben oder weiterführende Informationen hinterlegt. Diese zusätzlichen Inhalte öffnest du über **Verlinkungen** (blau unterstrichen) oder Symbole. Je nach Symbol unterscheidet sich die Darstellung des Inhaltes. Die Hilfen, Vertiefungsaufgaben oder zusätzlichen Informationen erscheinen am rechten Bildschirmrand. Probleme es aus:  
 - Hilfe als Video - Hilfe als Text - Hilfe als Tonspur - Hilfe als Bild . **Dabei stehen nicht immer alle Darstellungsformen für jede Hilfe zur Verfügung.** Es gibt gelegentlich aber auch Vertiefungsaufgaben, die durch das Symbol gekennzeichnet sind.

**Hinweise:** Du solltest das Tablet beim Öffnen eines Arbeitsblatts bereits in der gewünschten Position verwenden, da das Drehen des Tablets nach dem Öffnen eines Arbeitsblatts zu Problemen beim Schreiben auf den Touchpads führen kann.

**Gefahrenstoffe im Alltag:**  
 In unserem Alltag befinden sich eine Vielzahl von verschiedenen Gefahrenstoffen, auch wenn es auf den ersten Blick nicht so scheint. Hier siehst du zum Beispiel die Rückseite eines Geschirrspülmittels. Dort steht geschrieben "Verursacht schwere Augenreizungen. Darf nicht in die Hände von Kindern gelangen [...]". Gleichzeitig ist ein großes Symbol mit einem Ausrufezeichen zu sehen.

Das war nur ein Beispiel. Du findest viele weitere Gefahrenstoffhinweise auf allen möglichen alltäglichen Verbrauchsgegenständen. Insgesamt gibt es neun Kategorien für Gefahrenstoffe. Jede Kategorie wird dabei durch ein spezielles Symbol gekennzeichnet.

Symbol	Beschreibung	Beispiele für gefährliche Eigenschaften
	Explosionsgefährlich	instabile explosive Stoffe
	Flammable	entzündbare Stoffe
	Flammable liquid	entzündende Stoffe
	Gas	Gas unter Druck
	Corrosive	schwere Augenschädigungen, hautreizend
	Toxic	akute Toxizität
	Health hazard	hautreizend, augenreizend
	Health hazard	diverse Gesundheitsgefahren
	Environment	gewässergefährdend


**Hinweise:** Die folgenden Toucher können du entweder mit der virtuellen Tastatur auslösen (Standardbedienung) oder mit dem Apple Pencil. Klicke dazu auf das Symbol. Mit der Wischgeste kannst du zwischen dem Reader und Pinset wechseln, die Farbe einstellen und die Pinsetgröße ändern. Achtung: Über den Button wird die gesamte Tafelbildschirm in das Tablet- und Pinset-Symbol auch über einem Bild zu sehen. Du kannst dann das Bild beschriften (s. Aufgabe 1).

**Wichtig: Du musst deine Lösungen speichern, bevor du das Tablet schließt oder die Seite neu lädst. Drücke dazu den Button "Lösungen speichern" am Ende der Seite.**

**Aufgaben:**

1. Vervollständige das Gefahrensymbol "Dickes Ausrufezeichen":

2. Folgende Stoffe sind gegeben:



a)

b)

**Natronlauge**

c)

**Deodorant**

Ordne jedem Stoff ein Gefahrensymbol zu. Schreibe dazu die Bezeichnung des Gefahrensymbols auf.

3. Vertiefungsaufgabe

Quelle: [https://www.knipschlag.de/Dateien/Arbeitsblätter/Arbeitsblatt\\_Sicherheit\\_im\\_Chemieunterricht.pdf](https://www.knipschlag.de/Dateien/Arbeitsblätter/Arbeitsblatt_Sicherheit_im_Chemieunterricht.pdf)

Lösungen speichern
Arbeitsblatt drucken

Abbildung 140: HyperDoc der 1. Stunde der Unterrichtsreihe für die Oberstufe (3. Pilotierung). Gasflaschen nach Bru-nO (2017). Gefahrensymbole nach Wikipedia Foundation Inc. (2008).

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN HyperDocSystems
Arbeitsblätter Hilfen Benutzerverwaltung Monitoring
FAQ Eingeloggt als fitting

## Farbanalyse mit dem Handy Oberstufe (neu)

**Erinnerung:** Du solltest das Tablet beim Öffnen eines Arbeitsblattes bereits in der gewünschten Position verwenden, da das Drehen des Tablets nach dem Öffnen eines Arbeitsblattes zu Problemen beim Schreiben auf den Textfeldern führen kann. Drehe das Tablet ggf. in eine neue Position und öffne das Arbeitsblatt erneut.

**Versuch**

Du kannst du hier nochmal die Versuchsdurchführung anschauen, anhören oder durchlesen:

**Tabelle RGB-Werte**

Konzentration [mg / l]	Rot-Werte	Grün-Wert	Blau-Werte
20 mg/l			
30 mg/l			
40 mg/l			
50 mg/l			
60 mg/l			
verd. Sirup unb. Konz.			

**Diagramm RGB-Werte**

**Aufgaben**

1. Trage die mit dem Handy gemessenen Werte in die Tabelle ein.
2. Erstelle die Eichgeraden für die drei RGB-Werte.   
Tipp: Verwende gegebenenfalls ein Lineal, um eine gerade Linie zu ziehen.
3. Bestimme den Azorubin-Gehalt möglichst genau - begründe, welche Wertreihe sich hierfür am besten eignet.

**Schätzung anhand der**

R-Werte	G-Werte	B-Werte

**Konzentration**

--	--	--

**Begründung:**

4. Vertiefungsaufgabe

Lösungen speichern
Arbeitsblatt drucken

Abbildung 141: HyperDoc der 2. Stunde der Unterrichtsreihe für die Oberstufe (3. Pilotierung).  
Verändert nach Nieß et al. (o. D.).

Seite | 488

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN HyperDocSystems Arbeitsblätter Hilfen Benutzerverwaltung Monitoring FAQ Eingeloggt als fitting

## Das Spektralphotometer

Aufbau eines Photometers



**ABBILDUNG 1**


**Information**





Im Labor der Universität wurde das Spektralphotometer zunächst dazu benutzt, das Absorptionsverhalten der beiden Azofarbstoffe zu untersuchen. Dazu wurden die Lösungen beider Stoffe jeweils im Photometer bzgl. ihrer Extinktion bei allen Wellenlängen im sichtbaren Bereich untersucht. Die Darstellung der Extinktionswerte über den betreffenden Wellenlängenbereich nennt man Absorptionsspektrum (s. **Abbildung 2**). Aus ihm wurde abgeleitet, dass die Konzentrationsbestimmung bei Tartrazin bei einer Wellenlänge von 425 nm gemessen werden muss. Durch die Verwendung spezifischer Wellenlängen, die dem Absorptionsmaximum des Stoffes entsprechen, können auch verschiedene Stoffe in einer Mischung nebeneinander analysiert werden. Die Konzentrationsbestimmung mit dem Spektralphotometer erfolgt analog der Konzentrationsbestimmung, die du mit dem Smartphotometer schon durchgeführt hast. Es werden auch hier zunächst die Werte für Lösungen bekannter Konzentrationen bestimmt („Eichlösungen“). Aus diesen ergibt sich dann eine Eichgerade, mit deren Hilfe die Konzentration der unbekannteren Probe ermittelt werden kann. Die Messung ergab im trinkfitigen Snup für Tartrazin eine Konzentration von 19 mg/l.


**ABBILDUNG 2**



**Aufgaben**


- Lies dir die beiden Informationssätze durch und beschrifte die **Abbildung 1**.  
- Lies den Text im Info-Kasten und vergleiche mit der **Abbildung 2**.
- Erkläre den Begriff „Extinktion“ und erläutere am Beispiel der Untersuchung von Tartrazin, wie man mit dem Photometer ein Absorptionsspektrum (s. **Abbildung 2**) aufnimmt und wie man eine Konzentrationsbestimmung durchführt.





- Vergleiche das Spektralphotometer mit dem „Smartphotometer“. Beschreibe die Unterschiede.    



- Nenne Möglichkeiten, wie die Bestimmung der Azobubin-Konzentration mit dem „Smartphotometer“ noch weiter optimiert werden könnte.  



- Vertiefungsaufgabe  


Lösungen speichern  Arbeitsblatt drucken

Abbildung 142: HyperDoc der 3. Stunde der Unterrichtsreihe für die Oberstufe (3. Pilotierung).  
Verändert nach Nieß et al. (o. D.).

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN HyperDocSystems Arbeitsblätter Hilfen Benutzerverwaltung Monitoring FAQ Eingeloggt als fitting

## Optimierte Farbanalyse mit dem Handy

Tabelle Grün-Werte

Konzentration [mg/l]	Grün-Werte
20	<input type="text"/>
30	<input type="text"/>
40	<input type="text"/>
50	<input type="text"/>
60	<input type="text"/>
verd. Sirup (Probe)	<input type="text"/>

**Versuchsaufbau**

Diagramm Grün-Werte

Konzentration Probe:

**Aufgaben**

- Führe die Azorubin-Bestimmung nochmals durch (s. Versuchsaufbau). Verwende als Referenzwert die bereitgestellte Tartrazin-Lösung ( $c = 19 \text{ mg/l}$ ). Zur Beleuchtung wird grünes Licht verwendet. Erzeuge dazu in Flashscreen eine Beleuchtung, bei der die R- bzw. B-Werte auf Null gesetzt werden und der Grün-Wert maximal eingestellt wird.
- Berechne das Volumen des fertigen Getränks und die Anzahl an Gläsern (250 ml), das entsprechend des ADI-Wertes (40 mg/kg Körpergewicht und Tag) für Azorubin täglich von dir getrunken werden dürfte.
- Vertiefungsaufgabe

Abbildung 143: HyperDoc der 4. Stunde der Unterrichtsreihe für die Oberstufe (3. Pilotierung).  
Verändert nach Nieß et al. (o. D.).

## 9.2.3 Hauptkomponentenanalyse Messinstrument Cognitive Load

Tabelle 178: Korrelationsmatrix der Hauptkomponentenanalyse.

Korrelationsmatrix

		CL1	CL2	CL3	CL4	CL5	CL6	CL7	CL8	CL9	CL11	CL10INV
Korrelation	CL1	1,000	,492	,437	,242	,009	,644	,247	,693	,139	,025	-,305
	CL2	,492	1,000	,615	,855	,163	,268	,750	,292	,076	,067	-,061
	CL3	,437	,615	1,000	,435	,560	,098	,410	,164	-,068	-,111	-,074
	CL4	,242	,855	,435	1,000	,177	,328	,867	,193	,118	,156	-,115
	CL5	,009	,163	,560	,177	1,000	,022	,180	-,058	-,054	-,216	,012
	CL6	,644	,268	,098	,328	,022	1,000	,418	,776	,201	,028	-,612
	CL7	,247	,750	,410	,867	,180	,418	1,000	,305	,038	,152	-,229
	CL8	,693	,292	,164	,193	-,058	,776	,305	1,000	-,164	,122	-,649
	CL9	,139	,076	-,068	,118	-,054	,201	,038	-,164	1,000	,230	,008
	CL11	,025	,067	-,111	,156	-,216	,028	,152	,122	,230	1,000	-,010
	CL10INV	-,305	-,061	-,074	-,115	,012	-,612	-,229	-,649	,008	-,010	1,000

Tabelle 179: Anti-Image-Matrizen der Hauptkomponentenanalyse.

Anti-Image-Matrizen

		CL1	CL2	CL3	CL4	CL5	CL6	CL7	CL8	CL9	CL11	CL10INV
Anti-Image-Kovarianz	CL1	,182	-,056	-,097	,041	,078	-,070	,048	-,047	-,058	-,005	-,085
	CL2	-,056	,094	-,042	-,072	,036	,058	-,025	-,036	-,050	,076	-,034
	CL3	-,097	-,042	,281	,015	-,247	,041	-,028	,034	,059	-,011	,096
	CL4	,041	-,072	,015	,101	-,017	-,049	-,056	,034	,027	-,068	,014
	CL5	,078	,036	-,247	-,017	,521	-,042	,005	-,012	-,029	,089	-,076
	CL6	-,070	,058	,041	-,049	-,042	,139	-,048	-,074	-,123	,119	,028
	CL7	,048	-,025	-,028	-,056	,005	-,048	,199	,002	,045	-,052	,009
	CL8	-,047	-,036	,034	,034	-,012	-,074	,002	,129	,164	-,130	,113
	CL9	-,058	-,050	,059	,027	-,029	-,123	,045	,164	,474	-,250	,103
	CL11	-,005	,076	-,011	-,068	,089	,119	-,052	-,130	-,250	,658	-,091
	CL10INV	-,085	-,034	,096	,014	-,076	,028	,009	,113	,103	-,091	,426
	Anti-Image-Korrelation	CL1	,607 <sup>a</sup>	-,430	-,431	,306	,253	-,439	,253	-,308	-,198	-,013
CL2		-,430	,601 <sup>a</sup>	-,260	-,740	,162	,508	-,180	-,329	-,237	,305	-,172
CL3		-,431	-,260	,599 <sup>a</sup>	,091	-,645	,207	-,119	,180	,162	-,026	,278
CL4		,306	-,740	,091	,629 <sup>a</sup>	-,073	-,410	-,392	,296	,124	-,265	,068
CL5		,253	,162	-,645	-,073	,437 <sup>a</sup>	-,155	,016	-,046	-,058	,153	-,160
CL6		-,439	,508	,207	-,410	-,155	,550 <sup>a</sup>	-,288	-,554	-,477	,395	,113
CL7		,253	-,180	-,119	-,392	,016	-,288	,831 <sup>a</sup>	,015	,145	-,142	,032
CL8		-,308	-,329	,180	,296	-,046	-,554	,015	,544 <sup>a</sup>	,664	-,446	,481
CL9		-,198	-,237	,162	,124	-,058	-,477	,145	,664	,135 <sup>a</sup>	-,447	,229
CL11		-,013	,305	-,026	-,265	,153	,395	-,142	-,446	-,447	,186 <sup>a</sup>	-,172
CL10INV		-,306	-,172	,278	,068	-,160	,113	,032	,481	,229	-,172	,633 <sup>a</sup>

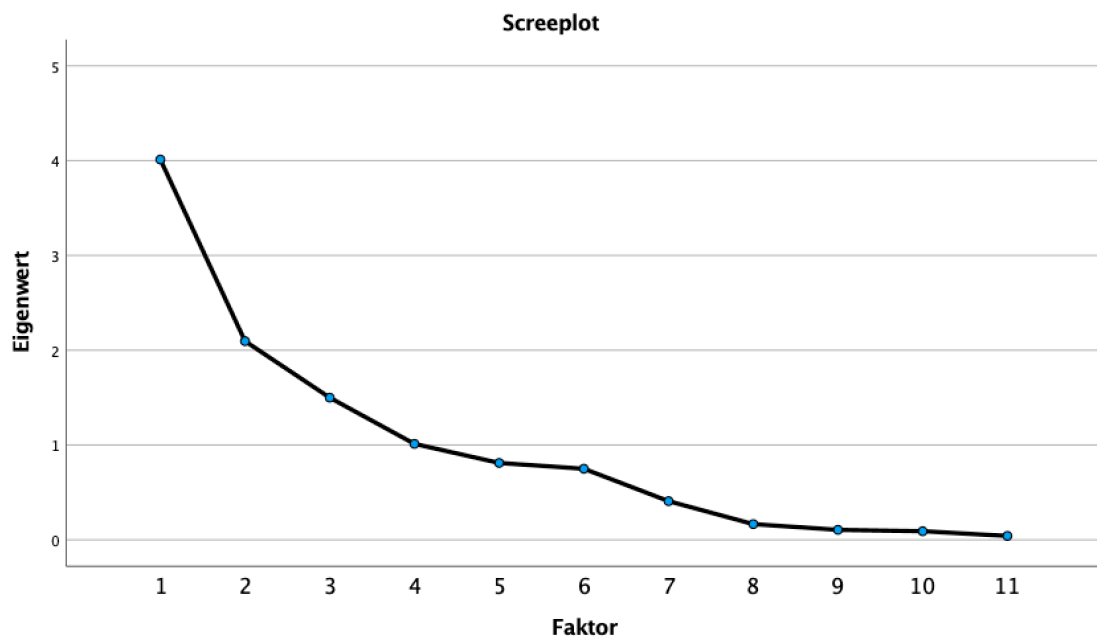
a. Maß der Stichprobeneignung

Tabelle 180: Erklärte Gesamtvarianz der Hauptkomponentenanalyse.

Erklärte Gesamtvarianz

Komponente	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	4,012	36,475	36,475	4,012	36,475	36,475	2,955	26,864	26,864
2	2,097	19,060	55,535	2,097	19,060	55,535	2,869	26,082	52,946
3	1,500	13,640	69,175	1,500	13,640	69,175	1,631	14,825	67,771
4	1,013	9,207	78,383	1,013	9,207	78,383	1,167	10,611	78,383
5	,812	7,381	85,763						
6	,751	6,823	92,586						
7	,409	3,716	96,303						
8	,167	1,514	97,817						
9	,106	,967	98,784						
10	,092	,833	99,617						
11	,042	,383	100,000						

Extraktionsmethode: Hauptkomponentenanalyse.



*Abbildung 144: Screepplot der Hauptkomponentenanalyse.*

## 9.2.4 Lösungswahrscheinlichkeit der Aufgaben in der 3. Pilotierung

### Mittelstufe

*Tabelle 181: Lösungswahrscheinlichkeiten der Aufgaben in den jeweiligen Stunden der 3. Pilotierung in der Mittelstufe.*

	Stunde 1 (n = 21)			
	Aufgabe 1	Aufgabe 2	Aufgabe 3	Vertiefung
Nicht beantwortet	9.52 %	9.52 %		19.05 %
Falsch beantwortet	0.00 %	9.52 %		4.76 %
Unvollständig beantwortet	0.00 %	9.52 %	nicht vorhanden	4.76 %
Richtig beantwortet	90.48%	71.43 %		71.43 %
	Stunde 2 (n = 19)			
	Aufgabe 1	Aufgabe 2	Aufgabe 3	Vertiefung
Nicht beantwortet	5.26 %	0.00 %	5.26 %	36.84 %
Falsch beantwortet	36.82 %	15.79 %	31.58 %	10.53 %
Unvollständig beantwortet	5.26 %	0.00 %	57.89 %	31.58 %
Richtig beantwortet	52.63 %	84.21 %	5.26 %	21.05 %
	Stunde 3 (n = 23)			
	Aufgabe 1	Aufgabe 2	Aufgabe 3	Vertiefung
Nicht beantwortet	8.70 %	4.35 %	82.61 %	95.65 %
Falsch beantwortet	0.00 %	21.74 %	17.39 %	0.00 %
Unvollständig beantwortet	0.00 %	17.39 %	0.00 %	0.00 %
Richtig beantwortet	91.30 %	56.52 %	0.00 %	4.35 %
	Stunde 4 (n = 22)			
	Aufgabe 1	Aufgabe 2	Aufgabe 3	Vertiefung
Nicht beantwortet	9.09 %	50 %	59.09 %	100 %
Falsch beantwortet	0.00 %	9.09 %	36.36 %	0.00 %
Unvollständig beantwortet	22.73 %	13.64 %	4.55 %	0.00 %
Richtig beantwortet	68.19 %	4.55 %	0.00 %	0.00 %

**Oberstufe***Tabelle 182: Lösungswahrscheinlichkeiten der Aufgaben in den jeweiligen Stunden der 3. Pilotierung in der Oberstufe.*

Stunde 1 (n = 6)					
	Aufgabe 1	Aufgabe 2	Aufgabe 3	Vertiefung	
Nicht beantwortet	0.00 %	0.00 %		0.00 %	
Falsch beantwortet	0.00 %	0.00 %		0.00 %	
Unvollständig beantwortet	0.00 %	0.00 %	nicht vorhanden		0.00 %
Richtig beantwortet	100.00 %	100.00 %		100.00 %	
Stunde 2 (n = 6)					
	Aufgabe 1	Aufgabe 2	Aufgabe 3	Vertiefung	
Nicht beantwortet	16.67 %	16.67 %	0.00 %	33.33 %	
Falsch beantwortet	0.00 %	0.00 %	0.00 %	16.67 %	
Unvollständig beantwortet	0.00 %	16.67 %	0.00 %	0.00 %	
Richtig beantwortet	83.33 %	66.67 %	100.00 %	50.00 %	
Stunde 3 (n = 6)					
	Aufgabe 1	Aufgabe 3*	Aufgabe 4	Aufgabe 5	Vertiefung
Nicht beantwortet	0.00 %	0.00 %	0.00 %	0.00 %	50.00 %
Falsch beantwortet	0.00 %	0.00 %	0.00 %	16.67 %	0.00 %
Unvollständig beantwortet	0.00 %	33.33 %	33.33 %	66.67 %	16.67 %
Richtig beantwortet	100.00 %	66.67 %	66.67 %	16.67 %	33.33 %
Stunde 4 (n = 6)					
	Aufgabe 1	Aufgabe 2	Aufgabe 3	Vertiefung	
Nicht beantwortet	0.00 %	0.00 %	0.00 %	83.33 %	
Falsch beantwortet	16.67 %	0.00 %	0.00 %	0.00 %	
Unvollständig beantwortet	83.33 %	16.67 %	16.67 %	0.00 %	
Richtig beantwortet	0.00 %	83.33 %	83.33 %	16.67 %	

\*In Aufgabe 2 sollte lediglich sein Text gelesen werden, weshalb sie hier nicht aufgeführt ist.



## 9.3 Ergebnisse

### 9.3.1 Beschreibung der Stichprobe

*Tabelle 183: Nutzungsangebot digitaler Endgeräte im häuslichen Umfeld. n = 198 (Gymnasium), n = 108 (Gesamtschule).*

	Nutzung Smartphone		Nutzung Tablet		Nutzung Computer	
	w	m	w	m	w	m
Gymnasium	96	94	58	56	83	94
Gesamtschule	48	60	21	23	36	49

*Tabelle 184: Nutzungsangebot digitaler Endgeräte im schulischen Umfeld. Mittelstufe: n = 198 (Gymnasium), n = 108 (Gesamtschule).*

	Nutzung Smartphone		Nutzung Tablet		Nutzung Computer	
	w	m	w	m	w	m
Gymnasium	57	60	13	22	64	50
Gesamtschule	37	44	28	40	23	32

*Tabelle 185: Nutzungsangebot digitaler Endgeräte im häuslichen Umfeld. Oberstufe: n = 101 (Gymnasium), n = 8 (Gesamtschule).*

	Nutzung Smartphone		Nutzung Tablet		Nutzung Computer	
	w	m	w	m	w	m
Gymnasium	43	49	25	27	41	49
Gesamtschule	8	/	8	/	8	/

*Tabelle 186: Nutzungsangebot digitaler Endgeräte im schulischen Umfeld. Oberstufe: n = 101 (Gymnasium), n = 8 (Gesamtschule).*

	Nutzung Smartphone		Nutzung Tablet		Nutzung Computer	
	w	m	w	m	w	m
Gymnasium	21	34	5	8	22	32
Gesamtschule	22	/	8	/	1	/

### 9.3.2 Intrinsische Motivation und Interesse (Mittelstufe)

#### Interesse (SDT)

Tabelle 187: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Gruppe“ (hier: Interventionsgruppe).

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	0.14	0.025	< 0.001*
1 – 3	0.37	0.010	< 0.001*
1 – 4	0.41	0.009	< 0.001*
2 – 3	0.22	0.017	< 0.001*
2 – 4	0.24	0.013	< 0.001*
3 – 4	0.01	0.050	0.784

Tabelle 188: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Gruppe“ (hier: Vergleichsgruppe).

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i> <i>tiert)</i>	<i>p-Wert</i>
1 – 2	-0.13	0.010	0.052
1 – 3	-0.07	0.025	0.345
1 – 4	0.07	0.050	0.373
2 – 3	0.12	0.013	0.062
2 – 4	0.25	0.009	< 0.001*
3 – 4	0.09	0.017	0.068

**Wahrgenommene Wahlfreiheit**

*Tabelle 189: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Geschlecht“ (hier: weiblich).*

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	0.16	0.050	0.002*
1 – 3	0.42	0.013	< 0.001*
1 – 4	0.60	0.009	< 0.001*
2 – 3	0.26	0.017	< 0.001*
2 – 4	0.42	0.010	< 0.001*
3 – 4	0.17	0.025	0.001*

*Tabelle 190: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Geschlecht“ (hier: männlich).*

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	0.06	0.050	0.125
1 – 3	0.25	0.010	< 0.001*
1 – 4	0.39	0.009	< 0.001*
2 – 3	0.16	0.017	< 0.001*
2 – 4	0.27	0.013	< 0.001*
3 – 4	0.12	0.025	0.037

*Tabelle 191: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Gruppe“ (hier: Interventionsgruppe).*

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	0.05	0.025	0.215
1 – 3	0.35	0.009	< 0.001*
1 – 4	0.43	0.010	< 0.001*
2 – 3	0.31	0.017	< 0.001*
2 – 4	0.35	0.013	< 0.001*
3 – 4	0.05	0.050	0.250

Tabelle 192: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Gruppe“ (hier: Vergleichsgruppe).

Stundenvergleich	$\hat{\Psi}$	p-Wert-Schwelle (adjus-	p-Wert
1 – 2	0.20	0.025	0.003*
1 – 3	0.28	0.017	< 0.001*
1 – 4	0.61	0.009	< 0.001*
2 – 3	0.05	0.050	0.411
2 – 4	0.34	0.010	< 0.001*
3 – 4	0.29	0.013	< 0.001*

### 9.3.3 Intrinsische Motivation und Interesse (Oberstufe)

#### Interesse (SDT)

Tabelle 193: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Geschlecht“ (hier: weiblich).

Stundenvergleich	$\hat{\Psi}$	p-Wert-Schwelle (adjus- tiert)	p-Wert
1 – 2	0.33	0.010	0.001*
1 – 3	0.68	0.009	< 0.001*
1 – 4	0.38	0.013	0.003*
2 – 3	0.27	0.017	0.009*
2 – 4	-0.02	0.050	0.834
3 – 4	-0.24	0.025	0.018*

*Tabelle 194: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Geschlecht“ (hier: männlich).*

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	0.09	0.025	0.128
1 – 3	0.31	0.009	0.001*
1 – 4	0.19	0.013	0.043
2 – 3	0.16	0.010	0.032
2 – 4	0.06	0.050	0.512
	-0.16	0.017	0.056

*Tabelle 195: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Gruppe“ (hier: Interventionsgruppe).*

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	0.20	0.017	0.002*
1 – 3	0.51	0.009	< 0.001*
1 – 4	0.44	0.010	< 0.001*
2 – 3	0.24	0.013	0.002*
2 – 4	0.13	0.025	0.034
3 – 4	-0.09	0.050	0.146

*Tabelle 196: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Gruppe“ (hier: Vergleichsgruppe).*

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	0.11	0.050	0.446
1 – 3	0.30	0.013	0.054
1 – 4	-0.21	0.017	0.175
2 – 3	0.10	0.025	0.426
2 – 4	-0.33	0.010	0.015
3 – 4	-0.47	0.009	< 0.001*

**Wahrgenommene Wahlfreiheit***Tabelle 197: Post-hoc-Analyse des Haupteffekts „Messzeitpunkt“ (Interventionsgruppe – Vergleichsgruppe).*

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	0.24	0.010	< 0.001*
1 – 3	0.40	0.009	< 0.001*
1 – 4	0.25	0.013	0.002*
2 – 3	0.10	0.025	0.024*
2 – 4	0.04	0.050	0.468
3 – 4	-0.10	0.017	0.020

**9.3.4 Intrinsische Motivation und Interesse (Mittelstufe - Schulform)***Tabelle 198: Post-hoc-Analyse des Haupteffekts „Messzeitpunkt“ (Gymnasium - Gesamtschule).*

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	0.12	0.025	< 0.001*
1 – 3	0.36	0.010	< 0.001*
1 – 4	0.42	0.009	< 0.001*
2 – 3	0.27	0.017	< 0.001*
2 – 4	0.26	0.013	< 0.001*
3 – 4	0.02	0.050	0.567

### 9.3.5 Lernhilfen und Vertiefungsaufgaben (Mittelstufe)

#### Nützlichkeit der Lernhilfen

Tabelle 199: Deskriptive Statistik der Variable "Nützlichkeit der Lernhilfen" in der Interventions- und Vergleichsgruppe der Mittelstufe.

	<i>Interventionsgruppe</i>			
Messzeitpunkt	1	2	3	4
Stichprobenumfang	262	267	231	229
Mittelwert	4.47	4.38	3.97	3.85
(Standardabweichung)	(0.80)	(0.89)	(1.14)	(1.25)
	<i>Vergleichsgruppe</i>			
Messzeitpunkt	1	2	3	4
Stichprobenumfang	99	101	88	87
Mittelwert	3.56	3.76	3.11	2.95
(Standardabweichung)	(1.34)	(1.12)	(1.34)	(1.37)

#### Bevorzugte Nutzungsvariante der Lernhilfen (digital / analog)

Tabelle 200: Deskriptive Statistik der bevorzugten Art der Hilfenutzung (digital / analog) in der Mittelstufe.

	<i>Interventionsgruppe</i>	
Messzeitpunkt	1	2
Stichprobenumfang	242	265
Mittelwert	4.40 (0.95)	4.43 (1.19)
(Standardabweichung)		

### **Gründe für die Nutzung der Lernhilfen und Vertiefungsaufgaben (Interventionsgruppe)**

*Tabelle 201: Deskriptive Statistik für die Gründe der Nutzung der Lernhilfen in der Interventionsgruppe der Mittelstufe.*

<b>H3</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	260	265	231	227
Mittelwert	3.36	2.88	2.71	2.76
(Standardabweichung)	(1.34)	(1.45)	(1.45)	(1.50)
<b>H4</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	260	265	231	227
Mittelwert	2.58	3.54	3.36	3.69
(Standardabweichung)	(1.37)	(1.30)	(1.38)	(1.32)
<b>H5</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	260	266	231	229
Mittelwert	1.85	1.92	1.93	2.14
(Standardabweichung)	(1.17)	(1.28)	(1.19)	(1.34)
<b>H6</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	261	265	231	228
Mittelwert	3.49	3.39	3.303	3.08
(Standardabweichung)	(1.31)	(1.42)	(1.43)	(1.51)



*Tabelle 202: Deskriptive Statistik für die Gründe der Nutzung der Vertiefungsaufgabe in der Interventionsgruppe der Mittelstufe.*

<b>V1</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	282	257	204	143
Mittelwert	3.35	3.22	2.92	2.89
(Standardabweichung)	(1.47)	(1.55)	(1.54)	(1.54)
<b>V2</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	282	257	203	144
Mittelwert	3.83	3.68	3.31	2.96
(Standardabweichung)	(1.26)	(1.35)	(1.47)	(1.49)
<b>V3</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	282	254	204	144
Mittelwert	1.88	2.13	2.23	2.26
(Standardabweichung)	(1.21)	(1.40)	(1.43)	(1.36)
<b>V4</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	282	253	204	144
Mittelwert	3.20	3.33	3.02	2.98
(Standardabweichung)	(1.29)	(1.39)	(1.43)	(1.48)
<b>V5</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	281	256	204	144
Mittelwert	3.82	3.70	3.58	3.12
(Standardabweichung)	(1.45)	(1.49)	(1.49)	(1.52)

### **Gründe für die Nutzung der Lernhilfen und Vertiefungsaufgaben (Vergleichsgruppe)**

*Tabelle 203: Deskriptive Statistik für die Gründe der Nutzung der Lernhilfen in Vergleichsgruppe der Mittelstufe.*

<b>H3</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	99	103	88	88
Mittelwert	2.61	2.68	2.43	2.22
(Standardabweichung)	(1.56)	(1.40)	(1.29)	(1.29)
<b>H4</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	99	103	87	87
Mittelwert	2.18	3.22	3.25	3.56
(Standardabweichung)	(1.40)	(1.37)	(1.43)	(1.40)
<b>H5</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	100	103	86	88
Mittelwert	2.17	2.05	1.96	2.10
(Standardabweichung)	(1.46)	(1.22)	(1.17)	(1.24)
<b>H6</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	100	103	86	88
Mittelwert	2.67	2.85	2.46	2.44
(Standardabweichung)	(1.47)	(1.44)	(1.35)	(1.36)

*Tabelle 204: Deskriptive Statistik für die Gründe der Nutzung der Vertiefungsaufgabe in der Vergleichsgruppe der Mittelstufe.*

<b>V1</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	144	135	103	64
Mittelwert	3.53	3.24	3.36	2.81
(Standardabweichung)	(1.49)	(1.56)	(1.47)	(1.67)
<b>V2</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	143	135	103	64
Mittelwert	3.71	3.39	3.26	2.97
(Standardabweichung)	(1.36)	(1.42)	(1.48)	(1.63)
<b>V3</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	144	135	103	64
Mittelwert	2.59	2.53	2.71	2.41
(Standardabweichung)	(1.47)	(1.44)	(1.39)	(1.50)
<b>V4</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	144	145	103	64
Mittelwert	2.83	3.06	3.00	2.62
(Standardabweichung)	(1.39)	(1.34)	(1.44)	(1.63)
<b>V5</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	144	136	103	64
Mittelwert	4.24	3.92	3.50	3.33
(Standardabweichung)	(1.16)	(1.24)	(1.43)	(1.66)

**Variablen für die zur Verfügung stehende Zeit (Interventionsgruppe)***Tabelle 205: Deskriptive Statistik für die Variablen der Zeit in der Interventionsgruppe der Mittelstufe.*

Messzeitpunkt	<i>Z1</i>			
	1	2	3	4
Stichprobenumfang	317	317	317	317
Mittelwert	4.53	4.29	4.24	3.93
(Standardabweichung)	(0.88)	(1.06)	(1.08)	(1.25)

Messzeitpunkt	<i>Z2</i>			
	1	2	3	4
Stichprobenumfang	317	317	317	317
Mittelwert	4.22	4.00	4.15	3.94
(Standardabweichung)	(1.15)	(1.29)	(1.20)	(1.30)

**Variablen für die zur Verfügung stehende Zeit (Vergleichsgruppe)***Tabelle 206: Deskriptive Statistik für die Variablen der Zeit in der Vergleichsgruppe der Mittelstufe.*

Messzeitpunkt	<i>Z1</i>			
	1	2	3	4
Stichprobenumfang	148	148	148	148
Mittelwert	4.80	4.60	4.38	4.04
(Standardabweichung)	(0.53)	(0.78)	(0.99)	(1.18)

Messzeitpunkt	<i>Z2</i>			
	1	2	3	4
Stichprobenumfang	148	148	148	148
Mittelwert	4.76	4.51	4.36	4.06
(Standardabweichung)	(0.71)	(0.86)	(1.07)	(1.13)

**Nutzungshäufigkeit Lernhilfen**

Tabelle 207: Unkorrigierte Hilfenutzung pro Schüler:in pro Stunde am Gymnasium (n = 205).

	<i>Stunde 1</i>	<i>Stunde 2</i>	<i>Stunde 3</i>	<i>Stunde 4</i>
Mittelwert (Standardabweichung)	3.41 (2.33)	4.42 (3.17)	2.30 (2.02)	2.30 (2.17)

Tabelle 208: Unkorrigierte Hilfenutzung pro Schüler:in pro Stunde an der Gesamtschule (n = 115).

	<i>Stunde 1</i>	<i>Stunde 2</i>	<i>Stunde 3</i>	<i>Stunde 4</i>
Mittelwert (Standardabweichung)	3.02 (2.58)	4.99 (3.59)	1.57 (1.55)	2.28 (2.11)

Tabelle 209: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Geschlecht“ (hier: weiblich).

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	-0.005	0.017	< 0.001*
1 – 3	0.002	0.025	0.026
1 – 4	< -0.001	0.050	0.871
2 – 3	0.007	0.009	< 0.001*
2 – 4	0.005	0.010	< 0.001*
3 – 4	- 0.003	0.013	< 0.001*

Tabelle 210: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Geschlecht“ (hier: männlich).

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	- 0.005	0.050	0.640
1 – 3	0.003	0.013	< 0.001*
1 – 4	0.003	0.017	0.001*
2 – 3	0.004	0.009	< 0.001*
2 – 4	0.004	0.010	< 0.001*
3 – 4	< -0.001	0.025	0.076

### 9.3.6 Lernhilfen und Vertiefungsaufgaben (Oberstufe)

#### Nutzungshäufigkeit Lernhilfen

Tabelle 211: Unkorrigierte Hilfenutzung pro Schüler:in pro Stunde am Gymnasium (n = 101).

	<i>Stunde 1</i>	<i>Stunde 2</i>	<i>Stunde 3</i>	<i>Stunde 4</i>
Mittelwert (Standardabweichung)	3.50 (2.86)	2.69 (2.98)	5.73 (5.26)	3.67 (3.37)

#### Nützlichkeit der Lernhilfen

Tabelle 212: Deskriptive Statistik der Variable "Nützlichkeit der Lernhilfen" in der Interventions- und Vergleichsgruppe der Oberstufe.

	<i>Interventionsgruppe</i>			
Messzeitpunkt	1	2	3	4
Stichprobenumfang	84	82	86	79
Mittelwert (Standardabweichung)	4.13 (1.01)	4.22 (0.97)	4.38 (0.91)	4.39 (0.85)
	<i>Vergleichsgruppe</i>			
Messzeitpunkt	1	2	3	4
Stichprobenumfang	26	26	26	28
Mittelwert (Standardabweichung)	3.65 (1.38)	3.54 (1.17)	4.00 (1.10)	3.46 (1.32)

#### Bevorzugte Nutzungsvariante der Lernhilfen (digital / analog)

Tabelle 213: Deskriptive Statistik der bevorzugten Art der Hilfenutzung (digital / analog) in der Oberstufe.

	<i>Interventionsgruppe</i>	
Messzeitpunkt	1	2
Stichprobenumfang	91	78
Mittelwert (Standardabweichung)	4.41 (0.94)	4.42 (0.89)

### **Gründe für die Nutzung der Lernhilfen und Vertiefungsaufgaben (Interventionsgruppe)**

*Tabelle 214: Deskriptive Statistik für die Gründe der Nutzung der Lernhilfen in der Interventionsgruppe der Oberstufe.*

<b>H3</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	84	82	86	79
Mittelwert	3.63	3.38	3.00	2.98
(Standardabweichung)	(1.37)	(1.39)	(1.42)	(1.40)
<b>H4</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	83	82	86	78
Mittelwert	1.94	3.08	3.60	3.65
(Standardabweichung)	(1.14)	(1.36)	(1.27)	(1.38)
<b>H5</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	84	82	86	79
Mittelwert	2.35	2.12	2.26	2.08
(Standardabweichung)	(1.48)	(1.33)	(1.47)	(1.32)
<b>H6</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	84	82	86	78
Mittelwert	3.42	3.43	3.36	3.30
(Standardabweichung)	(1.31)	(1.32)	(1.34)	(1.36)

*Tabelle 215: Deskriptive Statistik für die Gründe der Nutzung der Vertiefungsaufgabe in der Interventionsgruppe der Oberstufe.*

<b>V1</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	103	57	48	41
Mittelwert	3.70	3.79	3.58	3.32
(Standardabweichung)	(1.35)	(1.48)	(1.60)	(1.52)
<b>V2</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	103	57	48	40
Mittelwert	4.13	3.91	3.31	3.32
(Standardabweichung)	(1.06)	(1.18)	(1.55)	(1.62)
<b>V3</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	103	57	48	40
Mittelwert	2.27	2.30	2.04	2.25
(Standardabweichung)	(1.34)	(1.36)	(1.40)	(1.41)
<b>V4</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	103	57	48	40
Mittelwert	3.26	2.98	3.00	3.08
(Standardabweichung)	(1.20)	(1.34)	(1.47)	(1.38)
<b>V5</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	103	57	48	40
Mittelwert	4.34	3.74	3.10	3.08
(Standardabweichung)	(1.09)	(1.37)	(1.64)	(1.53)



## Gründe für die Nutzung der Lernhilfen und Vertiefungsaufgaben (Vergleichsgruppe)

Tabelle 216: Deskriptive Statistik für die Gründe der Nutzung der Lernhilfen in Vergleichsgruppe der Oberstufe.

<b>H3</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	29	26	30	30
Mittelwert	3.55	2.69	2.37	2.23
(Standardabweichung)	(1.62)	(1.59)	(1.38)	(1.50)
<b>H4</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	30	26	30	30
Mittelwert	1.77	3.00	3.30	3.27
(Standardabweichung)	(1.19)	(1.44)	(1.44)	(1.60)
<b>H5</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	29	25	27	30
Mittelwert	2.66	1.72	1.96	1.80
(Standardabweichung)	(1.56)	(1.00)	(1.34)	(1.22)
<b>H6</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	29	26	28	30
Mittelwert	3.17	3.00	2.68	2.17
(Standardabweichung)	(1.42)	(1.36)	(1.36)	(1.46)

*Tabelle 217: Deskriptive Statistik für die Gründe der Nutzung der Vertiefungsaufgabe in der Vergleichsgruppe der Mittelstufe.*

<b>V1</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	49	27	19	17
Mittelwert	3.88	4.11	3.37	3.59
(Standardabweichung)	(1.41)	(1.22)	(1.57)	(1.80)
<b>V2</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	51	28	19	17
Mittelwert	4.12	3.82	3.42	3.47
(Standardabweichung)	(0.99)	(1.16)	(1.46)	(1.77)
<b>V3</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	41	28	18	16
Mittelwert	2.63	2.25	2.17	2.25
(Standardabweichung)	(1.56)	(1.30)	(1.30)	(1.69)
<b>V4</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	50	28	19	17
Mittelwert	3.14	3.29	2.90	2.47
(Standardabweichung)	(1.34)	(1.41)	(1.52)	(1.77)
<b>V5</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	51	28	19	18
Mittelwert	4.41	3.54	3.21	2.94
(Standardabweichung)	(1.15)	(1.40)	(1.69)	(1.73)

**Variablen für die zur Verfügung stehende Zeit (Interventionsgruppe)***Tabelle 218: Deskriptive Statistik für die Variablen der Zeit in der Interventionsgruppe der Oberstufe.*

	<i>Z1</i>			
	1	2	3	4
Messzeitpunkt				
Stichprobenumfang	108	108	108	108
Mittelwert	4.74	4.41	4.13	4.23
(Standardabweichung)	(0.57)	(0.89)	(0.99)	(0.95)

	<i>Z2</i>			
	1	2	3	4
Messzeitpunkt				
Stichprobenumfang	108	108	108	108
Mittelwert	4.52	4.26	3.88	4.12
(Standardabweichung)	(0.77)	(0.95)	(1.18)	(1.17)

**Variablen für die zur Verfügung stehende Zeit (Vergleichsgruppe)***Tabelle 219: Deskriptive Statistik für die Variablen der Zeit in der Vergleichsgruppe der Oberstufe.*

	<i>Z1</i>			
	1	2	3	4
Messzeitpunkt				
Stichprobenumfang	51	51	51	51
Mittelwert	4.67	3.67	3.47	4.06
(Standardabweichung)	(0.86)	(1.35)	(1.54)	(1.10)

	<i>Z2</i>			
	1	2	3	4
Messzeitpunkt				
Stichprobenumfang	51	51	51	51
Mittelwert	4.71	3.41	3.65	3.98
(Standardabweichung)	(0.83)	(1.53)	(1.47)	(1.26)

### 9.3.7 Tabletbezogene Variablen (Mittelstufe)

Tabelle 220: Deskriptive Statistik für die Variablen TAB1 und TAB 2 in der Interventionsgruppe der Mittelstufe.

<b>TAB 1</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	324			
Mittelwert	1.47	1.42	1.49	1.46
(Standardabweichung)	(0.82)	(0.81)	(0.95)	(0.98)

<b>TAB 2</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	324			
Mittelwert	1.42	1.38	1.50	1.42
(Standardabweichung)	(0.94)	(0.76)	(0.94)	(0.89)

Tabelle 221: Deskriptive Statistik für die Variable TAB3 in der Interventionsgruppe der Mittelstufe.

<b>TAB 3</b>		
Messzeitpunkt	1	2
Stichprobenumfang	300	
Mittelwert	1.30 (0.70)	1.53 (0.92)
(Standardabweichung)		

### 9.3.8 Tabletbezogene Variablen (Oberstufe)

*Tabelle 222: Deskriptive Statistik für die Variablen TAB 1 und TAB 2 in der Interventionsgruppe der Oberstufe.*

<b>TAB 1</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	109			
Mittelwert	1.69	1.62	1.75	1.71
(Standardabweichung)	(0.94)	(1.04)	(1.21)	(1.23)

<b>TAB 2</b>				
Messzeitpunkt	1	2	3	4
Stichprobenumfang	109			
Mittelwert	1.53	1.54	1.78	1.68
(Standardabweichung)	(0.77)	(0.93)	(1.20)	(1.19)

*Tabelle 223: Deskriptive Statistik für die Variable TAB3 in der Interventionsgruppe der Oberstufe.*

<b>TAB 3</b>		
Messzeitpunkt	1	2
Stichprobenumfang	100	
Mittelwert	1.45 (0.86)	1.64 (0.94)
(Standardabweichung)		

### 9.3.9 Cognitive Load und Mental Effort (Mittelstufe)

#### Intrinsic Load

Tabelle 224: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Gruppe“ (hier: Interventionsgruppe).

Stundenvergleich	$\hat{\Psi}$	p-Wert-Schwelle (adjus-	p-Wert
1 – 2	-0.31	0.013	< 0.001*
1 – 3	-0.54	0.010	< 0.001*
1 – 4	-0.70	0.009	< 0.001*
2 – 3	-0.16	0.025	0.015
2 – 4	-0.29	0.017	< 0.001*
3 – 4	-0.13	0.050	0.049

Tabelle 225: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Gruppe“ (hier: Vergleichsgruppe).

Stundenvergleich	$\hat{\Psi}$	p-Wert-Schwelle (adjus-	p-Wert
1 – 2	-0.58	0.010	< 0.001*
1 – 3	-0.53	0.017	< 0.001*
1 – 4	-0.99	0.009	< 0.001*
2 – 3	0.064	0.050	0.220
2 – 4	-0.29	0.025	< 0.001*
3 – 4	-0.40	0.013	< 0.001*

Tabelle 226: Post-hoc-Analyse Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Geschlecht“ (hier: weiblich).

Stundenvergleich	$\hat{\Psi}$	p-Wert-Schwelle (adjus-	p-Wert
1 – 2	-0.50	0.010	< 0.001*
1 – 3	-0.59	0.013	< 0.001*
1 – 4	-1.01	0.009	< 0.001*
2 – 3	-0.10	0.050	0.0185*
2 – 4	-0.43	0.025	< 0.001*
3 – 4	-0.52	0.017	< 0.001*

Tabelle 227: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Geschlecht“ (hier: männlich).

Stundenvergleich	$\hat{\Psi}$	p-Wert-Schwelle (adjus-	p-Wert
1 – 2	-0.34	0.013	< 0.001*
1 – 3	-0.48	0.010	< 0.001*
1 – 4	-0.62	0.009	< 0.001*
2 – 3	-0.018	0.050	0.735
2 – 4	-0.18	0.017	0.012*
3 – 4	-0.13	0.025	0.076

### **Extraneous Load**

Tabelle 228: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Geschlecht“ (hier: weiblich).

Stundenvergleich	$\hat{\Psi}$	p-Wert-Schwelle (adjus-	p-Wert
1 – 2	-0.17	0.025	0.008*
1 – 3	-0.40	0.009	< 0.001*
1 – 4	-0.43	0.010	< 0.001*
2 – 3	-0.25	0.017	< 0.001*
2 – 4	-0.31	0.013	< 0.001*
3 – 4	-0.06	0.050	0.410

Tabelle 229: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Geschlecht“ (hier: männlich).

Stundenvergleich	$\hat{\Psi}$	p-Wert-Schwelle (adjus-	p-Wert
1 – 2	-0.15	0.025	0.043
1 – 3	-0.15	0.017	0.035
1 – 4	-0.36	0.009	< 0.001*
2 – 3	0.03	0.050	0.559
2 – 4	-0.18	0.013	0.012*
3 – 4	-0.22	0.010	< 0.001*

*Tabelle 230: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Gruppe“ (hier: Interventionsgruppe).*

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	-0.06	0.025	0.147
1 – 3	-0.29	0.013	< 0.001*
1 – 4	-0.31	0.009	< 0.001*
2 – 3	-0.16	0.017	0.013*
2 – 4	-0.23	0.010	< 0.001*
3 – 4	-0.08	0.050	0.222

*Tabelle 231: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Gruppe“ (hier: Vergleichsgruppe).*

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	-0.32	0.010	< 0.001*
1 – 3	-0.30	0.013	< 0.001*
1 – 4	-0.53	0.009	< 0.001*
2 – 3	-0.03	0.050	0.619
2 – 4	-0.26	0.017	0.002*
3 – 4	-0.24	0.025	0.002*

### **Mental Effort**

*Tabelle 232: Post-hoc-Analyse des Haupteffekts „Messzeitpunkt“ (Interventionsgruppe - Vergleichsgruppe).*

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	-0.37	0.013	< 0.001*
1 – 3	-0.64	0.010	< 0.001*
1 – 4	-0.68	0.009	< 0.001*
2 – 3	-0.18	0.025	0.008*
2 – 4	-0.20	0.017	0.004*
3 – 4	-0.07	0.050	0.239



### 9.3.10 Cognitive Load und Mental Effort (Oberstufe)

#### Mental Effort

Tabelle 233: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Gruppe“ (hier: Interventionsgruppe).

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	0.54	0.013	< 0.001*
1 – 3	-0.95	0.009	< 0.001*
1 – 4	-0.54	0.010	< 0.001*
2 – 3	-0.41	0.025	0.002*
2 – 4	-0.04	0.050	0.725
3 – 4	0.27	0.017	< 0.001*

Tabelle 234: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Gruppe“ (hier: Vergleichsgruppe).

<i>Stundenvergleich</i>	$\hat{\Psi}$	<i>p-Wert-Schwelle (adjus-</i>	<i>p-Wert</i>
1 – 2	-1.29	0.009	< 0.001*
1 – 3	-1.77	0.010	< 0.001*
1 – 4	-1.11	0.013	< 0.001*
2 – 3	-0.32	0.025	0.253
2 – 4	0.12	0.050	0.525
3 – 4	0.60	0.017	0.008*

### 9.3.11 Cognitive Load und Mental Effort (Schulform - Mittelstufe)

#### Intrinsic Load

Tabelle 235: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Schulform“ (hier: Gymnasium).

Stundenvergleich	$\hat{\Psi}$	p-Wert-Schwelle (adjus-	p-Wert
1 – 2	-0.31	0.013	< 0.001*
1 – 3	-0.54	0.010	< 0.001*
1 – 4	-0.70	0.009	< 0.001*
2 – 3	-0.16	0.025	0.015*
2 – 4	-0.29	0.017	< 0.001*
3 – 4	-0.13	0.050	0.049*

Tabelle 236: Post-hoc-Analyse des Innersubjektfaktors „Messzeitpunkt“ der Interaktion „Messzeitpunkt\*Schulform“ (hier: Gesamtschule).

Stundenvergleich	$\hat{\Psi}$	p-Wert-Schwelle (adjus-	p-Wert
1 – 2	0.013	0.013	< 0.001*
1 – 3	0.009	0.009	< 0.001*
1 – 4	0.010	0.010	< 0.001*
2 – 3	0.025	0.025	0.141
2 – 4	0.050	0.050	0.682
3 – 4	0.017	0.017	0.002

#### Mental Effort

Tabelle 237: Post-hoc-Analyse des Haupteffekts „Messzeitpunkt“ (Gymnasium – Gesamtschule).

Stundenvergleich	$\hat{\Psi}$	p-Wert-Schwelle (adjus-	p-Wert
1 – 2	-0.35	0.009	< 0.001*
1 – 3	-0.42	0.013	< 0.001*
1 – 4	-0.43	0.010	< 0.001*
2 – 3	-0.13	0.017	0.072
2 – 4	-0.09	0.025	0.188
3 – 4	< -0.01	0.050	0.942