

Professionelle Unterrichtswahrnehmung in der Lehrpersonenbildung fördern

Videobasierte Analyse motivationsrelevanter Aspekte im Unterricht

genehmigte Dissertation

zur Erlangung des akademischen Grades eines Doktors der Philosophie
am Fachbereich Erziehungswissenschaften
der RPTU Rheinland-Pfälzisch Technischen Universität Kaiserslautern-Landau
vorgelegt am 04.07.2025

von Madana Treiber

Mündliche Prüfung am 10.11.2025

Erstgutachterin: Prof. Dr. Gerlinde Lenske

Zweitgutachter: Prof. Dr. Jürgen Roth

Vorsitz des Promotionsausschusses: Prof. Dr. Pascal Bastian

Lebenslauf – Madana Treiber

BILDUNGSWEG

10/2010 – 11/2015 Diplom-Medizinpädagogin an der Charité-Universitätsmedizin Berlin

BERUFLICHE STATIONEN

Seit 10/2025 Wissenschaftliche Mitarbeiterin im Institut für Bildung im Kindes- und Jugendalter im Arbeitsbereich Grundschulpädagogik an der RPTU am Standort Landau

Seit 07/2023 Wissenschaftliche Mitarbeiterin im Projekt WebStuPlan im Zentrum für Lehrkräftebildung an der RPTU am Standort Landau

04/2016 – 06/2023 Wissenschaftliche Mitarbeiterin im Projekt DiAmant M an der Universität Koblenz · Landau am Standort Landau (neu fusioniert zur RPTU)

Inhaltsverzeichnis

Danksagung.....	iv
Zusammenfassung.....	vii
1 Einleitung.....	1
2 Motivation und Motivierung im Unterricht	5
2.1 Motivation durch Bedürfnisbefriedigung.....	5
2.2 Motivierung durch Zielorientierung im Lernkontext.....	10
2.2.1 Vom TARGET- zum IBAS-Modell	11
2.2.2 Vier Dimensionen zur Förderung der Motivation	12
2.3 Messung der Motivation im Lernkontext	16
2.4 Zusammenfassung in Bezug auf die vorliegende Erhebung.....	19
3 Professionelle Unterrichtswahrnehmung.....	21
3.1 Überblick angrenzender Konstrukte	21
3.2 Prozess der Professionellen Unterrichtswahrnehmung	26
3.3 Zusammenfassung in Bezug auf die vorliegende Erhebung.....	29
3.4 Forschungslage zur Professionellen Unterrichtswahrnehmung.....	31
3.4.1 Professionellen Unterrichtswahrnehmung: Erfassen und Fördern	31
3.4.2 Expertiseforschung	35
3.4.3 Zusammenfassung in Bezug auf die vorliegende Erhebung.....	36
4 Einsatz von Videos in der Lehre	38
4.1 Die fünf Heuristiken	38
4.2 Studienlage.....	45
4.2.1 Studien zum OBSERVER	45
4.2.2 Weitere Studien zur Förderung der Professionellen Wahrnehmung.....	51
4.3 Das Videotool ViviAn.....	56
4.3.1 Lehransatz und Lernziele	56
4.3.2 Beschreibung der Lernumgebung.....	57
4.3.3 Bisherige Forschungsarbeiten mit ViviAn	59
4.4 Zusammenfassung in Bezug auf die vorliegende Erhebung.....	62

5	Gütekriterien der Testentwicklung.....	66
5.1	Objektivität	66
5.2	Reliabilität	68
5.3	Validität	69
5.4	Nebenkriterien	71
5.5	Zusammenfassung in Bezug auf die vorliegende Erhebung.....	72
6	Zielsetzung und Fragestellung.....	80
7	Entwicklung und Validierung der Videovignetten	84
7.1	Stichprobe	84
7.2	Methodisches Vorgehen	84
7.2.1	Videografie mit Unterrichtsadaption.....	85
7.2.2	Validierung der Videovignetten.....	87
7.3	Ergebnisse	92
7.4	Interpretation und Schlussfolgerungen	101
7.5	Beschreibung der Vignetten	102
8	Entwicklung und Validierung der Analyseaufträge	106
8.1	Entwicklung der Analyseaufträge	106
8.2	Validierung der Analyseaufträge	110
8.2.1	Stichprobe.....	110
8.2.2	Methodisches Vorgehen	111
8.2.3	Ergebnisse und Interpretation	112
8.3	Implementierung in ViviAn.....	113
9	Akzeptanz des Tools ViviAn – Pilotierungsstudie.....	115
9.1	Stichprobe	115
9.2	Methodisches Vorgehen	115
9.3	Ergebnisse	117
9.4	Interpretation und Schlussfolgerungen	118

10	Validierung des Videotools Teil I: Raschskalierung	120
10.1	Stichprobe	120
10.2	Methodisches Vorgehen	121
10.2.1	Item-Response-Theorie	121
10.2.2	Raschskalierung	126
10.3	Ergebnisse	133
10.4	Interpretation und Schlussfolgerungen	145
11	Validierung des Videotools Teil II: Erhebung der Professionellen Wahrnehmung	150
11.1	Stichprobe	150
11.2	Methodisches Vorgehen	151
11.3	Ergebnisse	152
11.4	Interpretation und Schlussfolgerungen	153
12	Validierung des Tools Teil III: Förderung der Professionellen Wahrnehmung	156
12.1	Stichprobe	156
12.2	Beschreibung des Seminars	157
12.3	Methodisches Vorgehen	157
12.4	Ergebnisse	159
12.5	Interpretation und Schlussfolgerungen	160
13	Zusammenfassung der Ergebnisse.....	164
14	Diskussion	171
15	Perspektiven.....	184
	Literaturverzeichnis.....	xxi
	Abbildungsverzeichnis.....	xxxviii
	Tabellenverzeichnis.....	xl
	Anhang	xxiii

Danksagung

Zu Beginn kaum vorstellbar: 71.000 Wörter zu schreiben, mehr als eine halbe Million Zeichen zu tippen – oder vermutlich noch mehr, wenn man all die gelöschten, umformulierten und überarbeiteten Worte mitzählt. In den letzten Jahren durfte ich am Institut für Mathematik mit vielen Zahlen in Berührung kommen und das, obwohl dies eigentlich nicht mein eigener Fachbereich ist. Umso dankbarer bin ich für die herzliche Aufnahme, die tatkräftige Unterstützung und den Humor, der mir dort den Arbeitsalltag versüßt hat.

Mein besonderer Dank gilt meiner Betreuerin **Prof. Dr. Gerlinde Lenske** und meinem Betreuer **Prof. Dr. Jürgen Roth**. Danke für eure fachliche Expertise, die konstruktiven Gespräche auf Augenhöhe und das Vertrauen in meine Arbeit. Eure wertvollen Impulse haben das Projekt und mich weitergebracht, euer offenes Ohr war eine große Unterstützung. Danke, dass ihr mir die Freiheit gelassen habt, meine eigenen Wege zu gehen, und mich dabei gleichzeitig verlässlich begleitet habt.

Natürlich gilt mein größter Dank meiner Familie. Ihr wart immer mit Geduld, Verständnis und der richtigen Portion Zuspruch an meiner Seite.

Moritz – du hast mich so oft besänftigt und holst mich immer wieder genau dort ab, wo ich gerade stehe. Du bist mein Ruhepol und größter Rückhalt. Bei dir kann ich all meine Sorgen, Zweifel und Probleme aussprechen – du hörst zu, sprichst mir unendlich Mut zu, stärkst mich und gibst mir das Gefühl, alles schaffen zu können. Es ist ein großes Geschenk, dich als so wunderbaren Begleiter an meiner Seite zu wissen, auf diesem langen Weg und für alles, was noch kommt.

Ulrike und Martin – danke für den legendären Großeltern-Mittwoch, für die liebevolle Betreuung der Rasselbande, das Milchreis-Kochen, das Wunscherfüllen und die spürbare Entlastung. Insbesondere danke ich dir, Ulrike, für das sorgfältige Korrekturlesen: das Aufspüren verlorengegangener Kommata, Buchstabenverdreher und das Korrigieren meiner berüchtigten Schwäche bei den Fällen – man sagt, dass wir aus Berlin etwas entspannter mit den Grammatikregeln umgehen.

Mama – du hast immer an mich geglaubt! Eigentlich muss ich gar nicht mehr sagen. Ohne dich würde ich heute nicht hier stehen. Auf dich konnte ich mich immer verlassen und du hast mich bei allem unterstützt!

Dietmar – danke, dass du unsere Familie so selbstverständlich bereicherst und die Rolle „Opi aus Berlin“ für unsere Kinder herzlich angenommen hast. Du bist immer für uns da, wenn Hilfe gebraucht wird. Es ist schön, dich an unserer Seite zu wissen.

Liebe **Patrizia** – auch dir gilt ein riesiges Dankeschön! Gemeinsam haben wir unzählige Stunden im Büro verbracht, methodisch diskutiert, uns in R durchgebissen und nie die Freude am Austausch verloren. Deine Unterstützung bis zur letzten Seite ist für mich unbezahlbar. Es ist ein großes Glück, nicht nur eine Kollegin, sondern eine so wunderbare Freundin gefunden zu haben.

Vielen Dank an alle **Mathematiker:innen** und ganz besonders an **Tanja**, die mich in den letzten Jahren mit viel Herz und einem stets offenen Ohr begleitet hat.

Und natürlich ein ganz großes Dankeschön an all die vielen anderen wunderbaren Menschen in unserem Leben, die uns begleiten, unterstützen und einfach da sind. Schön, dass es euch gibt!

Für Johan, Aron, Luisa und Mathilda.

Glaubt an eure Träume, eure Stärke und eure Fähigkeiten – ihr seid wundervoll!

Zusammenfassung

Die vorliegende Arbeit verbindet zwei wichtige Aspekte des Unterrichts miteinander. Zum einen die **Motivation und Motivierung**, die eine zentrale Rolle für den Lernerfolg spielen. Eine Lehrperson kann über motivationsförderliches Handeln Lernprozesse initiieren und begleiten. Sowohl die Motivation der Lernenden selbst als auch die Motivierung sind herausfordernde Aufgaben für Lehrpersonen. Sie müssen über das Wissen von motivationsförderlichen Strategien verfügen und dieses im Unterrichtsalltag auch tatsächlich anwenden können. Zudem ist die Motivation der Lernenden, die in den letzten Jahren insgesamt stetig abgenommen hat, als innerer Prozess nur von außen deutbar oder erfragbar.

Zum anderen kommt der **Professionellen Unterrichtswahrnehmung** dabei eine besondere Bedeutung zu. Hierbei handelt es sich um die Fähigkeit, relevante Aspekte wahrzunehmen (Noticing) und wissensbasiert zu beschreiben, zu erklären und Vorhersagen für das weitere Unterrichtsgeschehen treffen zu können (drei Teilprozesse des Knowledge-based Reasonings). Studien zeigen, dass die Professionelle Unterrichtswahrnehmung mit zunehmender Unterrichtserfahrung deutlich höher ausgeprägt ist als bei Noviz:innen. Um bereits Lehramtsstudierenden und Referendar:innen die Möglichkeit zu geben, diese wichtigen Fähigkeiten zu erlernen, kommen in der Forschung häufig Videovignetten mit Analyseaufträgen zu speziellen Themen wie zum Beispiel der Klassenführung und lernunterstützende Maßnahmen zum Einsatz.

Das Video-Tool ViviAn („Videovignetten zur Analyse von Unterrichtsprozessen“, <https://vivian-training.de>) bietet eine Lernumgebung, in die unter anderem Unterrichtsvideos, Arbeitsmaterialien und Analyseaufträge eingebettet werden können. Dieses Videotool wurde in der vorliegenden Arbeit um den Themenbereich der Motivierung mit den Aspekten der Relevanz, Leistungsdifferenzierung, individuelle Bezugsnormorientierung, Unterstützung der Autonomie sowie angemessene Strukturierung des Unterrichts erweitert. Die Validierung dieses Abschnitts innerhalb von ViviAn ist das zentrale Anliegen dieser Arbeit, um am Ende den Studierenden eine erweiterte videobasierte Lernumgebung zur Analyse ihrer Professionellen Unterrichtswahrnehmung hinsichtlich motivationsrelevanter Aspekte zur Verfügung stellen zu können. Der mehrstufige Prozess der Validierung setzt bei den entwickelten Videovignetten mit einem neuen Ansatz an. Die Videosequenzen wurden nicht wie bisher in der Forschung über Expert:innenurteile ausgewählt, sondern konnten über die tatsächlich erlebte Motivation der Schüler:innen inhaltlich validiert werden. Die Passung der Analyseaufträge sind über Expert:innenratings bestätigt. Mit einer IRT-Raschskalierung konnte die Konstruktvalidität im Sinne einer erwarteten Differenzierung zwischen den drei Phasen des Knowledge-based Reasonings mit reliablen Skalenwerten und modellkonformen Itemfitwerten aufgezeigt werden. Hinsichtlich der theoriegeleiteten Konstruktvalidität musste festgestellt werden, dass die Schwierigkeitsgrade innerhalb des Knowledge-based Reasoning-Prozesses nicht wie erwartet kontinuierlich vom Beschreiben über das Erklären zum Vorhersagen abnehmen, sondern dass

das Beschreiben für die Noviz:innen am schwierigsten war. Gründe dafür können im Zusammenhang mit den theoriebasierten Schwerpunkten und der besonderen Herausforderung der vorangestellten Wahrnehmung relevanter Ereignisse im Unterricht liegen. Zudem wurde bezüglich der gruppendifferenzierten Konstruktvalidität nachgewiesen, dass die Studierenden im Verlauf des Studiums keinen Fähigkeitszuwachs im Beschreiben und im Erklären erzielen. Dies lässt vermuten, dass sich erst mit zunehmender Erfahrung im Lehren die Professionelle Unterrichtswahrnehmung in diesen Bereichen verbessert. Jedoch zeigte sich im Vorhersagen, dass die Studierenden höheren Semesters besser Auswirkungen auf den weiteren Lehr-Lernverlauf ableiten können als Studienanfänger:innen. Das Zusammenspiel mehrerer Faktoren wie z.B. implizite Praxiserfahrungen, leichteres situatives Antizipieren oder Priming-Effekte kann dies möglicherweise bewirken.

Das Kriterium der Änderungssensitivität, inwieweit eine Entwicklung der Professionellen Unterrichtswahrnehmung nach einer Intervention in Form eines themenbezogenen Seminars aufgezeigt werden kann, wurde in einem Pre-Post-Design untersucht. In allen drei Bereichen des Knowledge-based Reasonings sind Veränderungen zu verzeichnen. Entgegen der Erwartung entwickelt sich die Ausprägung der Professionellen Unterrichtswahrnehmung im Beschreiben negativ, möglicherweise auf Grund der Schwerpunkte innerhalb des Seminarkonzepts. Beim Erklären und Vorhersagen sind die Studierenden beim Post-Test im Mittel besser geworden.

Diese Arbeit bestätigt neben den bisherigen Forschungsarbeiten, dass mit ViviAn eine gut gestaltete, relevant und interessant wahrgenommene Lernumgebung existiert, die in der universitären Lehramtsbildung als Analysetool mit direkten Anknüpfungspunkten zur Praxis eingesetzt werden kann.

Teil 1
Theoretische Grundlagen

1 Einleitung

Die Motivation bildet seit Jahrzehnten ein zentrales Thema der Bildungsforschung, sie ist bedeutsam für den Lernprozess und eine erfolgreiche Lernentwicklung (Deci und Ryan 1993, S. 235–236; Schiefele und Streblov 2006, S. 232; Spinath 2005, S. 203; Krapp und Hascher 2014b, S. 269). Zudem gilt sie als Prädiktor für die Schulleistung (Steinmayr und Spinath 2007). In einer einfachen Formel dargelegt, setzt sich der Lernerfolg durch zwei Faktoren zusammen (Stangl 2003, S. 9):

$$\text{Lernerfolg} = \text{investierte Mittel} \times \text{motivationaler Faktor}$$

Die Motivation als innere Einstellung zum Unterricht wie zum Beispiel das Interesse am Thema, die emotionale Bindung zur Lehrperson oder die Freude am Lernen sind durch das Zusammenwirken mit den investierten Mitteln, die sich aus den Ressourcen wie Vorwissen, Lernzeit, Konzentrationsfähigkeit bilden, zentral für ein gelungenes Lernen (Stangl 2003, S. 9). Wenn man davon ausgeht, dass Lehrpersonen bzw. das Bildungssystem zwar einen relevanten Einfluss auf die investierten Mittel wie zum Beispiel die Arbeitszeit haben, die motivationalen Faktoren der Schüler:innen aber nur im geringen Ausmaß durch die Lehrpersonen beeinflusst werden können (Stangl 2003, S. 9), dann bleiben bei geringer intrinsischer Motivation selbst große Anstrengungen wirkungslos. Umso prekärer ist es, dass die Motivation der Schüler:innen nicht nur im Verlauf der Schullaufbahn abnimmt (Gottfried et al. 2001, S. 10; Freiberger und Spinath 2014), sondern dass auch im 10 Jahres-Vergleich dieser Abwärtstrend sichtbar ist (Spinath et al. 2016). Darüber hinaus sind Kinder mit einer geringen Motivation zu Beginn ihrer Schulzeit auch im weiteren Bildungsverlauf stärker benachteiligt (Gottfried et al. 2001, S. 10), weil das aktuelle Schulsystem wenig Raum für individualisiertes und selbstgesteuertes Lernen bietet (Stangl 2003, S. 9–10). Um diesem Verlauf entgegenzusteuern, können Lehrpersonen durch Berücksichtigung von Aspekten wie Beziehungsaufbau, Relevanz, Selbstbestimmung, Fehlerklima und Aktivierung den Lernerfolg der Schüler:innen positiv beeinflussen (Stangl 2003, S. 9–12). Somit hat die Gestaltung des Unterrichts letztlich doch direkt Auswirkungen auf die Motivation der Lernenden (Helmke 2015; Kunter et al. 2011; Brünken et al. 2019, S. 186). „Ohne Motivation bzw. ohne ihre Förderung und Aufrechterhaltung ist Unterricht auf Dauer nicht lernwirksam (Florio-Hansen 2014, S. 61).“ Dieser Aussage schließen sich auch Brünken et al. (2019, S. 186) an. Demnach spielt die Motivierung eine zentrale Rolle für das erfolgreiche Lernen. Wenn es gelingt, dass sich Schüler:innen durch unterstützende Rahmenbedingungen und motivationsförderliches Handeln mit Lerninhalten aktiv auseinandersetzen und sie als relevant wahrnehmen, dann trägt dies wesentlich zum wirksamen Lernprozess bei. Das Erlernen und Erleben der eigenen Motivation im Rahmen der

Bildungsprozesse hat nicht nur Effekte auf den aktuellen Unterricht, sondern beeinflusst das lebenslange Lernen (Brünken et al. 2019, S. 186).

In dieser Arbeit wird von dem Ansatz ausgegangen, dass dem Trend der wahrgenommenen sinkenden Motivation durch die Förderung motivationsunterstützenden Handelns der Lehrpersonen entgegengewirkt werden kann. Nicht nur das Wissen über Motivationstheorien und Aspekte motivierenden Unterrichtshandelns, sondern auch die Fähigkeit, Unterricht analysieren zu können, ist für ein erfolgreiches Handeln entscheidend (Kersting et al. 2010; Schäfer und Seidel 2015; Seidel et al. 2010). Lehrpersonen müssen (motivations-)relevante Ereignisse im Unterricht gezielt wahrnehmen und wissensbasiert verarbeiten, als Voraussetzung dafür, flexibel und angemessen auf das Geschehen und Lernen der Schüler:innen reagieren zu können (Seidel et al. 2010; Sherin und van Es 2008). Diese Fähigkeit ist stark von der Unterrichtserfahrung abhängig (Berliner 2001, 1987; Carter et al. 1988; Seidel und Prenzel 2007) und wird in der Literatur als Professionelle Unterrichtswahrnehmung bezeichnet (Seidel et al. 2010; Sherin und van Es 2008). Insbesondere Studierende, Referendar:innen und Noviz:innen im Beruf fällt das Wahrnehmen, Beschreiben, Erklären relevanter Situationen sowie das Schlussfolgern deutlich schwerer als erfahrenen Lehrpersonen (Seidel und Prenzel 2007, S. 211–213). Gerade deshalb ist eine Förderung dieser Kompetenzen von besonderer Bedeutung. Es zeigt sich, dass die Professionelle Unterrichtswahrnehmung insbesondere durch videobasierte Instrumente gezielt gestärkt werden kann (u.a. Frommelt et al. 2016, S. 365; Gold und Hasselhorn 2013, S. 149; Hellermann et al. 2015, S. 104; Santagata und Guarino 2011, S. 142–143; Sunder et al. 2016, S. 9).

Die vorliegende Arbeit setzt an diesem Punkt an und stellt ein videobasiertes Analysetool vor, mit dem vor allem angehende Lehrpersonen die Möglichkeit haben sollen, Unterrichtssituationen besser beschreiben, erklären und Vorhersagen für den weiteren Lehr-Lernprozess treffen zu können. Dabei soll den Studierenden eine erleichterte Verknüpfung zwischen dem praktischen Lehrerhandeln und den dahinterliegenden Theorien ermöglicht werden.

Einordnung der Dissertationsarbeit

Die Dissertationsarbeit ist im Rahmen des BMBF-geförderten Projektes DiAmant (Diagnose und Adaptation medial analysieren und nachhaltig trainieren) in einem von vier Teilprojekten an der Universität Koblenz-Landau angesiedelt. Im Zuge der Hochschulstrukturreform in Rheinland-Pfalz wurde der Standort Landau von der bisherigen Universitätsstruktur getrennt und ging in die neu gegründete Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (RPTU) über, die aus der Fusion mit der Technischen Universität Kaiserslautern hervorging. Infolge der strukturellen Neuausrichtung ist das Dissertationsprojekt nun an der RPTU verortet.

Ziel des Projektes DiAmant ist die Analyse von Einflussfaktoren auf die Diagnosekompetenz und Fähigkeit der Adaption von Lehrpersonen, um valide Analyse- und Trainingsprogramme entwickeln zu können. Der Fokus liegt auf der Untersuchung diagnostischer Kompetenzen im Hinblick auf domänenspezifischen und fächerübergreifende Aspekte. Mit der Entwicklung des universitären Videotools ViviAn - Videovignetten zur Analyse von Unterrichtsprozessen (<https://vivian-training.de>) steht eine videobasierte Lernumgebung zur Messung und Förderung diagnostischer Kompetenzen zur Verfügung (Bartel und Roth 2017). Im Rahmen der Teilprojekte (Abb. 1) soll ViviAn weiterentwickelt und evaluiert werden.

DiAmant M	DiAmant A	DiAmant B	DiAmant S
Auswirkungen von diagnostischer Kompetenz auf die prozessuale Unterrichtsqualität am Beispiel der Motivierung	Aufmerksamkeitsverhalten während Schülerexperimentierphasen	Mathematische Begriffsbildung und zieladäquate Materialnutzung	Diagnose von Strukturiertheit und kognitiver Aktiviertheit beim naturwissenschaftlichen Problemlösen

Abbildung 1 Teilprojekte von DiAmant - Diagnose und Adaption medial analysieren und nachhaltig trainieren

Das Teilprojekt DiAmant M untersucht die Auswirkungen von diagnostischer Kompetenz auf die prozessuale Unterrichtsqualität am Beispiel der Motivierung. Im Rahmen dieses Projektes ist die vorliegende Arbeit einzuordnen.

Aufbau der Dissertation

Der erste Teil „Theoretische Grundlagen“ enthält neben der *Einleitung* (Kapitel 1) vier relevante Themenkomplexe für die vorliegende Dissertation: Im Kapitel 2 *Motivation und Motivierung im Unterricht* wird auf verschiedene Motivationstheorien sowie die Förderung der Motivation im Lernkontext eingegangen. Das anschließende Kapitel 3 widmet sich der *Professionellen Unterrichtswahrnehmung*. Hier werden die Phasen des *Noticings und Knowledge-based Reasonings* unter Berücksichtigung der aktuellen Forschungslage erläutert. Das entwickelte Videotool soll der Analyse und Förderung der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung im Unterricht dienen und in der ersten Phase der Lehrer:innenbildung zum Einsatz kommen. Dazu wird im Kapitel 4 der *Einsatz von Videos in der Lehre* thematisiert und verschiedene Videostudien zur Förderung der Professionellen Unterrichtswahrnehmung vorgestellt. Das hier entwickelte videobasierte Instrument wird in das universitäre Videotool *ViviAn* der RPTU am Standort Landau integriert. Daher erfolgt eine ausführliche Beschreibung zum Tool *ViviAn*. Die Entwicklung des Videotools erfolgt unter Berücksichtigung der *Gütekriterien der Testentwicklung*, die im Kapitel 5 beschrieben werden. Dabei wird geprüft, welche der genannten Kriterien angewendet werden können.

Der zweite Teil „Eigene empirische Erhebungen“ beginnt mit der Ableitung der Forschungsfragen. In diesem Abschnitt wird schrittweise die Entwicklung und Validierung des Videotools beschrieben. Diese gliedert sich in vier Phasen auf. Zu Beginn erfolgt im Kapitel 7 die Beschreibung des Prozesses zur *Entwicklung der Videovignetten* und deren *Validierung* über die Aktuelle Lernmotivation der Schüler:innen in den einzelnen gefilmten Unterrichtseinheiten. Die Ergänzung der Videosequenzen durch die *Entwicklung der Analyseaufträge* sowie die *Validierung* dieser über Expert:innenurteile schließt sich im Kapitel 8 an. Die Videovignetten mit Zusatzmaterialien und den passenden Analyseaufträgen ergeben zusammen das videobasierte Instrument, welches in der Pilotierungsstudie eingesetzt und zunächst auf *Akzeptanz des Tools* (Kapitel 9) seitens der Studierenden beurteilt wird. In der nächsten Phase der Validierung wird das Videotool zunächst im Kapitel 10 durch eine *Raschskalierung* überprüft. Anschließend erfolgt die *Erfassung der Professionellen Unterrichtswahrnehmung* (Kapitel 11) sowie die *Förderung der Professionellen Unterrichtswahrnehmung* (Kapitel 12) von Studierenden und Lehramtsanwärter:innen als weitere Validierungsschritte.

Im dritten Teil der Arbeit werden die vorliegenden *Ergebnisse zusammengefasst* (Kapitel 13), in der *Diskussion* im Kapitel 14 interpretiert, kritisch reflektiert sowie die Eignung des Tools zur Analyse der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung beurteilt. Darüber hinaus werden die Limitationen der Arbeit und Implikationen für künftige Forschungen vorgestellt.

2 Motivation und Motivierung im Unterricht

Motivation als zentraler Ausgangspunkt für das Verhalten eines Individuums oder einer Gruppe (Krapp und Hascher 2014a, S. 234) leitet sich aus dem lateinischen „movere“ ab und bedeutet „bewegen, antreiben“ (Brünken et al. 2019, S. 186). Daraus ergibt sich, dass die Motivation den Menschen zu einer gewissen Handlung bewegt. Dabei kann es sich um „die aktivierende Ausrichtung des momentanen Lebensvollzuges auf einen positiv bewerteten Zielzustand bzw. auf das Vermeiden eines negativ bewerteten Zustandes“ (Rheinberg und Vollmeyer 2019, S. 17) handeln. Schiefele und Schaffner (2020, S. 164) weisen jedoch daraufhin, dass diese weit verbreitete Definition nicht ausreichend ist, um das Lernen sowie die Leistung ausreichend erklären zu können. Die Lernmotivation ist ein bewusster Prozess, der die Verwirklichung eines Lernzieles intendiert (Krapp und Hascher 2014b, S. 253). Innerhalb dieses Lernprozesses wird durch die Leistungsmotivation das Erreichen einer gewissen Leistung angetrieben (Krapp und Hascher 2014b, S. 253). Zur Erklärung der dahinterstehenden Motivation kann zwischen der intrinsische und extrinsische Motivation differenziert werden (Schiefele und Schaffner 2020, S. 165). Hierbei sind Motivationstheorien wertvoll, um zu verstehen, wie Motivation entwickelt wird und was sie beeinflusst. Einen Überblick zum Forschungsfeld Motivation, insbesondere auch zur Lern- und Leistungsmotivation geben u.a. Brünken et al. (2019, S. 185–204), Krapp und Hascher (2014a, 2014b) oder Schiefele und Schaffner (2015, S. 154–175). In diesem Kapitel werden auszugsweise Theorien und Modelle vorgestellt, die als theoretischer Rahmen für die Konzeption des videobasierten Instruments zur Erfassung der Professionellen Wahrnehmung hinsichtlich der Motivation und Motivierung im Unterricht dienen. Schwerpunktmäßig liegt der Fokus auf der Selbstbestimmungstheorie nach Deci und Ryan (1993) und dem IBAS-Modell von Benning et al. (2019) in den Abschnitten 2.1 *Motivation durch Bedürfnisbefriedigung* und 2.2. *Motivierung durch Zielorientierung im Lernkontext*. Eine besondere Aufgabe steht den Lehrpersonen im Berufsalltag bevor, wenn es um die *Messung der Motivation im Lernkontext* geht (siehe Abschnitt 2.3).

2.1 Motivation durch Bedürfnisbefriedigung

Eine Theorie zur Erklärung der Motivation anhand der Befriedigung von Bedürfnissen haben Deci und Ryan (1993) entwickelt. Die Theorie der Selbstbestimmung (siehe Abbildung 2) wird in der vorliegenden Arbeit aus mehreren Perspektiven beleuchtet und als ein Axiom für das videobasierte Instrument zur Förderung der Professionellen Unterrichtswahrnehmung gesehen. Dementsprechend wird an dieser Stelle die Theorie näher beschrieben und anhand der Perspektiven: Studierende, Lehrende und Lernende beispielhaft skizziert.

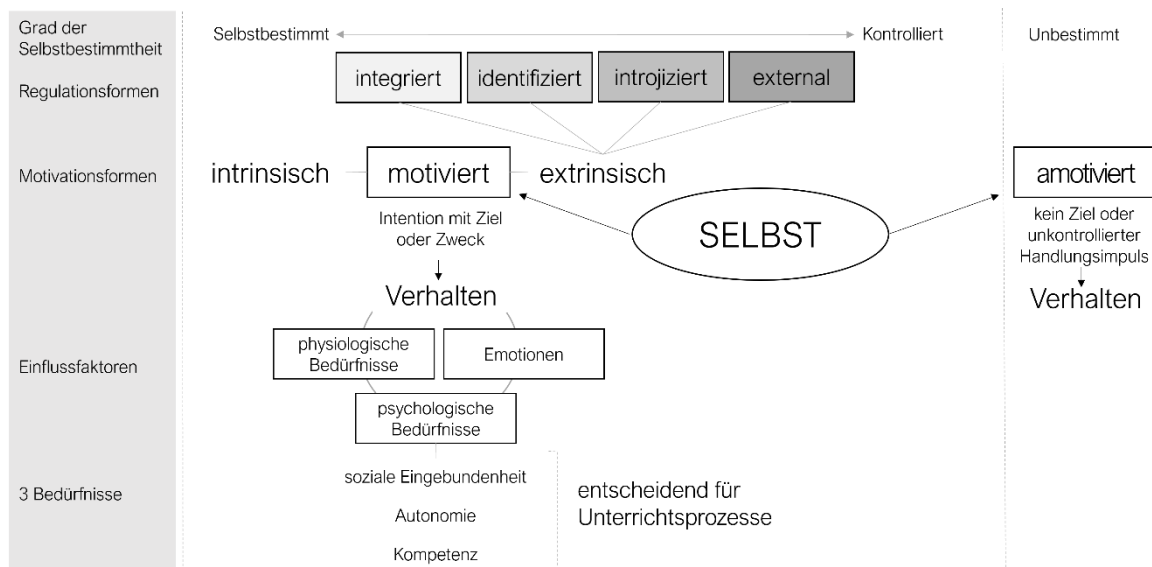


Abbildung 2 Selbstbestimmungstheorie (eigene Darstellung) in Anlehnung an Deci und Ryan (1993)

Die Selbstbestimmungstheorie nach Deci und Ryan (1993) postuliert, dass jedes Individuum nach einem selbstbestimmten, zielorientierten Handeln strebt (andernfalls gilt es als amotiviert) (Deci und Ryan 1993, S. 224). Dieses motivierte Verhalten kann hinsichtlich des Ausmaßes der Selbstbestimmtheit analysiert und als intrinsische oder extrinsische Motivation eingestuft werden (Deci und Ryan 1993, S. 225–228). Dabei ist die intrinsische Motivation der innere Motor, sich stets weiterzuentwickeln, sich Herausforderungen zu stellen, aktiv zu sein und Neues zu entdecken (Ryan und Deci 2000, S. 70). Dem gegenübergestellt wird die extrinsische Motivation, die zur Erreichung eines Ziels das Verhalten von außen beeinflusst (Ryan und Deci 2000, S. 71). Beide Formen der Motivation stehen im Unterricht miteinander im Einklang. Wünschenswert ist ein hoher Anteil intrinsischer Motivation, die ein Lernen für sich selbst, mit Freude und aus Eigeninitiative bedeutet (Deci und Ryan 1993, S. 225). Dies gilt für Schüler:innen in Schule oder Ausbildung zum Beispiel bei der Aneignung von neuen Fähigkeiten, für Studierende bei der selbstbestimmten Auseinandersetzung mit Theorien sowie für Lehrkräfte bei der Reflektion des eigenen Handelns oder bei der Weiterbildung zu relevanten Thematiken. Das Lernen aus intrinsischer Überzeugung ist effektiv und nachhaltig (Gottfried et al. 2001, S. 9–10). Studien, wie in der Metaanalyse von Cerasoli et al. (2014) herausgestellt, belegen einen positiven Zusammenhang zwischen intrinsischer Motivation und Leistung (183 Studien, $N = 212.468$, $r = 0.21$) (Cerasoli et al. 2014, S. 989). Je größer die intrinsische Motivation, desto besser ist die Schulleistung. Sowohl die intrinsische als auch die extrinsische Motivation können ein hohes Maß an selbstbestimmtem Handeln hervorrufen (Deci und Ryan 1993, S. 226–227). Dabei spielt die graduelle Klassifikation der extrinsischen Motivation entsprechend der Regulationsformen integriert,

identifiziert, introjiziert und external - mit abnehmendem Grad der Selbstbestimmtheit - eine Rolle (Deci und Ryan 1993, S. 226–227). In der Abbildung 3 *Regulationsformen* werden die einzelnen Formen der extrinsischen Motivation kurz beschrieben.

integriert	Das Verhalten erfolgt bei der integrierten Regulation durch die Übernahme der erwarteten Ziele und Werte in das Selbstkonzept. Eine Person handelt aus innerer Überzeugung. Damit wird ein hohes Maß an Selbstbestimmtheit erreicht.
identifiziert	Eine Person identifiziert sich mit den Zielen oder dem Zweck einer Erwartung und eignet sich diesen Aspekt an. Das Handeln erfolgt auf Grund der wahrgenommenen Bedeutsamkeit für sich selbst.
introjiziert	Der Wunsch bestimmten Verhaltensweisen zu folgen wird bei der introjizierten Regulation durch erwartete Handlungsmuster oder Werte von außen angetrieben.
external	Das Verhalten erfolgt auf Basis von Reaktionen oder Verstärkungen von außen. Dabei kann es sich um positive Verstärkungen in Form von Belohnungen oder eine negative Verstärkung durch den Entzug von Bestrafungen handeln.

Abbildung 3 *Regulationsformen* (eigene Darstellung in Anlehnung an (Deci und Ryan 1993, S. 227–228))

Eine externale Regulation kann beispielsweise bei einer Schülerin durch eine monetäre Belohnung der Eltern bei einer guten Note dazu führen, mehr Zeit in die Ausarbeitung eines Referatsthemas zu investieren. Es steigt die Motivation auf Grund der positiven Verstärkung von außen. Zielführend für ein erfolgreiches Lernen sollte jedoch die Überführung in eine andere Regulationsstufe sein. In diesem Fall kann durch eine Betonung der Relevanz des Referatsthemas eine identifizierte Regulation erreicht werden. Damit steigt das Empfinden, nicht nur für Geld etwas zu tun, sondern auch für sich selbst. Es wird deutlich, dass je nach Motivations- und Regulationsform ein unterschiedliches Ausmaß an Internalisierung erreicht wird (Ryan und Deci 2000, S. 71). Mit zunehmendem Grad der Selbstbestimmtheit kann das Lernen erwünschter Werte, Perspektiven, Kenntnisse oder Handlungen gefördert werden (Ryan und Deci 2000, S. 71–73). Das bedeutet für den Lernkontext, dass Lehrpersonen ihre Schüler:innen und Dozierende ihre Studierenden weniger external, sondern möglichst selbstbestimmt motivieren sollten. Wie Cerasoli et al. (2014) können Richardson et al. (2012) in ihrer Metaanalyse ($N = 5326$) für den Kontext Studium einen positiven Zusammenhang zwischen der Leistung (Benotung) und der intrinsischen Motivation ($r = 0.17$) im Gegensatz zur extrinsischen Motivation ($r = 0.01$) aufzeigen (Richardson et al. 2012, S. 368). Auch in diesem Fall bedeutet dies, je höher der Grad der Selbstbestimmtheit, desto besser werden die Noten. Dementsprechend sollte auch im Studium die intrinsische Motivation gefördert werden. Dies kann beispielsweise durch die Verwendung von Videos mit einem erhöhten Anwendungsbezug gelingen (Syring et al. 2015, S. 682; Ertl-Schmuck 2019, S. 48–49).

Vorteile des Einsatzes von Videos sowie Faktoren, die berücksichtigt werden sollten, sind im Kapitel 4 *Einsatz von Videos in der Lehre* beschrieben.

Grundsätzlich stellt sich die Frage, wie intrinsische Motivation erreicht werden kann bzw. welche Faktoren bei der Motivation eine Rolle spielen. In der Selbstbestimmungstheorie werden dazu drei Einflussfaktoren beschrieben. Ein motiviertes Handeln wird durch physiologische und psychologische Bedürfnisse sowie durch Emotionen hervorgerufen (Deci und Ryan 1993, S. 229). Sobald physiologische Grundbedürfnisse nach Nahrung, Schlaf, Toilettengängen oder auch Sicherheit nicht erfüllt sind, hat dies demnach enorme Auswirkungen auf die Motivation und das Verhalten. Daher müssen alle beteiligten Personen am Lernprozess hinsichtlich dieser Bedürfnisse befriedigt sein, um ein erfolgreiches Lernen zu ermöglichen. Unausgeglichene Emotionalität zum Beispiel durch Wut, Trauer, Angst werde sich ebenfalls auf die Motivation auswirken und das Lernen beeinflussen. Die Erfüllung der psychologischen Bedürfnisse nach Autonomie, Kompetenz und soziale Eingebundenheit kommt laut Deci & Ryan Deci und Ryan (1993, S. 229) einer besonderen Bedeutung bei der Motivation und Zielsetzung zu. Der Mensch strebt von Natur aus nach sozialer Zugehörigkeit, einem sozialen Umfeld, in dem man miteinander agiert, seine Fähigkeiten und Fertigkeiten einbringen möchte, aber sich dennoch autonom fühlen kann (Deci und Ryan 1993, S. 229). In zahlreichen Studien konnten die Postulate der Selbstbestimmungstheorien hinsichtlich der Bedeutsamkeit der Bedürfnisse für den Unterrichtsalltag belegt werden. Beispielhaft werden an dieser Stelle drei Studien beschrieben, an die im Abschnitt 2.5 *Zusammenfassung in Bezug auf die vorliegende Erhebung* angeknüpft wird.

Bieg und Mittag (2011) haben die Auswirkung der Förderung der selbstbestimmten Motivation auf die Leistung untersucht. In ihrer Intervention im Fach Deutsch in der Sekundarstufe kann durch eine Weiterbildung der Lehrpersonen zu motivationsfördernden Maßnahmen, Konzepten und Verhaltensweisen im Unterricht sowie durch speziell konzipiertes Lern- und Unterrichtsmaterial für die Schüler:innen auf Basis der drei psychologischen Bedürfnisse nach Deci & Ryan (Bieg und Mittag 2011, S. 225–226) ein positiver Einfluss auf die Leistung der Schüler:innen festgestellt werden (Bieg und Mittag 2011, S. 231).

Eine andere Facette des Schulalltags betrachten Kaiser et al. (2021), die das Thema Druck und Kontrolle im Zusammenhang mit der Befriedigung der drei psychologischen Grundbedürfnisse nach Deci & Ryan analysieren. Die Erhebung wurde mit Lernenden der Sekundarstufe II aus Nordrhein-Westfalen ($N = 715$) durchgeführt (Kaiser et al. 2021, S. 1414). Die Ergebnisse werden getrennt nach den drei faktorenanalytisch überprüften Dimensionen des Druckerlebens (Kaiser et al. 2021, S. 1411) zusammengefasst (Kaiser et al. 2021, S. 1419–1420):

- 1) Der *Druck durch zeitliche Anforderungen* (z.B. Umfang der Hausaufgaben, Vor- und Nachbereitungen, Arbeitsaufwand im Unterricht) wirkt sich negativ auf das Autonomieerleben und die soziale Eingebundenheit aus. Das Kompetenzerleben wird durch den Zeitdruck nicht signifikant beeinflusst.
- 2) Das Kompetenzerleben sinkt stattdessen durch den *Druck durch Leistungsanforderungen*. Eine Auswirkung dieses Druckerlebens auf die soziale Eingebundenheit und die Autonomie können nicht nachgewiesen werden.
- 3) Der *Druck durch die Lehrperson* wirkt sich insbesondere negativ auf das Autonomiegefühl aus. Zusätzlich kann das Kompetenzerleben durch den Einsatz von Druck durch die Lehrperson gehemmt werden. Die soziale Eingebundenheit wird jedoch nicht beeinflusst.

Damit können Kaiser et al. (2021) aufzeigen, dass sich Druck im Unterricht auf die physiologischen Grundbedürfnisse negativ auswirken kann.

Ein besonderes Augenmerk in der vorliegenden Arbeit wird auf die Studie von Luo et al. (2021) gelegt, die die Motivation sowie die Grundbedürfnisse von College-Studierenden ($N = 227$) im online-basierten Lernsetting untersucht. Es zeigt sich (siehe Abbildung 4 *Ergebnisse der PLS-SEM*), dass

- 1) es einen positiven Zusammenhang zwischen den drei Grundbedürfnissen wahrgenommene Autonomie (PA), Kompetenz (PC), soziale Eingebundenheit (PR) und der intrinsischen Motivation (wahrgenommene Freude - PE) im online Lernsetting gibt,
- 2) mit Befriedigung der drei Grundbedürfnisse die Studierenden eine höhere Akzeptanz (wahrgenommene Nützlichkeit - PU) des onlinebasierten Lernsettings aufweisen und die Motivation (PE und PU) steigt, die Aufgaben zu erledigen (Lernsetting weiterhin nutzen – CI),
- 3) die wahrgenommene Kompetenz (PC) sowohl die intrinsische (PE) als auch die extrinsische Motivation (PU) am stärksten beeinflusst (Luo et al. 2021, S. 1391–1392).

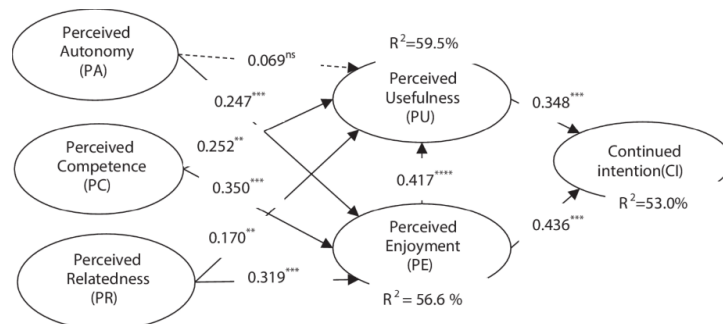


Abbildung 4 Ergebnisse der PLS-SEM (Luo et al. 2021, S. 1390)

Ein online-basiertes Instrument eignet sich demnach, um die Motivation, das Interesse, die Bereitschaft zum Lernen und die Leistung zu erhöhen (Luo et al. 2021, S. 1392–1393). Das gelingt, sofern das Instrument die Autonomie, soziale Eingebundenheit und Kompetenz fördert und es interessant, benutzerfreundlich und relevant aufgebaut ist (Luo et al. 2021, S. 1393).

Demnach ist die Befriedigung dieser Bedürfnisse entscheidend für Unterrichtsprozesse und sie sollten im Lernsetting unbedingt berücksichtigt werden. Ein hilfreiches Modell für Lehrpersonen zur Umsetzung motivationsförderlicher Maßnahmen ist das IBAS-Modell nach Benning et al. (2019) und wird im Abschnitt 2.3 *Motivationsförderung im Lernkontext* näher beschrieben.

2.2 Motivierung durch Zielorientierung im Lernkontext

Entsprechend der Selbstbestimmungstheorie nach Decy & Ryan (siehe Abschnitt 2.1.) wird deutlich, dass jedem motivational geleiteten Handeln ein Ziel zu Grunde liegt. Im Lernkontext wird in der Literatur klassisch zwischen Lernziel- und Leistungszielorientierungen unterschieden (Heckhausen und Heckhausen 2018, S. 430; Wild und Möller 2009, S. 275). Leistungszielorientierungen kennzeichnen sich durch den Wunsch, dass die eigenen Kompetenzen positiv bewertet werden (Heckhausen und Heckhausen 2018, S. 578). Dabei wird es vermieden, Schwächen zu zeigen. Die Beschäftigung mit Aufgaben dient dazu, die vorhandenen Kompetenzen zum Ausdruck zu bringen und zu bestätigen (Heckhausen und Heckhausen 2018, S. 578) (Heckhausen 2018, S. 578). Die Leistungsziele können differenziert nach Annäherungs- und Vermeidungszielen betrachtet werden (Schiefele und Schaffner 2011, S. 21). Dabei handelt es sich um das Ziel eines direkten Kompetenzvergleichs mit anderen Personen (Annäherungsleistungsziel) oder um das Verbergen der eigenen Schwächen (Vermeidungsleistungsziel) (Schiefele und Schaffner 2015, S. 162).

Im Vergleich zu den Leistungszielen mit der Demonstration bzw. dem Verbergen der eigenen Fähigkeiten, dienen Lernziele dazu, sich intensiv mit neuen Herausforderungen zu beschäftigen und einen Lern- und Kompetenzzuwachs zu erreichen (Heckhausen und Heckhausen 2018, S. 578). Die Lernzielorientierungen lassen sich entsprechend ihrer Bedeutung als intrinsische und extrinsische Lernmotivation einordnen (Heckhausen und Heckhausen 2018, S. 430): Bei einer intrinsischen Lernzielorientierung erfolgt der Wunsch, etwas zu lernen von innen heraus, während bei der extrinsischen Lernzielorientierung die Motivation des Lernens darin begründet ist, anderen Menschen zu zeigen bzw. zu vergleichen, wie kompetent man ist. In Abgrenzung zur Leistungszielorientierung ist das Bestreben bei extrinsisch motivierten Lernzielen das wirkliche Erreichen eines Kompetenzzuwachses. Die Motivation zu lernen erfolgt jedoch durch äußere Reize zum Beispiel in Form der Anerkennung. Wie bereits dargestellt, ist für den Erfolg des Lernens in der Schule insbesondere die Erreichung

einer intrinsischen Motivation von Bedeutung. Es stellt sich die Frage, wie die Lehrpersonen im Unterrichtsalltag dies erreichen können. Daher wird in diesem Abschnitt insbesondere auf zwei Modelle zur Förderung einer Lernzielkultur im Unterricht eingegangen: dem Target- und dem IBAS-Modell.

2.2.1 Vom TARGET- zum IBAS-Modell

Zur Förderung der Motivation und Lernentwicklung der Schüler:innen hat Epstein (1988) sechs relevante Strukturen herausgestellt: „task, authority, reward, grouping, evaluation, and time (TARGET) structures“ (Epstein 1988, S. 92–93). Diese TARGET-Strukturen bilden eine Möglichkeit, Anpassungen im Lernsetting vorzunehmen, um bessere Lernbedingungen zu realisieren. Ames (1992a, 1992b) betont in ihren Arbeiten die Bedeutsamkeit der Leistungszielorientierung im Unterricht und analysiert die TARGET-Strukturen hinsichtlich dieser zur Förderung der Motivation. In der Literatur werden die Ergebnisse dieser Forschungsarbeit häufig als TARGET-Modell beschrieben. Ausgangspunkt in der vorliegenden Arbeit stellt die Beschreibung der TARGET-Strukturen nach Epstein (1988) dar, die durch die wissenschaftliche Arbeit von Ames (1992a, 1992b) ergänzt werden. Die Arbeitsgruppe um Benning (2019) hat auf Grundlage der Selbstbestimmungstheorie und des TARGET-Modells nach Ames das IBAS-Modell konzipiert. Ziel ist es, mit dem IBAS-Modell motivationsfördernde Aspekte für unterrichtliches Handeln vereinfacht und fokussiert aufzeigen zu können (Benning et al. 2019, S. 528–529). Dabei lenken sie ihr Forschungsinteresse vor allem auf die Umsetzung einer Lernzielstruktur im Unterricht, da diese das Lernen und die Motivation der Schüler:innen positiv beeinflusst (Benning et al. 2019, S. 523–524).

Im IBAS-Modell werden vier relevante Dimensionen für die Umsetzung einer Lernzielkultur beschrieben: Inhalts-, Bewertungs-, Autonomiedimension und Soziale Dimension (Benning et al. 2019, S. 528). Die IBAS-Dimensionen entsprechen den TARGET-Strukturen wie in Abbildung 5 aufgezeigt.

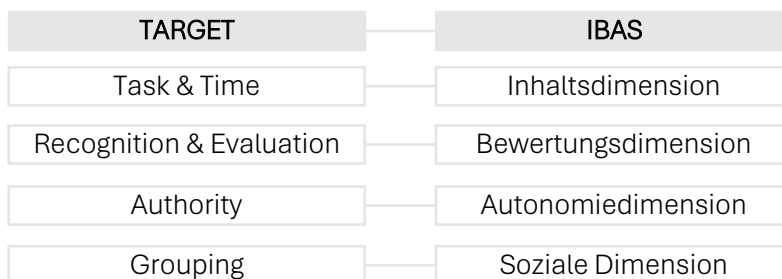


Abbildung 5 TARGET und IBAS – Strukturen (Benning et al. 2019, S. 528)

Insgesamt werden den vier IBAS-Dimensionen 13 unterrichtliche Handlungen zugeordnet (siehe Abbildung 6 IBAS-Modell), welche auf Grundlage theoretischer Annahmen identifiziert und begründet worden sind (Benning et al. 2019, S. 529–530). Diese Unterrichtshandlungen bieten Lehrpersonen und Lehramtsstudierenden eine gute

Möglichkeit, ihren eigenen Unterricht zu analysieren, zu planen und anzupassen, um die Motivation zu erhöhen und die physiologischen Bedürfnisse nach Decy & Ryan zu befriedigen. Konkrete Vorschläge zur Förderung einzelner Subdimensionen lassen sich bei Schiefele (2014, S. 255–258) finden. Die im IBAS-Modell dargestellten motivationsförderlichen Aspekte im Unterricht können schwer separat betrachtet werden, denn sie interagieren. Zusammenhänge zwischen einzelnen Facetten stellt Rakoczy (2006) in ihrer Arbeit dar. Unter anderem konnte sie belegen, dass die Gewährung von Freiräumen für eigene Entscheidungen (Wahlmöglichkeiten), wertschätzende Beziehungen und konstruktive Rückmeldungen als Unterstützung der Autonomie wahrgenommen werden (Rakoczy 2006, S. 839). Ebenso motivationsförderlich wirkt sich die Kognitive Aktivierung aus, allerdings nur im Zusammenhang mit Maßnahmen der Individualisierung (Rakoczy 2006, S. 837). Das IBAS-Modell stellt eine wesentliche Komponente in der Entwicklung der vorliegenden Videovignetten dar (siehe Abschnitt 4.2).

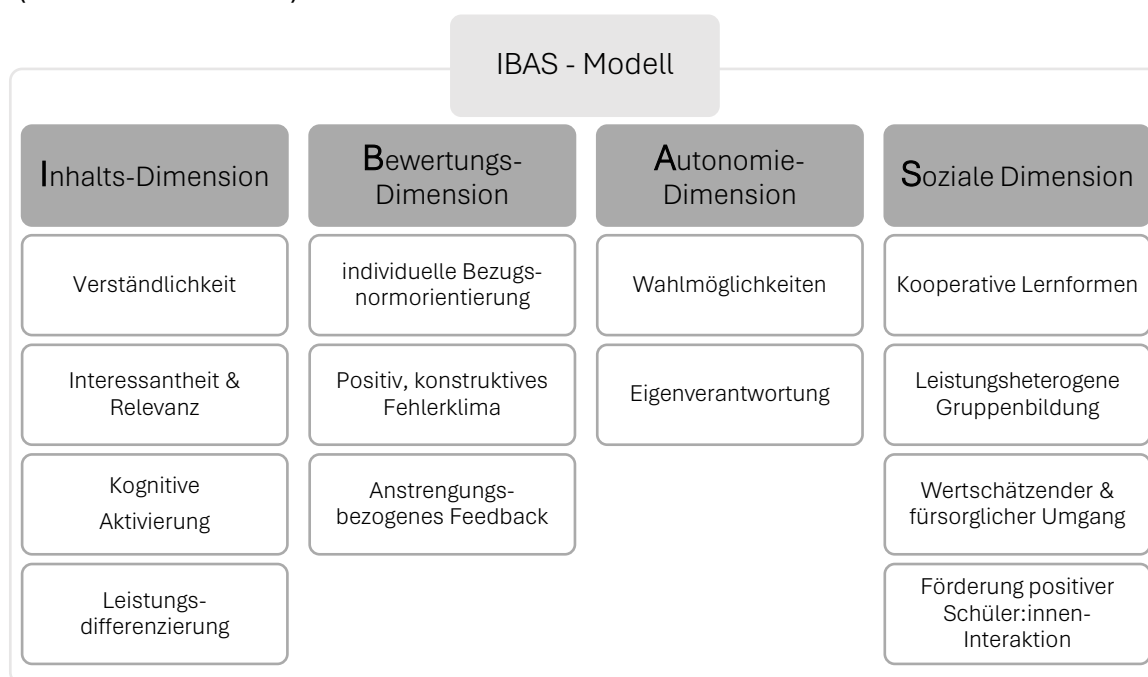


Abbildung 6 IBAS-Modell (Vgl. Benning, Praetorius, Janke, Dickhäuser & Dresel, 2019, S. 529)

2.2.2 Vier Dimensionen zur Förderung der Motivation

Eine inhaltliche Gegenüberstellung der Dimensionen des TARGET- und IBAS-Modells soll zum einen die Grundannahmen, zum anderen die Gemeinsamkeiten und Unterschiede aufdecken (Abbildung 5 TARGET und IBAS – Strukturen).

Inhaltsdimension

„Task and Time“ im TARGET-Modell bezieht sich auf die Aufgabenstruktur, ausgehend vom übergeordneten Zeitrahmen eines Lehrplans bis zu einer einzelnen Unterrichtssequenz (Epstein 1988, S. 93). Dabei sollen die Aufgabenstrukturen

entsprechend der einzelnen Schüler:innen divers gestaltet werden (Epstein 1988, S. 93): Möglichkeiten der Adaption liegen in der Verwendung verschiedener Materialien und Schwierigkeitsgrade, unterschiedlicher Themen mit entsprechender Relevanz, in der Struktur der Aufgaben von logisch aufbauend bis spiralförmig, im Umfang der Aufgaben (Epstein 1988, S. 93) sowie in der Zeitvorgabe für die Erledigung der Aufgaben (Epstein 1988, S. 97–98). Zusätzlich kann zwischen verschiedenen Lehr-Lernmethoden wie z.B. Lehr- und Lernendenvortrag, Projekt- und Gruppenarbeit, computer-gestütztes Lernen oder Einzelarbeit gewählt werden (Epstein 1988, S. 93). Im IBAS-Modell werden der Inhaltsdimension vier Subdimensionen zugeordnet (siehe Abbildung 2): Verständlichkeit, Interessantheit & Relevanz, Kognitive Aktivierung und Leistungsdifferenzierung (Benning et al. 2019, S. 529). Anhand dieser Subdimensionen werden die Überschneidungen mit dem TARGET-Modell deutlich. Der Unterricht sollte interessant und relevant sein, aber auch die Aufgaben entsprechend dem Niveau der Schüler:innen gestaltet werden (Benning et al. 2019, S. 529). Das Interesse an einem Lerngegenstand geht mit positiven Emotionen einher und wirkt sich deshalb förderlich auf die intrinsische Motivation aus (Schiefele und Schaffner 2015, S. 158; Wild und Möller 2015, S. 2015). Im engen Zusammenhang damit steht die wahrgenommene Relevanz der Lerninhalte. Authentische, relevante und reale Probleme tragen zur Lernmotivation bei und befähigen die Lernenden dazu, sich intensiv einem Thema zu widmen (Reinmann und Mandl 2006, S. 640). Die Relevanz betrifft nicht nur das Wecken des Interesses am Thema durch einen Alltagsbezug, sondern auch die Vermittlung der Bedeutsamkeit für die individuellen Zielsetzungen (Assor et al. 2002, S. 262). Dadurch soll den Schüler:innen bewusst werden, dass sie für sich selbst lernen, die Eigenverantwortung für ihr Handeln tragen und auf dieser Grundlage befähigt werden, eigene Entscheidungen zu treffen (Assor et al. 2002, S. 262). Jede:r Schüler:in bringt nicht nur eigene Erfahrungen (Paradies und Linser 2009, S. 41), sondern auch eigene Voraussetzungen für das Lernen mit (Paradies und Linser 2009, S. 34). Daher ist es notwendig, im Unterricht zu differenzieren (Paradies und Linser 2009, S. 34), wie es Epstein (1988, S. 93) beispielhaft beschreibt (siehe oben). Im IBAS-Modell wird ein Fokus auf die kognitive Aktivierung gelegt (Benning et al. 2019, S. 529). Dadurch sollen sich die Schüler:innen aktiv, konzentriert und vertieft mit einem Lerninhalt befassen (Heymann 2015, S. 7). Vorteile einer kognitiven Aktivierung sind die erleichterte Verknüpfung mit dem bisherigen Wissen, das Verstehen von Zusammenhängen, die Förderung des Wissenstransfers auf andere Aufgabenbereiche sowie das Verstehen des Inhalts überhaupt (Heymann 2015, S. 7). Die Auseinandersetzung mit herausfordernden Aufgaben ist bedeutsam für die Förderung der Kompetenzen (Benning et al. 2019, S. 529). Dabei spielt die Differenzierung hinsichtlich der Leistung, Lernstile, Lerntempi eine Rolle (Paradies und Linser 2009, S. 36–37). Im IBAS-Modell wird zusätzlich die Verständlichkeit als eine Subdimension aufgegriffen. Die Verständlichkeit kommt der „Inhaltlichen Klarheit“ als Merkmal guten

Unterrichts nach Meyer (2020, S. 55) nahe. Dabei geht es um eine Lernförderung durch die Formulierung verständlicher Aufgaben und verständlicher Abläufe (Meyer 2020, S. 55–57). Die nachvollziehbare Bewertung (Meyer 2020, S. 58–59) wird eher der Bewertungsdimension zugeordnet.

Bewertungs-Dimension

Die Bewertungsstruktur im IBAS-Modell wird aufgegliedert in die individuelle Bezugsnormorientierung, dem positiv, konstruktiven Fehlerklima und dem anstrengungsbezogenen Feedback (Benning et al. 2019, S. 529). Wenn die Bewertung entsprechend der individuellen Leistungen erfolgt, kann die Lern- und Leistungsmotivation der Schüler:innen gesteigert werden (Rheinberg und Vollmeyer 2019, S. 99). Der Umgang mit Fehlern wirkt sich positiv auf den Lernzuwachs aus, sofern Fehler als zulässig gelten und reflektiert werden (Benning et al. 2019, S. 529–530). Benning et al. (2019, S. 529) verweisen bei der Bedeutsamkeit von anstrengungsbezogenen Rückmeldungen auf Ames (1992a), dass diese den individuellen Bemühungen und Lernerfolgen entsprechen. Hinsichtlich der Anerkennungs- und Bewertungsstruktur „Recognition & Evaluation“ wird im Target-Modell von Epstein (1988, S. 95–97) deutlich gemacht, dass es keine festgelegten Kriterien für Lob oder Anerkennung im Gegensatz zur Bewertung gibt. Für eine Bewertung wird klar festgelegt, wofür, wie oft und in welcher Weise diese erfolgt (Epstein 1988, S. 95). Belohnungen haben stets Auswirkungen auf die Schüler:innen. Sie können zur Steigerung oder Minderung der Lernmotivation führen (Epstein 1988, S. 95). Individuelle Belohnungen oder Bewertungen sollten immer in Abhängigkeit vom Ausgangspunkt, den Zielen und den Ergebnissen erfolgen, nur dann ist der Lernverlauf beurteilbar (Epstein 1988, S. 95–97). Hier wird vor die individuelle Bezugsnorm (siehe IBAS-Modell) deutlich. Rückmeldungen sind motivations- und lernförderlich, wenn sie auf realistischen, transparenten und gerechten Kriterien basieren und die individuelle Entwicklung entsprechend der individuellen Ziele aufzeigen können (Epstein 1988, S. 97).

Autonomie-Dimension

Epstein (1988, S. 94) zeigt ein weiter gefasstes Verständnis für „Authority“ in der Schule auf. Mitbestimmung muss sich nicht nur auf die Schüler:innen einer Klasse beziehen, wenn sie bei den Abläufen im Unterricht, den Aufgaben, den Regeln zum Sozialverhalten, den Evaluationen oder den Bewertungen mitwirken dürfen (Epstein 1988, S. 94). Darüber hinaus können Eltern, Firmen, Vereine, weiteres Schulpersonal bei Entscheidungsfindungen und Arbeitsprozessen integriert werden (Epstein 1988, S. 94). Ebenso sollten aktuelle Forschungsarbeiten und Literatur einbezogen werden, um einen bestmöglichen Lernerfolg garantieren zu können (Epstein 1988, S. 94).

Dies bedeutet, dass nicht nur die Lehrperson die Autorität innehat und über alles selbst entscheiden sollte. Vielmehr handelt es sich um eine geteilte Verantwortung mehrerer Parteien. Zudem profitiert unser Bildungssystem von evidenzbasierter Lehre. Eine erhöhte Selbst- oder Mitbestimmung der Schüler:innen wirkt sich positiv auf deren Motivation, deren Selbstwert, die aktive Teilnahme und die Effektivität des Lernprozesses aus (Epstein 1988, S. 94). Im IBAS-Modell wird die Autonomie-Dimension mit den Bereichen der Wahlmöglichkeiten und Eigenverantwortung zur Unterstützung einer Lernzielkultur näher erläutert (Benning et al. 2019, S. 530). Dabei geht es um die Schaffung adäquater Freiräume im Unterricht durch Beteiligung der Schüler:innen an Entscheidungsprozessen und die Übernahme der Verantwortung für den eigenen Lernerfolg (Benning et al. 2019, S. 530). Im Gegensatz dazu gibt es Handlungsweisen, die kontrollierend wirken und die Lernenden extrinsisch motivieren. Dazu zählen ein direkter Sprachstil, eine Zeiterfassung bei Aufgaben, das Aufzeigen von Lösungen, statt sie zu besprechen, sowie das Formulieren von Kontrollfragen (Reeve und Jang 2006, S. 216). Lässt man eine hohe Autonomie zu, handeln die Lernenden mit einem hohen Selbstbestimmungsgrad und von innen heraus und es entsteht weniger das Gefühl, etwas auferlegt zu bekommen (Reeve und Jang 2006, S. 216). In ihrer Untersuchung haben Reeve und Jang (2006, S. 215–216) autonomieunterstützende Verhaltensweisen im Lehralltag identifiziert: selbstgesteuerte und autonome Arbeitsphasen, Wertschätzen mit informativen Rückmeldungen, Ermutigen und Hilfestellung geben. Auch Rakoczy (2006, S. 839) bestätigt mit ihrer Arbeit die Förderung der Autonomiewahrnehmung durch das Schaffen von Entscheidungsfreiräumen und das Geben eines positiv-konstruktiven Feedbacks. Ebenfalls tragen soziale Komponenten mit einem unterstützenden Lernklima, wie z.B. einem wertschätzenden Umgang und Klassenführungsaspekte wie Störungsfreiheit und klare Regeln zu einem Autonomiegefühl bei (Rakoczy 2006, S. 835–836).

Soziale Dimension

Die soziale Dimension im IBAS-Modell umfasst neben dem wertschätzenden und fürsorglichen Umgang mit den Schüler:innen auch den Einsatz kooperativer Lernformen, eine leistungsheterogene Gruppenbildung sowie die Förderung einer positiven Interaktion zwischen den Lernenden (Benning et al. 2019, S. 530). Diese sozialen Aspekte sind für den Aufbau einer Lernzielkultur wichtig (Benning et al. 2019, S. 530). Epstein (1988, S. 95–96) fokussiert sich beim „grouping“ vor allem auf die Aspekte der Gruppenzugehörigkeit und Gruppenzusammensetzung, die durch die Lehrperson beeinflusst werden kann und Auswirkungen auf das Lernen hat. Eine Differenzierung nach leistungsstarken und leistungsschwächeren Schüler:innen ermöglicht eine bedarfsgerechte Förderung (Epstein 1988, S. 96). Dabei betont Epstein (1988, S. 96) die Notwendigkeit eines flexiblen Umgangs mit der Gruppenzuweisung, um die

Motivation für das Lernen aufrecht halten zu können. Eine solche leistungshomogene Gruppenzusammensetzung wird in der Literatur nicht nur positiv, sondern auch kritisch betrachtet. Laut Bandura (1997, S. 175) wirkt sich diese vor allem für die leistungsschwächeren oder nicht gut vorbereiteten Lernenden als motivationshemmend aus und erzeugt bei den Lernenden eine Negativspirale der Selbstwirksamkeit und des Lernerfolgs. Leistungsstärkere Schüler:innen wiederum können sowohl in homogenen Gruppen als auch beim individuell geförderten Lernen hinsichtlich des Lernzuwachses profitieren und ihr Kompetenzbedürfnis erfüllen (Vock und Gronostaj 2017, S. 69). Allerdings besteht die Gefahr, auf Dauer an die individuelle Belastungsgrenze zu kommen sowie in den sozialen Beziehungen innerhalb des Klassenverbands Einbußen wahrnehmen zu müssen (Vock und Gronostaj 2017, S. 69–70). Im IBAS-Modell werden leistungsheterogene Gruppenzusammensetzungen und kooperative Lernformen für den Aufbau einer Lernzielkultur präferiert. Kooperative Lernformen setzen eine kollegiale und verantwortungsbewusste Teamarbeit voraus, um einen größtmöglichen Lernerfolg für alle Gruppenmitglieder zu gewährleisten (Florio-Hansen 2014, S. 126–128). Diese Teamarbeit fördert die Befriedigung des Bedürfnisses nach sozialer Eingebundenheit. Gleichzeitig ist sie Voraussetzung für den Lernerfolg (Hattie 2018, S. 151). In seiner Metaanalyse konnte Hattie (2018, S. 251–254) herausstellen, dass kooperative Lernformen effektiver sind als andere Lernformen wie beispielsweise heterogenes Lernen. Peer-Beziehungen, aber vor allem die Beziehung zur Lehrkraft stellen wichtige Indikatoren für ein positives Verhältnis zur Schule dar (Markus und Schwab 2021, S. 361) und sind für den Lernerfolg entscheidend (Wettstein und Raufelder 2021, S. 17). Lehrkräfte tragen mit ihrer Klassenführung und dem Aufbau einer angenehmen Lernatmosphäre dazu bei, das Bedürfnis nach sozialer Eingebundenheit befriedigen zu können (Markus und Schwab 2021, S. 361).

2.3 Messung der Motivation im Lernkontext

Die besondere Herausforderung liegt für Lehrkräfte darin, nicht nur Theorien zur Motivation und Maßnahmen zur Förderung einer Lernzielkultur zu kennen, sondern auch die Motivation der Lernenden zu erfassen. Motivation ist ein innerer Prozess, der von außenstehenden Personen nicht direkt beobachtet werden kann (Florio-Hansen 2014, S. 57; Krapp und Hascher 2014a, S. 234). Lehrkräfte haben die Möglichkeit, Rückschlüsse vom Verhalten eines Lernenden auf die Motivation zu ziehen. Die innere Motivation der Lernenden bleibt dennoch im Verborgenen. Zudem ist die Motivation ein komplexes Konstrukt (Rheinberg und Vollmeyer 2019, S. 261; Krapp und Hascher 2014a, S. 234), das erst durch „kognitive Prozesse wie Zielbildungen, Erwartungen, Ursachenerklärungen, Wertorientierungen, Selbstbilder, Selbstüberwachung etc., aber auch anreizverleihende Affekte samt ihrer neuro-hormonellen

Grundlagen“ (Rheinberg und Vollmeyer 2019, S. 261) zum Verhalten führt. In der Konsequenz müssten alle Facetten einzeln diagnostiziert und final zusammengefasst werden (Rheinberg und Vollmeyer 2019, S. 262). Eine Möglichkeit, die Motivation zu erfassen, ist die Selbsteinschätzung (Krapp und Hascher 2014a, S. 234). Hierzu lassen sich in der Literatur verschiedene Fragebögen finden. Ein Auszug einiger Instrumente aus Forschungsarbeiten zum Thema Motivation sind in der folgenden Tabelle 1 dargestellt:

Tabelle 1 Instrumente zur Erfassung der Motivation im Lehr-Lernkontext

Testgliederung	Beispielitems
SELLMO - Skalen zur Erfassung der Lern- und Leistungsmotivation, Spinath et al. (2012)	
31 Items, 4 Zielarten:	In der Schule/im Studium geht es mir darum...
<ul style="list-style-type: none"> - Lernziele - Annäherungs-Leistungsziele - Vermeidungs-Leistungsziele - Arbeitsvermeidung 	<ul style="list-style-type: none"> - neue Ideen zu bekommen. (LZ) - zu zeigen, dass ich bei einer Sache gut bin. (ALZ) - dass andere Schüler/Studierende nicht denken, ich sei dumm. (VLZ) - keine schwierigen Tests o. Arbeiten zu haben. (AV)
Aktuelle Motivation, Rakoczy et al. (2005)	
15 Items, 5 Faktoren:	In diesen Mathematikstunden...
<ul style="list-style-type: none"> - amotiviert - external - introjiziert - identifiziert - intrinsisch 	<ul style="list-style-type: none"> - machte das Lernen/Arbeiten Spaß. (INTR) - habe ich mitgearbeitet, weil ich die Sache später brauchen kann. (ID) - habe ich nur das getan, wozu mich der Lehrer aufgefordert hat. (EX) - war mir alles egal. (AM) - war ich aufmerksam, weil ich immer aufpasse. (IJ)
Motivationsunterstützung im Mathematikunterricht, Rakoczy (2008)	
11 Items, 3 Faktoren:	Im Mathematikunterricht...
<ul style="list-style-type: none"> - Autonomie - Kompetenz - Soziale Eingebundenheit 	<ul style="list-style-type: none"> - kann ich selber entscheiden, wie ich arbeiten will. (A) - werde ich für gute Leistungen gelobt. (K) - habe ich das Gefühl dazuzugehören. (SE)
4 Items:	In diesen Mathematikstunden...
<ul style="list-style-type: none"> - Intrinsische Motivation 	<ul style="list-style-type: none"> - verging die Zeit wie im Flug. (INTR)
2 Items, 2 Faktoren:	Individuelle Lernvoraussetzungen
<ul style="list-style-type: none"> - Interesse - Kompetenzüberzeugung 	<ul style="list-style-type: none"> - Mathe ist spannend. (IT) - Ich bin unbegabt für Mathe. (KÜ)

Multiple Stait- und Trait-Ziele im Kontext Unterricht, Bürger (2013)

12 Items, 4 Faktoren:	In der Schule verfolge ich im Allgemeinen das Ziel...
<ul style="list-style-type: none"> - Lernziele (MAS) - Annäherungs-Leistungsziel (PAP) - Arbeitsvermeidungs-Ziel (WOA) - Affiliations-Ziel (AFL) 	<ul style="list-style-type: none"> - viele neue Fähigkeiten zu erwerben. (MAS) - Arbeiten besser zu schaffen als andere. (PAP) - nicht so schwer zu arbeiten. (WOA) - mich mit Klassenkameraden auszutauschen. (AFL)

FAM – Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen, Rheinberg et al. (2001)

18 Items, 4 Faktoren:	Meine momentane Einstellung
<ul style="list-style-type: none"> - Herausforderung - Interesse - Erfolgswahrscheinlichkeit - Misserfolgsbefürchtung 	<ul style="list-style-type: none"> - Ich mag solche Rätsel und Knobeleyen. (IT) - Ich glaube, dass kann jeder schaffen. (EW) - Es ist mir etwas peinlich, hier zu versagen. (MB) - Die Aufgabe ist eine richtige Herausforderung für mich. (H)

Die Förderung von motivationsunterstützendem Unterricht, Kramer (2002)

38 Items, 6 Faktoren:	In dieser Unterrichtsstunde...
<ul style="list-style-type: none"> - Wahrgenommene Unterrichtsbedingungen der Lernsituation/Überforderung - Inhaltliche Relevanz - Instruktionsqualität/Klarheit - Soziale Eingebundenheit - Inhaltliches Interesse beim Lehrenden - Kompetenzunterstützung - Autonomieunterstützung 	<ul style="list-style-type: none"> - wurde ich in Situationen gebracht, wo ich selbst merken konnte, wie wichtig der Stoff ist. (IR) - ging mir alles zu schnell. (Ü) - habe ich einen Überblick über die geplante Vorgehensweise erhalten. (KL) - hat mich die Begeisterung der Lehrkraft richtig angesteckt. (II) - war die Atmosphäre freundlich und entspannt. (SE) - wurde ich sachlich über meine Fortschritte informiert. (KU) - durfte ich Aufgaben auf meine Art erledigen. (AU)

36 Items, 6 Faktoren:	Beim Lernen/Arbeiten in dieser Stunde...
<ul style="list-style-type: none"> - Amotiviert - External - Intrinsisch - Identifiziert - Introjiziert - Interessiert <p>(Prenzel et al. 1996)</p>	<ul style="list-style-type: none"> - versuchte ich mich zu drücken. (A) - hätte ich ohne Druck von außen nichts getan. (E) - habe ich mich selbst unter Druck gesetzt, um alles möglichst richtig/gut zu machen. (IJ) - wollte ich selbst den Stoff verstehen/beherrschen. (ID) - machte das Lernen/Arbeiten richtig Spaß. (IT) - hat mich die Sache so fasziniert, dass ich mich voll einsetzte. (INTR)

16 Items, 3 Faktoren:	Das Lernen/Arbeiten empfand ich als...
<ul style="list-style-type: none"> - Wichtigkeitsempfinden - Positive Empfindungen - Negative Empfindungen 	<ul style="list-style-type: none"> - frustrierend. (NE) - wichtig für Prüfungen. (WE) - anregend. (PE)

Anhand dieser Testinstrumente wird die Vielfalt der Motivationsarten und Einflussfaktoren deutlich. Es besteht kein einheitliches Instrument zur Erfassung der Motivation. Vielmehr werden unterschiedliche Bereiche wie zum Beispiel das situationale Interesse, das emotionale Erleben, das Selbstkonzept, die Zielorientierung, die psychologischen Bedürfnisse und die Motivationsformen befragt. Andere Einflussfaktoren wie individuelle Motivationsdispositionen und Selbstkonzepte (Wild und Möller 2020, S. 169), persönliche Werte sowie die Änderung der Motivation je nach Kontext oder im Laufe der Zeit (Krapp und Hascher 2014a, S. 234) sind bei den aufgelisteten Instrumenten nicht berücksichtigt. Auch wenn diese Instrumente keine vollständige Motivation abbilden können, bieten sie dennoch den Lehrpersonen Ansatzpunkte, die aktuelle Motivation der Lernenden zu beurteilen. Mit diesem Wissen können Lehrpersonen beispielsweise Einfluss auf das Interesse, die Bereitschaft zum Lernen oder das Engagement nehmen. Für ein erfolgreiches Lernen muss der Unterricht so gestaltet sein, dass die Schüler:innen motiviert sind (Florio-Hansen 2014, S. 61), denn die Motivation wirkt sich positiv auf die Bereitschaft zum Lernen, den Lernerfolg und Kompetenzzuwachs aus (Krapp und Hascher 2014b, S. 269; Brünken et al. 2019, S. 186).

2.4 Zusammenfassung in Bezug auf die vorliegende Erhebung

Deutlich wird, dass die Motivation einen entscheidenden Einfluss auf das Lernen und die Leistung von Schüler:innen hat (Deci und Ryan 1993; Schiefele und Streblow 2006; Spinath 2005). „Ohne Motivation bzw. ohne ihre Förderung und Aufrechterhaltung ist Unterricht auf Dauer nicht lernwirksam“ (Florio-Hansen 2014, S. 61). Das bedeutet, Lehrpersonen müssen Wissen über Theorien zur Motivation und Motivierung kennen sowie Kompetenzen besitzen, die Motivation der Schüler:innen zu erkennen und geeignete Maßnahmen zur Förderung der Motivation anwenden können. Bereits im Lehramtsstudium sollen angehende Lehrkräfte diese Kompetenzen erwerben: „Die Absolventinnen und Absolventen [...] kennen Theorien der Lern- und Leistungsmotivation und Möglichkeiten, wie diese im Unterricht angewendet werden“ (KMK 2019, S. 8). Zudem sollen sie praktische Kompetenzen besitzen, um das Lernen der Schüler:innen durch motivationsförderliche Unterrichtsgestaltung positiv zu beeinflussen (KMK 2019, S. 8). Um diese Kompetenzen zu vermitteln kann beispielsweise ein online basiertes Video-Lerntool genutzt werden (siehe Kapitel 4 *Einsatz von Videos in der Lehre*). Ein videografiertes Unterrichtsereignis bietet die Möglichkeit, bereits im Studium das Unterrichtsgeschehen des späteren Berufsalltags zu veranschaulichen und zu analysieren (van Es 2009, S. 101). Der Vorteil von Videos ist zudem die Generierung einer hohen Theorie-Praxis-Verzahnung für die Studierenden (Vgl. Bartel und Roth 2017, S. 43–44; Blomberg et al. 2013, S. 106). Mit einem online-basierten Instrument kann die Steigerung der Motivation, des Interesses und der

Lernbereitschaft gelingen (Luo et al. 2021, S. 1392–1393). Folglich setzen sich die Studierenden intensiver mit der Thematik auseinander. Wie Luo et al. (2021, S. 1391–1392) aufzeigen, wird die intrinsische Motivation am stärksten durch die wahrgenommene Kompetenz beeinflusst. Darauf zielt die vorliegende Arbeit mit der Entwicklung eines online Tools durch eine Kombination aus videografiertem Unterricht und Analysetool zur professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung im Unterricht ab.

Dabei bleibt das Problem existent, dass die tatsächliche Motivation der Schüler:innen - auch im Video, durch die Lehrkräfte schwer zu erfassen ist (Florio-Hansen 2014, S. 57; Krapp und Hascher 2014a, S. 234). Deshalb wird im hier entwickelten videobasierten Tool die aktuelle Motivation der Lernenden über einen Fragebogen erfasst (siehe Abschnitt 7.1.2 *Methodisches Vorgehen zur Entwicklung und Validierung der Vignetten und Analyseaufträge*), der die fünf Motivationsarten entsprechend der Selbstbestimmungstheorie nach Deci und Ryan (1993) erhebt. Dazu werden verschiedene Instrumente zur Erhebung der Motivation (siehe Abschnitt 2.3 *Messung der Motivation im Lernkontext*) hinsichtlich der Reliabilität, Validität und Eignung für das Projektvorhaben analysiert. Entsprechend dieser Analyse (siehe Anhang A) scheint das verwendete Instrument aus der Interventionsstudie „Die Förderung von motivationsunterstützendem Unterricht“ (Kramer 2002) von Prenzel et al. (1996) am geeignetsten (Anhang B). Es bedarf lediglich geringer sprachlicher Anpassungen für die Erhebung in der Schule, da die ursprüngliche Konzeption im Ausbildungsbereich verortet ist.

Bieg und Mittag (2011, S. 225–231) können aufzeigen, dass sich die Weiterbildung zu motivationsförderlichen Maßnahmen und Verhaltensweisen im Unterricht auf Basis der Selbstbestimmungstheorie (Deci und Ryan 1993) positiv auf die Leistung der Schüler:innen auswirkt. Daher wird ein videobasiertes Instrument, das auf motivationsförderliche Maßnahmen zielt, als sinnvoll erachtet. Grundlage zur Entwicklung der Videoausschnitte bildet das IBAS-Modell (Benning et al. 2019), welches relevante Faktoren zur Steigerung der Motivation durch das Erreichen einer Lernzielkultur anhand von vier Dimensionen aufzeigt und klar strukturiert. Anhand dieses Modells wird deutlich, dass es hinsichtlich des motivationsförderlichen Handelns von Lehrpersonen viele Facetten gibt. Dabei ist für (angehende) Lehrpersonen zum einen die Kenntnis über die Maßnahmen wichtig, aber auch die Umsetzung im Unterricht: Wie kann ich die Motivation meiner Schüler:innen fördern? Das in dieser Arbeit vorgestellte Videotool greift einzelne Facetten aus dem IBAS-Modell auf, stellt mit Videos aus dem realen Unterricht einen Alltagsbezug her und verknüpft durch Analyseaufträge zur Motivierung die Theorie mit der Praxis.

3 Professionelle Unterrichtswahrnehmung

Um die Schüler:innen motivieren zu können, wird neben dem Wissen über Theorien und Modelle zur Motivation sowie dem Einsatz motivationsfördernder Aspekte im Unterricht auch die Fähigkeit, die Motivation der Schüler:innen wahrzunehmen benötigt. Die Herausforderung besteht darin, die tatsächliche Motivation zu erfassen, da sie als hoch-inferenter Prozess nicht eindeutig von außen erkennbar ist (Büttner 2008, S. 288). Wenn es gelingt, die Motivation wahrzunehmen und zu beschreiben, ist die Voraussetzung gegeben, adäquat und flexibel im Unterricht darauf reagieren zu können (Seidel et al. 2010; Sherin und van Es 2008). Dieser Prozess der Professionellen Unterrichtswahrnehmung bildet ein zentrales Element für das erfolgreiche pädagogische Handeln (Kersting et al. 2010; Schäfer und Seidel 2015; Seidel et al. 2010) und wird mit seinen Teilprozessen *Noticing und Reasoning* im Abschnitt 3.2 ausführlich beschrieben. Da es sich bei der professionellen Unterrichtswahrnehmung um einen komplexen Prozess handelt, sind *Unterschiede zwischen Noviz:innen und Expert:innen* nicht nur denkbar, sondern auch empirisch gesichert (siehe Abschnitt 3.3 *Forschungslage*). Mit der Entwicklung eines videobasierten Analysetools soll mit der vorliegenden Arbeit ein Beitrag zur Analyse und Förderung der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung angehender Lehrpersonen geleistet werden. Zunächst folgt ein kurzer *Überblick angrenzender Konstrukte* (siehe Abschnitt 3.1), die im Zusammenhang mit der Fähigkeit des Beurteilens von Lehrpersonen in der Literatur häufig verwendet werden.

3.1 Überblick angrenzender Konstrukte

In diesem Abschnitt liegt der Fokus darauf, einen Überblick über angrenzende Konstrukte zu bieten, die in Verbindung mit der Professionelle Unterrichtswahrnehmung stehen. Es werden kurz zentrale, verwandte Konzepte vorgestellt, um einen Einblick in das Themengebiet zu ermöglichen und zu verdeutlichen, welche weiteren Facetten bei der Betrachtung berücksichtigt werden können. Dabei werden verschiedene relevante Begriffe und Zusammenhänge kurz erläutert, ohne jedoch in eine detaillierte Vergleichsanalyse einzusteigen.

Neben dem Begriff der Professionellen Unterrichtswahrnehmung werden im deutschsprachigen Raum häufig die Begriffe Diagnostische Kompetenz - Pädagogische Diagnostik - Urteilsgenauigkeit - Diagnostische (Urteils-)Fähigkeit verwendet, um die Fähigkeit des Beurteilens einer Lehrkraft zu beschreiben. In der Literatur wird von unterschiedlichen Verständnissen dieser Fähigkeit ausgegangen. Als zentraler Begriff wird dabei die **Diagnose** genannt: „Das Wort leitet sich aus dem griechischen *diagnostikos* (zum Unterscheiden geschickt) ab. *Diagnostik* ist eine professionelle, systematische, wissenschaftlich und methodisch fundierte Tätigkeit mit dem

Ziel, Erkenntnisse über Merkmalsträger[:innen] zu gewinnen oder Entscheidungen über nachfolgende Maßnahmen treffen zu können. Dabei wird ein Ist-Zustand erfasst (Helmke 2015, S. 272).“ Eine solche diagnostische Beurteilung gehört zum Schulalltag dazu, „durch die bei einzelnen Lernenden und den in einer Gruppe Lernenden Voraussetzungen und Bedingungen planmäßiger Lehr- und Lernprozesse ermittelt, Lernprozesse analysiert und Lernergebnisse festgestellt werden, um individuelles Lernen zu optimieren (Ingenkamp und Lissmann 2008, S. 13).“ Ingenkamp und Lissmann (2008, S. 13) fassen solche diagnostischen Tätigkeiten als **Pädagogische Diagnostik** zusammen. Dabei soll die Pädagogische Diagnostik nicht nur bei individuellen Entscheidungs- und Förderungsmaßnahmen zum Einsatz kommen, sondern auch zur Analyse und Adaption der eigenen Unterrichtsgestaltung angewandt werden (Ingenkamp und Lissmann 2008, S. 13–14). Insgesamt wird der Prozess als Interaktion zwischen Lehrenden und Lernenden verstanden, der sich über mehrere Phasen vollzieht: von der Erwartung über die Informationsaufnahme, Urteilsbildung hin zur Urteilsreaktion und Handlung seitens der Lernenden (Ingenkamp und Lissmann 2008, S. 15–17). Die diagnostischen Tätigkeiten wie sie Ingenkamp und Lissmann (2008, S. 13) ausführen, lassen sich auch in den von der KMK formulierten Standards zur Lehrerbildung als Kompetenzformulierung finden: „Lehrkräfte diagnostizieren Lernvoraussetzungen und Lernprozesse von Schülerinnen und Schülern; sie fördern Schülerinnen und Schüler gezielt [...]“ (KMK 2019, S. 11). Zusätzlich wird bei der KMK (2019, S. 11) der Beratungsaspekt der Lernenden aufgegriffen. Die Kompetenz, die Personen zum Diagnostizieren benötigen, wird in der Literatur als **diagnostische Kompetenz** definiert. Helmke (2015, S. 119) beschreibt, dass "[u]nter diagnostischer Kompetenz [...] im schulischen Kontext heute meistens die Fähigkeit verstanden [wird], Personen oder Personengruppen (z.B. Schulklassen) zutreffend zu beurteilen bzw. genaue diagnostische Urteile abzugeben." Eine Erweiterung dieser Definition nehmen Artelt und Gräsel (2009, S. 157) vor, indem sie zusätzlich zur Beurteilung der Merkmale der Lernenden die Analyse der „Lern- und Aufgabenanforderungen“ (Artelt und Gräsel 2009, S. 157) als Aspekt der diagnostischen Kompetenz berücksichtigen. Die Definition nach Weinert (2000, S. 14) spiegelt die Forderungen der KMK am ehesten wieder, wenn er beschreibt, dass es sich bei diagnostischen Kompetenzen „um ein Bündel von Fähigkeiten [handelt], um den Kenntnisstand, die Lernfortschritte und die Leistungsprobleme der einzelnen Schüler[:innen] sowie die Schwierigkeiten verschiedener Lernaufgaben im Unterricht fortlaufend beurteilen zu können, sodass das didaktische Handeln auf diagnostischen Einsichten aufgebaut werden kann (Weinert 2000, S. 14)“. Weinert bezieht die Diagnostik zum einen auf der Schüler:innenebene, aber auch auf der Aufgabenebene und beschreibt Auswirkungen auf das Handeln der Lehrperson im Sinne eines fortlaufenden Prozesses. Die diagnostische Kompetenz stellt eine essentielle Voraussetzung für „die notwendige Individualisierung und Differenzierung“ (Weinert 2000, S. 14) dar. Damit geht Weinert (2000, S. 14)

implizit von der Folge Beobachtung, Beurteilung, Handeln aus. Die **diagnostischen Fähigkeiten** werden von Leuders et al. (2018) konkreter operationalisiert, nämlich als Prozess des Wahrnehmens, Interpretierens und Entscheidens. Dabei verstehen sie die diagnostischen Fähigkeiten als situationsspezifische, kognitive Fähigkeiten (Leuders et al. 2018, S. 8). Im direkten Vergleich können die diagnostischen Fähigkeiten als systematische Diagnostik verstanden werden, während diagnostische Kompetenzen allgemein den Beobachtungsprozess und das abgeleitete didaktische Handeln fokussieren.

Analysiert man die verschiedenen Definitionen, so wird deutlich, dass für das Handeln von Lehrpersonen sowohl Diagnosen von Merkmalen der Lernenden als auch Merkmale des Unterrichts erfasst werden müssen. Differenziert betrachtet können auf der Mikroebene die Lernenden hinsichtlich ihrer Lernfortschritte, Lernstände und Leistungsprobleme eingeschätzt werden (Schrader 2009; Weinert 2000, S. 14; KMK 2019, S. 11; Helmke 2015, S. 119; Ingenkamp und Lissmann 2008, S. 13). Diese Beurteilung wirkt sich auf die Mesoebene, also die Unterrichtsgestaltung aus, indem beispielsweise die Anforderungsgrade der Aufgaben und Inhalte auf die Lernenden abgestimmt werden, um eine individuelle Förderung gewährleisten zu können (Weinert 2000, S. 14; Artelt und Gräsel 2009, S. 157; Helmke 2015, S. 121). Die Beurteilung der diagnostischen Kompetenz erfolgt häufig über den Aspekt der „Urteils- oder Diagnosegenauigkeit“ (Schrader 2009, S. 237) oder den der „Akkuratheit von Lehrer[:innen]einschätzungen“ (Spinath 2005, S. 85). Dabei muss ebenso „methodisches und prozedurales Wissen (Verfügbarkeit von Methoden zur Einschätzung von Schülerleistungen und zur Selbstdiagnose) als auch konzeptuelles Wissen (Kenntnis von Urteilstendenzen und -fehlern)“ (Helmke 2015, S. 119) betrachtet werden, um „diagnostische Tätigkeiten“ (Ingenkamp und Lissmann 2008, S. 13) durchführen zu können.

Eingebettet in Modelle können die diagnostischen Fähigkeiten und Kompetenzen von Lehrpersonen systematisch erfasst, untersucht und gefördert werden. Im Folgenden werden drei in der Forschung relevante Modellansätze mit ihren unterschiedlichen Dimensionen und Teilprozessen beschrieben.

Modelle

Brunner et al. (2011) haben sich in ihrer Untersuchung im Rahmen der COACTIV-Studie mit den Diagnostischen Fähigkeiten von Lehrkräften im Bereich der Mathematik auseinandergesetzt. Grundlage für die Verortung des Begriffs der Diagnostischen Fähigkeit ist das Kompetenzmodell von COACTIV (Baumert und Kunter 2011). Im Rahmen von COACTIV konnte dieses Kompetenzmodell, basierend auf den bisherigen theoretischen Annahmen zur Kompetenz- und Professionsdebatte, entwickelt werden, um die Professionelle Handlungskompetenz von Lehrkräften zur Bewältigung

der alltäglichen Anforderungen im Beruf herauszustellen (Baumert und Kunter 2011, S. 30). Dem Professionswissen als einem Aspekt professioneller Kompetenz kommt hierbei eine besondere Bedeutung zu (Baumert und Kunter 2011, S. 33). Es wird durch verschiedene Kompetenzbereiche, die wiederum verschiedene Kompetenzfacetten abbilden, differenziert (siehe Abbildung 7). Die diagnostischen Fähigkeiten werden nach Brunner et al. (2011, S. 216) innerhalb der Kompetenzbereiche fachdidaktisches Wissen und pädagogisch-psychologisches Wissen im Aspekt Professionswissen angesiedelt (siehe gestrichelter Rahmen in Abbildung 7).

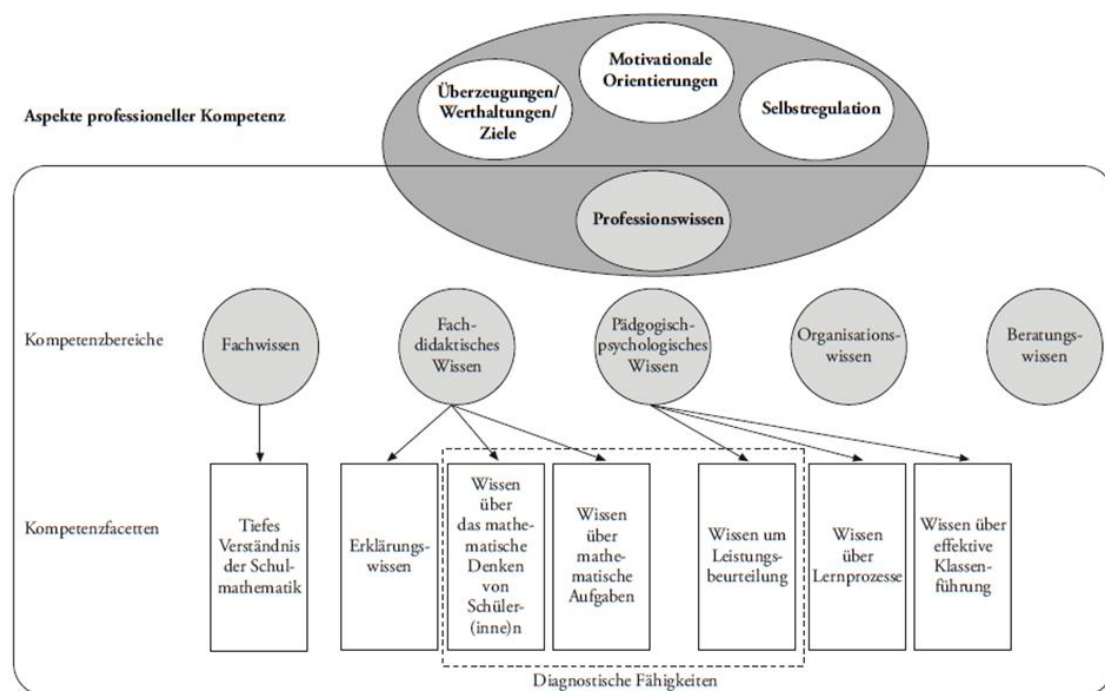


Abbildung 7 Professionelle Kompetenz von Lehrpersonen. Einordnung der Diagnostischen Fähigkeiten (Brunner et al. 2011, S. 217) in das Kompetenzmodell von COACTIV

Ausgehend von diesem Ansatz sind drei Facetten des Professionswissens für die diagnostischen Fähigkeiten von Lehrkräften relevant (Brunner et al. 2011, S. 217): „Wissen über das mathematische Denken von Schüler(inne)n“, „Wissen über mathematische Aufgaben“ und „Wissen um Leistungsbeurteilung“. Die weiteren Kompetenzfacetten des Professionswissens sowie die Aspekte der Überzeugungen, Motivationalen Orientierungen und Selbstregulation finden in dieser schematischen Darstellung keine Berücksichtigung, sind jedoch laut Ingenkamp & Lissmann (2008) ebenso entscheidend für den Prozess der diagnostischen Beurteilung. Solche Aspekte werden im Modell von Blömeke et al. (2015) (siehe unten) integriert. Der Fokus bei Beretz et al. (2017) liegt wiederum auf dem reinen Diagnoseprozess und wird daher als nächstes kurz beschrieben:

Zur Förderung der diagnostischen Kompetenz beschreiben Beretz et al. (2017) ein fünf-stufiges Modell zur Beschreibung des Diagnoseprozesses (siehe Abbildung 8).

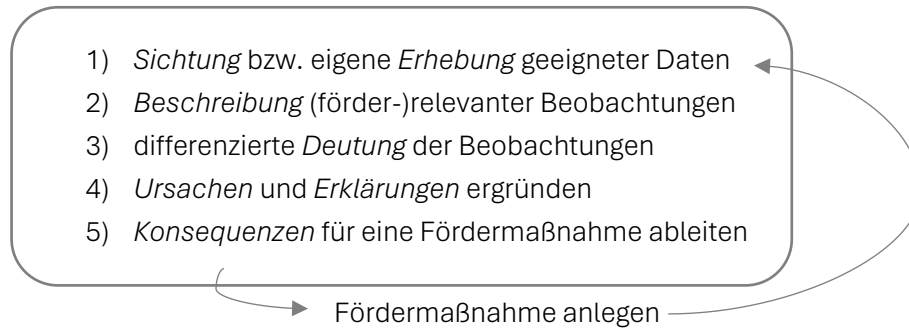


Abbildung 8 Diagnoseprozess, in Anlehnung an Beretz et al. (2017, S. 151)

Um die Studierenden an den Prozess des Diagnostizierens heranzuführen, ist es vorteilhaft, zunächst die diagnostischen Elemente von der Förderung zu trennen (Beretz et al. 2017, S. 151–152). Bei der Förderung diagnostischer Kompetenzen im Studium sollten die einzelnen Phasen klar voneinander abgegrenzt werden, um differenzierte Analysen vornehmen zu können (Beretz et al. 2017, S. 166), auch wenn es sich bei dem Modell um kein striktes Phasenmodell handelt, sondern um einen „iterativen Prozess“ (Beretz et al. 2017, S. 151). Der strukturierte Ablauf zur Analyse von Lernsituationen ermöglicht ein wirksames Ableiten von Fördermaßnahmen. Die Förderung der Lernenden spielt eine zentrale Rolle bei Beretz et al. (2017). Die Elemente des Beobachtens, Beschreibens, Erklärens und Ableitens von Konsequenzen lassen sich auch im Modell von Blömeke et al. (2015) wiederfinden.

Blömeke et al. (2015) betrachten in ihrem Modell *Competence as Continuum*, auch bekannt als *PID*-Modell (Perception, Interpretation, Decision-making), die Kompetenz von Lehrpersonen als einen mehrstufigen und zusammenhängenden Prozess, der sich zwischen den Voraussetzungen, situationsspezifischen Fähigkeiten und dem Handeln einer Lehrperson bewegt (siehe Abbildung 9). Dabei beschreiben die Autoren die Kompetenz als ein Kontinuum zwischen den Dispositionen auf der einen und der Performanz auf der anderen Seite (Blömeke et al. 2015, S. 7). Die situationsbezogenen Fähigkeiten, verlaufen als mediiender Prozess und setzen sich aus den drei Aspekten Wahrnehmung, Interpretation und Entscheidungsfindung zusammen (Blömeke et al. 2015, S. 7). Lehrpersonen müssen zunächst relevante Situationen gezielt wahrnehmen. Sie müssen diese im Anschluss fachlich und pädagogisch richtig deuten, um fundierte Entscheidungen für ihr Handeln ableiten zu können.

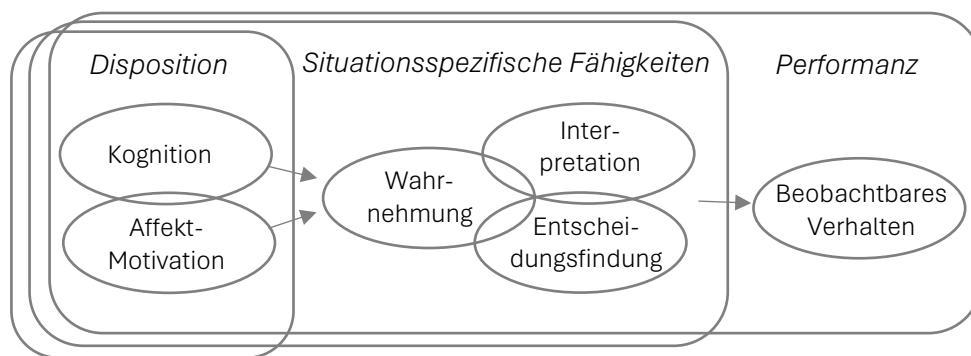


Abbildung 9 Situationsbezogene Fähigkeiten im Modell „Kompetenz als Kontinuum“, Übersetzung und Darstellung in Anlehnung an Blömeke et al. (2015, S. 7)

Die situationspezifischen Fähigkeiten sowie das professionelle Wissen spielen ebenfalls bei der Professionellen Unterrichtswahrnehmung nach Seidel et al. (2010) und Sherin und van Es (2008) eine zentrale Rolle. Diese wird im nächsten Abschnitt (3.2) detailliert vorgestellt und zu den bisher genannten Modellen in Beziehung gesetzt (siehe Abschnitt 3.3)

Insgesamt ist festzuhalten, dass unabhängig von der begrifflichen Zuweisung, die Förderung der Fähigkeit des Diagnostizierens bzw. der Professionellen Unterrichtswahrnehmung der Weiterentwicklung des professionellen Handelns der Lehrperson dient (Jahn et al. 2014) und damit positiv auf die Unterrichtsqualität und den Lernerfolg wirkt (Kersting et al. 2010; Seidel et al. 2010).

3.2 Prozess der Professionellen Unterrichtswahrnehmung

Die Professionelle Unterrichtswahrnehmung, im Englischen als *professional vision* bezeichnet, lässt sich auf Goodwin (1994) zurückführen. In seiner Arbeit zeigt Goodwin (1994, S. 606) anhand von berufsbezogenen Praktiken aus der Archäologie und der Justiz auf, wie die Aufmerksamkeit auf berufliche Ereignisse gelenkt und diese durch den gemeinsamen Diskurs in Wissensobjekte übertragen werden können. Die Wahrnehmung, Interpretation und Entwicklung der *professional vision* innerhalb verschiedener Berufsfelder erfolgt durch mentale Prozesse sowie komplexe soziale Interaktion im Handlungsfeld (Goodwin 1994, S. 627–628). Dies lässt vermuten, dass Lehrpersonen über eine spezifische Professionelle Wahrnehmung verfügen (müssen).

Die Übertragung und Differenzierung des Konzepts der *professional vision* auf den Bildungsbereich wird vor allem mit den Arbeiten von Sherin & Kolleg:innen (Sherin 2009; Sherin und van Es 2008; Es und Sherin 2002; Sherin 2002; Sherin und Han 2004) in Zusammenhang gebracht. Im deutschsprachigen Raum haben Seidel et al. (2010) die Professionelle Unterrichtswahrnehmung im Hinblick auf die Lehr-Lernforschung weiter etabliert. Dabei geht es weit über eine passive Beobachtung des Unterrichtsgeschehens hinaus (Sherin 2009, S. 384–385). Wie eingangs beschrieben, besteht

die Herausforderung im Berufsalltag einer Lehrperson daraus, das komplexe Unterrichtsgeschehen richtig wahrzunehmen, zu verarbeiten und adäquat darauf zu reagieren (Seidel et al. 2010; Sherin 2009, S. 384–385). Dies schließt mit ein, dass eine Lehrperson Vorhersagen für künftige Entscheidungen treffen bzw. den aktuellen Unterricht adaptieren und Auswirkungen auf den Lernprozess feststellen kann (Seidel et al. 2010; Sherin und van Es 2008). Dieser Prozess der Professionellen Unterrichtswahrnehmung bildet ein zentrales Element für das pädagogische Handeln (Schäfer & Seidel, 2015).

Die Professionelle Unterrichtswahrnehmung setzt sich aus zwei interagierenden Prozessen zusammen, dem „Noticing“ und dem „Knowledge-based Reasoning“ (Seidel et al., 2010; Sherin & van Es, 2008) (siehe Abbildung 10), die in einer dynamischen Interaktion aus top-down und bottom-up Prozessen vorkommen (Sherin 2009, S. 384).

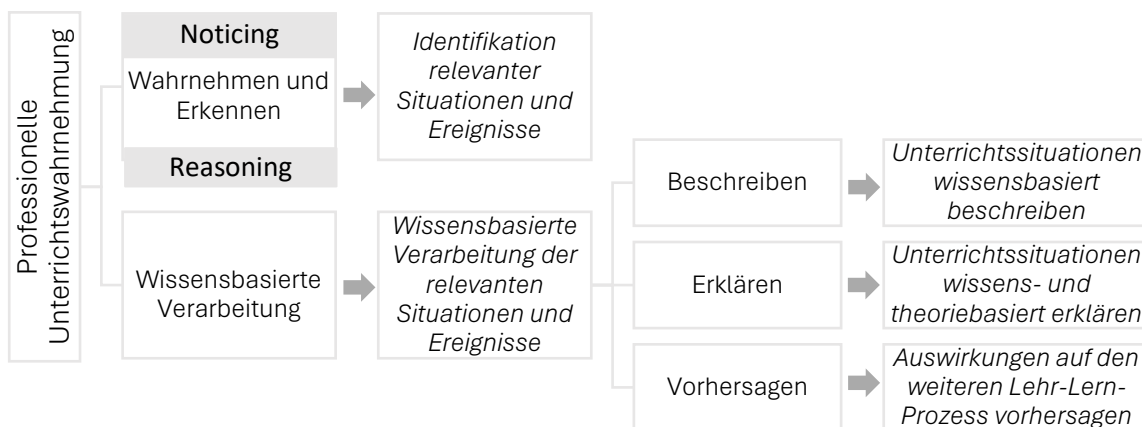


Abbildung 10 Professionelle Unterrichtswahrnehmung in Anlehnung an Seidel et al., 2010; Sherin & van Es, 2008; van Es, 2009.

Das Noticing beschreibt grundlegend das Wahrnehmen und Erkennen von Situationen und Ereignissen, die relevant für das Lehren und Lernen im Unterricht sind (u.a. Seidel et al. 2010, S. 297; Sherin 2009, S. 384). Die Herausforderung besteht darin, innerhalb des komplexen Unterrichtsgeschehens den Aufmerksamkeitsfokus stets anzupassen und auf die wesentlichen Ereignisse zu lenken (Sherin und van Es 2008, S. 22; Sherin 2009, S. 384). Dies erfordert ein hohes Maß an Selektionsfunktion. Die Lehrperson muss in der Lage sein, die wichtigen Informationen zu filtern und sich auf relevante Aspekte für den Unterrichtsprozess zu konzentrieren, z.B. auf Interaktionen zwischen Lernenden (Sherin 2009, S. 384). Die wissensbasierte Verarbeitung dieser wahrgenommenen Informationen erfolgt im Prozess des Knowledge-based Reasonings (Seidel et al. 2010; Sherin und van Es 2008; van Es 2009). Dadurch ist es möglich, dass Lehrpersonen basierend auf der Wahrnehmung relevanter Situationen fundierte pädagogische Entscheidungen treffen, indem sie beispielsweise auf ihr Fachwissen, Empfehlungen der Curricula oder Erfahrungen mit den Lernenden

zurückgreifen (Sherin und van Es 2008, S. 22). Damit können Lehrpersonen effektiver auf die individuellen Bedürfnisse ihrer Lernenden eingehen und ihren Unterricht entsprechend anpassen. Das Knowledge-based Reasoning kann in drei Aktivitäten eingeteilt werden: Beschreiben, Erklären und Vorhersagen (Seidel et al. 2010; Sherin und van Es 2008; van Es 2009). In der pädagogischen Praxis kann eine als relevant identifizierte Situation zunächst wissensbasiert beschrieben werden.

Zur Erläuterung ein Beispiel aus dem Unterricht:

Die Lehrerin hat im ersten Schritt eine relevante Situation für den Lehr-Lern-Prozess wahrgenommen und identifiziert. Ihre wissensbasierte Beschreibung sieht folgendermaßen aus. Ein Schüler liest mehrfach konzentriert die Aufgabenstellung, statt wie alle anderen Lernenden der Klasse mit der Bearbeitung der Aufgabe zu beginnen, klappt das Heft zu und schaut aus dem Fenster. Anmerkung: Dieser Schüler ist seit drei Wochen neu in der Klasse und hat aufgrund seines Migrationshintergrunds Schwierigkeiten mit dem Text- und Sprachverständnis.

Im Anschluss an die Beschreibung der Situation kann diese auf Basis von Wissen und Theorien erklärt werden.

Die Lehrerin kann auf Grund der Motivationstheorien und Modelle das Verhalten des Schülers analysieren und mögliche Ursachen identifizieren:

- 1) die sprachliche Schwierigkeit des Schülers erschwert das Verstehen der Aufgabe*
- 2) die Aufgabenstellung ist zu schwierig bzw. zu wenig differenziert für den Schüler*

Schließlich können aus der Analyse Auswirkungen für den weiteren Lehr-Lern-Prozess vorhergesagt und entsprechende Interventionsmöglichkeiten abgeleitet werden.

Die Lehrerin könnte basierend auf ihren Erkenntnissen gezielte Maßnahmen ergreifen, um den betroffenen Schüler zu unterstützen, die Aufgabenstellung zu verstehen. Dies könnte beispielsweise eine vereinfachte Sprache oder zusätzliche Erklärungen und Hilfestellungen umfassen. Durch die Anwendung von theoretischem Wissen und einer fundierten Analyse kann die Lehrerin den Unterricht gezielt anpassen und so den Lernerfolg des Schülers verbessern. Der hier beschriebene Prozess der Professionellen Unterrichtswahrnehmung der Lehrerin kann langfristig Auswirkung auf ihre Unterrichtsvorbereitung haben, da sie perspektivisch differenzierte Aufgabenstellungen nach Niveau und sprachlicher Kompetenz ihrer Lernenden formulieren möchte.

Der Prozess der Professionellen Unterrichtswahrnehmung zeigt, wie Lehrpersonen durch wissenschaftliche Beschreibung, theoriebasierte Erklärung und Vorhersage von Auswirkungen auf den Lehr-Lern-Prozess effektiv auf relevante Situationen im Unterricht reagieren können. Dabei erfordern das Noticing und Knowledge-based Reasoning eine kontinuierliche Reflexion über den eigenen Unterricht und die Bereitschaft, das eigene Handeln nicht nur zu überdenken, sondern auch anzupassen.

3.3 Zusammenfassung in Bezug auf die vorliegende Erhebung

In den ersten beiden Abschnitten dieses Kapitels wurden verschiedene Begriffe sowie zentrale Theorien und Modelle zur Diagnose und Professionellen Wahrnehmung von Lehrpersonen im Unterricht vorgestellt. Alle Konzepte thematisieren das professionelle Handeln von Lehrpersonen auf Grundlage von Beobachtungen und Analysen. Dabei kommt es zu fortlaufenden Anpassungen im Unterrichtsprozess.

Diese Arbeit folgt dem adaptierten Konzept der professional vision (Sherin und van Es 2008), im deutschsprachigen Raum als Professionelle Unterrichtswahrnehmung bezeichnet, auf Grundlage der Arbeit von Goodwin (1994) mit den Teilprozessen des Noticings und Knowledge-based Reasonings (Abb. 11). Im Vergleich zur Diagnostischen Kompetenz, die sich stärker auf die individuellen Voraussetzungen und Kenntnisse der Lernenden und die Gestaltung der Aufgaben bezieht, wird bei der Professionellen Unterrichtswahrnehmung das ganzheitliche Unterrichtsgeschehen mit Interaktionen und Lernprozessen näher betrachtet. Die Zielebene kann auch als Abgrenzung der Professionellen Unterrichtswahrnehmung zur diagnostischen Fähigkeit gesehen werden. Hier liegt der Fokus auf konkreten Teilhandlungen im Diagnoseprozess und dem Durchführen der schrittweisen Analyse.

Noticing	Wahrnehmen und Erkennen relevanter Situationen und Ereignisse im Unterrichtsgeschehen
Knowledge-based Reasoning	Verarbeitung der relevanten Unterrichtsgeschehnisse durch wissenschaftliche und theoriebasierte a) Beschreibungen b) Erklärungen c) Vorhersagen der Auswirkungen auf den weiteren Lehr-Lernprozess

Abbildung 11 Professionelle Unterrichtswahrnehmung mit den Teilprozessen (Vgl.: Seidel et al., 2010; Sherin & van Es, 2008; van Es, 2009)

Die Entscheidung für das Modell der Professionellen Unterrichtswahrnehmung erfolgt zudem auf Grund der Anschlussfähigkeit des vorliegenden Forschungsvorhabens an die bisherigen empirischen Arbeiten hinsichtlich der Analyse und Förderung der Professionellen Unterrichtswahrnehmung im Kontext videobasierter Lehr-Lernformate. Das Videotool Observer spielt dabei eine entscheidende Rolle. Es verdeutlicht die Eignung videobasierter Instrumente zur differenzierten Erfassung der

Professionellen Unterrichtswahrnehmung entsprechend des Modells nach Sherin und van Es (2008) bzw. Seidel et al. (2010).

Im Vergleich zu diesem Modell ist das beschriebene Coactiv Modell breiter in der Theorie angelegt. Es beschreibt die diagnostischen Fähigkeiten als Teilkompetenzen des fachdidaktischen Wissens und des pädagogisch-psychologischen Wissens, die ihrerseits zwei Kompetenzbereiche des Professionswissens sind (Brunner et al. 2011, S. 216). Das Professionswissen wird durch drei weitere Aspekte professioneller Kompetenz ergänzt: Überzeugungen/Werthaltungen/Ziele, motivationale Orientierungen und Selbstregulation (Brunner et al. 2011, S. 217). Diese komplexe theoretische Einordnung erlaubt eine umfassende Betrachtung der Kompetenzen, die eine Lehrperson benötigt. Für die vorliegende Arbeit ist dies zu allgemein und hinsichtlich der praktischen Umsetzung des Vorhabens zu wenig fokussiert auf die Prozesse des Wahrnehmens von Unterrichtssituationen und der kognitiven Verarbeitung dieser Informationen.

Demgegenüber heben sich das PID-Modell (Blömeke et al. 2015), das Stufenmodell diagnostischer Kompetenz nach Beretz et al. (2017) sowie das Modell der Professionellen Unterrichtswahrnehmung (Sherin und van Es 2008) durch eine klare Struktur der Diagnoseprozesse ab. Das PID-Modell und die Professionelle Unterrichtswahrnehmung integrieren explizit den Prozess der Wahrnehmung als einen wesentlichen Ausgangspunkt. Bevor Aspekte im Unterricht interpretiert und ein Handeln abgeleitet werden kann, braucht es eine gute Wahrnehmung von Situationen oder Merkmalen im Unterricht, die relevant für das Erreichen der Unterrichtsziele sind. Auch bei Beretz et al. (2017, S. 151) geht dem Beschreiben eine Stufe voraus: Die „Sichtung/eigene Erhebung geeigneter Daten“ kann dabei eher als Bereitstellung geeignetem Materials verstanden werden. Die Wahrnehmung relevanter Ereignisse wird dafür im folgenden Schritt als „Beschreibung (förder-)relevanter Beobachtungen“ (Beretz et al. 2017, S. 151) implizit als Komponente berücksichtigt. Allerdings bleibt die Beurteilung der Relevanz von Beobachtungen gleichzeitig mit dem Beschreiben verbunden und wird nicht als separater Teilprozess wie beim PID-Modell oder der Professionellen Unterrichtswahrnehmung gesehen. Die wahrgenommenen Situationen werden bei Blömeke et al. (2015) im Anschluss interpretiert. Die Interpretation wird dabei nicht weiter untergliedert und als mittlerer Teilprozess zwischen der Wahrnehmung und der Entscheidungsfindung gesehen (ebd.). Dabei werden diese drei Fähigkeiten gestützt durch die kognitiven und affektiv-motivationalen Dispositionen (ebd.). Das Wissen wird in der Professionellen Unterrichtswahrnehmung im Resoning durch die „wissensbasierte“ Verarbeitung der wahrgenommenen Ereignisse ebenfalls implementiert (Seidel et al. 2010). Wohingegen die affektiv-motivationalen Aspekte der Lehrperson keine offensichtliche Berücksichtigung finden. Das Interpretieren und die Entscheidungsfindung bleiben bei Blömeke et al. (2015) auf einem zwei-stufigen

Prozess, währenddessen Beretz et al. (2017) diesen Prozess weiter ausdifferenziert. In ihrer Arbeit werden relevante Beobachtungen beschrieben, differenziert gedeutet, Ursachen und Erklärungen ergründet und im Anschluss Konsequenzen abgeleitet (Beretz et al. 2017, S. 151). Dies ermöglicht eine differenzierte Analyse – vor allem auch im Hinblick auf die Entwicklung eines videobasierten Tools zur Beurteilung der Fähigkeiten der Studierenden oder auch der Lehrpersonen hinsichtlich der Professionellen Unterrichtswahrnehmung bzw. der diagnostischen Kompetenzen. Dadurch können Förderbedarfe zielgerichteter abgeleitet werden. Hinsichtlich der Überprüfung der empirischen Validität könnte eine feinere Differenzierung unter Umständen weniger zuverlässig die Dimensionen trennen. Möglicherweise bietet die Professionelle Unterrichtswahrnehmung mit einer Drei-Teilung des Reasonings (Vgl. Seidel et al., 2010; Sherin & van Es, 2008; van Es, 2009) eine geeignete mittlere Differenzierung sowohl für die Analyse der Fähigkeiten als auch für die Validierung des Videoinstruments. Der erfolgreiche Einsatz von videobasierten Lehr-Lernformaten auf theoretischer Basis der Professionellen Unterrichtswahrnehmung kann mit verschiedenen Forschungsarbeiten belegt werden (siehe folgender Abschnitt 3.4).

3.4 Forschungslage zur Professionellen Unterrichtswahrnehmung

Die Professionelle Unterrichtswahrnehmung ist in der Literatur sowie in der Lehr- und Lernforschung fest verankert. Zentrale forschungsrelevante Arbeiten und empirische Ergebnisse werden in den folgenden zwei Abschnitten vorgestellt. Zunächst werden Ansätze zur Erfassung und Förderung der Professionellen Unterrichtswahrnehmung kurz dargestellt. Im Anschluss werden die Erkenntnisse aus der Expertiseforschung, bei denen es um das systematische Aufzeigen von Unterschieden zwischen Noviz:innen und Expert:innen geht, beschrieben. Im letzten Abschnitt (3.4.3) werden diese Ergebnisse zusammengefasst und Konsequenzen für die vorliegende Arbeit abgeleitet.

3.4.1 Professionellen Unterrichtswahrnehmung: Erfassen und Fördern

Um herauszufinden, wie stark die Fähigkeit der Professionelle Unterrichtswahrnehmung einzelner Lehrpersonen, Lehramtsanwärter:innen oder Lehramtsstudierender ausgeprägt ist, braucht es geeignete Instrumente, die dies erfassen können. Wie zuvor ausgeführt wird, die Professionelle Unterrichtswahrnehmung in zwei Teilprozesse gegliedert. In Forschungsarbeiten wird daher der Fokus in der Regel entweder auf das Noticing oder das Knowledge-based Reasoning gelegt.

Bei der Erfassung des Noticings liegt die Herausforderung im Konstrukt selbst. Es handelt sich hierbei um einen Wahrnehmungsprozess, bei dem durch gezielte Aufmerksamkeitslenkung relevante Situationen im Unterricht identifiziert werden (u.a.

Seidel et al. 2010, S. 297; Sherin 2009, S. 384). Das methodische Vorgehen in den Anfängen der Forschung zur Professionellen Unterrichtswahrnehmung hinsichtlich des Noticings verlief klassischerweise folgendermaßen: Lehrpersonen sollten Unterrichtsszenen anhand von Bildern oder Videos beschreiben, um Rückschlüsse auf ihre Wahrnehmungsprozesse ziehen zu können (Huang et al. 2023, S. 124). Streng genommen handelt es sich bei diesem Vorgehen bereits um den Verarbeitungsprozess des Beschreibens im Sinne des Reasonings (Huang et al. 2023, S. 124). Nichtsdestotrotz können retrospektiv Urteile darüber getroffen werden, worauf die Personen ihren Fokus gelegt haben. Beispielsweise konnten Sherin und van Es (2008, S. 20) in ihrer Video Club Studie aufzeigen, dass Lehrpersonen durch das gemeinsame Analysieren von Videos als Fortbildungsformat ihre Fähigkeiten in der Professionellen Unterrichtswahrnehmung sowohl im Noticing als auch im Reasoning verbessern konnten. Über das Beschreiben von Videoclips zu den Inhalten, die sie wahrgenommen haben, konnten in einer Pre-Post-Messung Unterschiede im Noticing aufgezeigt werden (Sherin und van Es 2008, S. 26). Zu Beginn lag der Fokus weniger auf dem mathematischen Denken der Schüler:innen, sondern mehr auf den Aussagen der Lehrkräfte hinsichtlich pädagogischer Themen (Sherin und van Es 2008, S. 26). Dies änderte sich zum zweiten Messzeitpunkt, und die Aufmerksamkeit für das mathematische Denken, gemessen an der Anzahl initiiertes Ideeneinheiten während des Videoclubformats, erhöhte sich (Sherin und van Es 2008, S. 26).

Die Forschungsgruppe Jacobs et al. (2010, S. 174) haben strukturierte Interviews zur Erfassung des Noticings hinsichtlich spezifischer mathematischer Kenntnisse zum mathematischen Denken der Schüler:innen angewendet. Dazu haben die angehenden und bereits praktizierenden Lehrpersonen retrospektiv Fragen zu Bildern und Videos, auf welche Weise und in welchem Umfang sie das Denken der Lernenden wahrnehmen, beantwortet (Jacobs et al. 2010, S. 171).

Aktuelle Forschungsansätze zur Erfassung des Noticing-Prozesses verwenden häufig die Methode des Eye-trackings (Vgl. Metaanalyse Keskin et al. 2024). Diese Vorgehensweise ermöglicht es, eine Blickanalyse der Lehrperson während des Unterrichtsgeschehens vorzunehmen und Informationen über die Aufmerksamkeitslenkung und Dauer des Fokus auf eine Situation zu sammeln (Huang et al. 2023, S. 124). Durch die Auswertung des Blickes im Moment des Geschehens kann der Wahrnehmungsprozess im Sinne des Noticings erfasst werden und es gelingt eine Abgrenzung zur wissensbasierten Beschreibung und Erklärung. Keskin et al. (2024) haben 98 Studien mit dem Forschungsschwerpunkten Eyetracking und Noticing zusammengefasst und interpretiert. In den Studien wurden verschiedene Medien zur Analyse der Augenbewegung eingesetzt: Hauptsächlich fanden die Eye-Tracking-Erhebungen innerhalb des Klassenzimmers statt, alternativ kamen Videos, Fotos, Online-Unterricht und Virtual-Reality-Klassenraumsimulationen zum Einsatz (Keskin et al. 2024, S. 12). Ergebnisse im Hinblick auf Unterschiede zwischen erfahrenen und

weniger erfahrenen Lehrpersonen werden im nächsten Teilabschnitt 3.3.2 *Expertiseforschung* erläutert.

Die größte Studienlage hinsichtlich der Professionellen Unterrichtswahrnehmung findet sich zum Knowledge-based Reasoning. Zur standardisierten und vergleichbaren Erfassung dieses Teilprozesses werden Instrumente genutzt, die es ermöglichen, Unterrichtssituationen exemplarisch aufzugreifen. Dafür eignen sich video- oder textbasierte Fallbeispiele (Syring et al. 2015, S. 682). Da es sich um einen komplexen Prozess handelt, der im Unterricht stattfindet, wird bei der Erfassung der Professionellen Unterrichtswahrnehmung häufig auf videobasierte Instrumente zurückgegriffen. Diese ermöglichen unter anderem ein realitätsnahes Setting, das ein besseres Hineinversetzen in die Situation erlaubt (Syring et al. 2015, S. 682). Im vierten Kapitel wird ausführlich der *Einsatz von Videos in der Lehre* mit den Vorteilen, Anwendungsarten und Einschränkungen beschrieben. Die Ergebnisse der verschiedenen Video-Studien verdeutlichen, dass es geeignete Instrumente sowohl zum Erfassen als auch zum Fördern der Professionellen Unterrichtswahrnehmung und der diagnostischen Fähigkeiten gibt. Beispiele dieser Forschungsarbeiten sind im Abschnitt 4.2 *Studienlage* ausführlich beschrieben. Deswegen erfolgt an dieser Stelle lediglich eine Übersicht dieser Studien (siehe Tabelle 2). Es konnte unter anderem gezeigt werden, dass es innerhalb der drei Prozesse des Reasonings unterschiedliche Fähigkeitsniveaus gibt (Stürmer et al. 2015, S. 352; Schäfer und Seidel 2015, S. 50–51; Meschede et al. 2015, S. 328; Vgl. Frommelt et al. 2016, S. 369). Das Beschreiben ist demnach leichter als das Erklären und das Erklären gelingt besser als das Vorhersagen. Dieser Aspekt wird ebenfalls im folgenden Abschnitt (3.3.2) zur *Expertiseforschung* aufgegriffen und näher erläutert.

Tabelle 2 Übersicht über ausgewählte Studien zur Erforschung der Professionellen Unterrichtswahrnehmung und Diagnosefähigkeit mittels Videos

Autor:innen	Video-clips	Klassenstufe	Items	Erfasste Konzepte	Unterrichtsaspekt
Seidel et al 2010; Stürmer & Seidel 2014; Jahn et al 2014	Observer: 6 Clips, 2-4 Min. Länge	8./9. Klasse	Geschlossen: 112 Items, 4-stufig	Reasoning: Beschreiben, Erklären, Vorhersagen	Zielklarheit, Unterstützung der Lehrperson, Lernklima
Schäfer & Seidel 2015	Observer: 1 Clip, 3 Min.	7.Klasse	Offen: 1 Item Geschlossen: 36 Items	Noticing & Reasoning	Lehrstrategien

Seidel & Stürmer 2015	Observer Extended: 10 Clips, 2-4 Min.		Geschlossen: 41 Items, 4-stufig	Reasoning: Beschreiben, Erklären, Vor- hersagen	Zielsetzung, Orientierung, Lernaktivität, Regulation, Evaluation
Gold et al 2013	7 Clips, 1-3 Min.	Grund- schule	Geschlossen	Prof. Unter- richtswahr- nehmung	Klassenführung
Hellermann et al. 2015	eigene & 4 fremde Videos, 2-5 Min.	Grund- schule	Geschlossen: 47 Ratingitems	Beschreiben und Interpre- tieren	Klassenfüh- rung: Monito- ring, prozessua- le Strukturie- rung, Regeln und Routinen
Meschede et al. 2015; Sunder 2016	6 Clips	Grund- schule	Geschlossen: 68 Rating- Items	Beschreiben und Interpre- tieren	Lernunterstüt- zung
Frommelt et al 2016	Observer: 3 Clips & eigene Vi- deos	8./9. Klasse	Geschlossen: 4-stufig	Reasoning: Beschreiben, Erklären, Vor- hersagen	Zielklarheit, un- terstützende Lernbegleitung, positive Lernat- mosphäre
Santagata & Guarino 2011	1 Clip, 16 Min.	1. Klasse	Offen	Noticing, Be- schreiben, Kommentie- ren, Diskutie- ren	Lernrelevante Inhalte und Lehrstrategien
Star & Strick- land 2008	2 Clips, 45 und 50 Min.	8. Klasse	Offen & Geschlossen: Wahr/Falsch, multiple choice	Noticing	Klassenführung
Kersting et al 2010	13 Clips, 3-5 Min.	5./6. Klasse	Offen	Diskussion: Beschreiben, Erklären, Be- werten	Unterstützung der Lehrperson
Hofmann & Roth 2020	ViviAn	7./8.Klasse	Offen	Diagnosti- sche Fähig- keiten	Fehlkonzepte von Schüler:in- nen
Walz & Roth 2019	ViviAn: 3 Clips	Sekundar- stufe	14 Diagnose- aufträge	Diagnose, In- tervention, Reflexion	Gruppenar- beitsprozesse

Einige dieser Studien zeigen zudem auf, dass die Professionelle Unterrichtswahrnehmung gefördert werden kann und dies nicht erst mit zunehmender Praxiserfahrung, sondern bereits im Studium (ebenfalls Abschnitt 4.2 *Studienlage*). Dadurch können die künftigen Lehrpersonen frühzeitig ihre Fähigkeiten verbessern. Zudem kann mit dem Einsatz von Videos die Motivation der Studierenden in der universitären Veranstaltung erhöht werden (Syring et al. 2015, S. 682).

3.4.2 Expertiseforschung

Die Professionelle Unterrichtswahrnehmung ist von der Erfahrung der Lehrperson abhängig (Berliner 2001, 1987; Carter et al. 1988). Es konnte herausgestellt werden, dass Anfänger:innen im Lehrberuf, Referendar:innen und Lehramtsstudierende im Allgemeinen schlechtere Ergebnisse hinsichtlich der Ausprägung der Professionellen Unterrichtswahrnehmung erzielen als sogenannte Expert:innen im Beruf (Seidel und Prenzel 2007, S. 211–213). Dabei scheint es keinen signifikanten Unterschied zwischen Studierenden und Referendar:innen zu geben (Stürmer et al. 2015, S. 355). Dies deutet darauf hin, dass die Qualität der Unterrichtswahrnehmung mit zunehmender Berufserfahrung erst im Anschluss an die erste und zweite Phase der Lehrpersonenbildung steigt. Die Ausbildungsphase bleibt daher eine Zeit, in der anscheinend ohne gezielten Input die Professionelle Unterrichtswahrnehmung nicht gefördert wird. Wenn man sich beide Teilprozesse differenziert betrachtet, lassen sich bereits beim Noticing Unterschiede feststellen. Erfahrene Lehrpersonen können schneller zwischen relevanten Situationen wechseln, fokussieren sich dabei besser auf das Wesentliche und behalten einen breiteren Überblick über die Geschehnisse im Klassenraum (Huang et al. 2023, S. 131). In ihrer metaanalytischen Auswertung konnten Keskin et al. (2024, S. 20) diese Ergebnisse bestätigen: Noviz:innen lenken ihren Fokus weniger auf die Schüler:innen, sondern eher auf die Unterrichtsmaterialien oder unwichtigere Details im Klassenraum/Unterrichtsgeschehen im Vergleich zu Expert:innen. Dies kann dazu führen, dass sie möglicherweise nicht alle relevanten Informationen im Unterricht wahrnehmen und angemessen darauf reagieren können. Erfahrene Lehrpersonen weisen eine verbesserte selektive Aufmerksamkeit, stärkere Erkennung von Zusammenhängen sowie effizientere und schnellere Informationsverarbeitung auf (Huang et al. 2023, S. 132). Zudem zeigen die Studien zur Blickanalyse, dass erfahrene Lehrkräfte ihren Blick gleichmäßiger über alle Lernenden hinweg richten können (Keskin et al. 2024, S. 20). Dadurch fühlen sich nicht nur alle Schüler:innen gesehen und ernst genommen, sondern die Lehrperson kann früher Unterstützungsbedarfe oder Motivationsprobleme wahrnehmen. Jacobs et al. (2010, S. 194) haben ebenfalls festgestellt, dass angehende Lehrkräfte im komplexen Klassenraumgeschehen Strategien der Lernenden weniger erfolgreich wahrgenommen haben als erfahrene Lehrpersonen. Sie erkennen deutlich besser die

Denkprozesse der Schüler:innen und können somit frühzeitiger Erklärungen finden sowie Vorhersagen für den weiteren Lernprozess treffen und gegebenenfalls eher intervenieren als Studierende.

Forschungen zum Knowledge-based Reasoning zeigen ebenfalls Unterschiede zwischen Noviz:innen und Expert:innen. Analysiert man die drei Prozesse des Reasonings, belegen Studien, dass Anfänger:innen am besten in der Kategorie des Beschreibens abschneiden, gefolgt vom Erklären und Vorhersagen (Stürmer et al. 2015, S. 352; Schäfer und Seidel 2015, S. 50–51; Meschede et al. 2015, S. 328; Vgl. Frommelt et al. 2016, S. 369). Es wird angenommen, dass für die anderen Kategorien wie das Erklären und Vorhersagen mehr Wissen über Lehren und Lernen erforderlich ist (Stürmer et al. 2013, S. 478), das erst mit zunehmender Erfahrung aufgebaut werden kann (Carter et al. 1988, S. 31). Dies bedeutet, dass Anfänger:innen möglicherweise Schwierigkeiten haben, komplexe Zusammenhänge zu erkennen und angemessen zu erklären oder vorherzusagen. Zudem ist für Studierende die Herausforderung groß, die gelernten Theorien an praktischen Beispielen anzuwenden (Ertl-Schmuck 2019, S. 48). Es fällt ihnen schwerer, ihr fachliches und pädagogisches Wissen in den konkreten Unterrichtssituationen anzuwenden. Im Vergleich zu Expert:innen fällt innerhalb des Reasonings bereits beim Beschreiben auf, dass Anfänger:innen nicht auf dem gleichen Niveau sind (Seidel und Prenzel 2007, S. 211–213; Schäfer und Seidel 2015, S. 49; Seidel et al. 2010, S. 297–298). Sie treffen häufiger naive Annahmen mit weniger fachlicher Fundiertheit (Schäfer und Seidel 2015, S. 48). Dies könnte ebenfalls auf den anspruchsvollen Theorie-Praxis-Übertrag zurückzuführen sein. Expert:innen verfügen über ein tieferes Verständnis des Fachgebiets und sind damit möglicherweise besser in der Lage, ihre Entscheidungen und Einschätzungen auf solides Fachwissen zu stützen. Dies hat Konsequenzen für das wissenschaftsbasierte Erklären von Unterrichtssituationen. Es ist abhängig vom Professionswissen der Lehrperson, „das sich unter anderem aus fachdidaktischem Wissen, Erfahrungswissen sowie Fachwissen zusammensetzt“ (Oser et al. 2010, S. 24). Sie nutzen ihr Wissen, um komplexe Phänomene im Klassenzimmer zu verstehen und zu erklären (Oser et al. 2010, S. 24; Carter et al. 1988, S. 31). Im Vergleich dazu bleiben die Noviz:innen eher beim Beschreiben der Situation ohne tiefere Interpretation (Carter et al. 1988, S. 31). Konkrete Studienergebnisse zu den genannten Befunden sind im Abschnitt 4.2 *Studienlage* dargestellt.

3.4.3 Zusammenfassung in Bezug auf die vorliegende Erhebung

Die Professionelle Unterrichtswahrnehmung trägt mit ihren interagierenden Teilprozessen des Noticings und Reasonings entscheidend zum erfolgreichen pädagogischen Handeln von Lehrpersonen bei (Schäfer und Seidel 2015). Die Erfassung des Noticings gelingt besonders gut durch Eye-Tracking-Methoden (Vgl. Metaanalyse

Keskin et al. 2024siehe Abschnitt 3.4.1). Beim Reasoning eignen sich zur Analyse der Fähigkeiten Videoformate (Vgl. Diagnostische Fähigkeiten, Bartel und Roth 2017, S. 47), die zudem in der Lehrpersonenbildung zur Steigerung der Motivation (Syring et al. 2015, S. 682) und für einen erhöhten Anwendungsbezug eingesetzt werden können (Ertl-Schmuck 2019, S. 48–49). Auf Grund der komplexen Umsetzung eines Vorhabens, mit dem beide Teilprozesse erhoben werden können, fokussiert sich die vorliegende Arbeit auf den Reasoning-Prozess mit den drei Teilbereichen des Beschreibens, Erklärens und Vorhersagens. Zudem bietet das an der Universität Koblenz – Landau entwickelte Videotool ViviAn „Videovignetten zur Analyse von Unterrichtsprozessen“ eine passende Lernumgebung. Eine Erweiterung der Vignetten und Analyseaufträge in ViviAn ist nicht nur aus ökonomischer Sichtweise sinnvoll, sondern auch, um den Studierenden zusätzlich zu fachspezifischen Themen übergeordnete Inhalte beispielsweise zum Thema Motivierung anbieten zu können.

Anhand der Expertiseforschung handelt es sich bei der Professionellen Unterrichtswahrnehmung um eine Fähigkeit, die sich Lehrpersonen über Jahre angeeignet haben. Diese Tatsache spricht dafür, dass die Professionelle Unterrichtswahrnehmung eine wichtige Fähigkeit im Alltag einer Lehrperson zu sein scheint. Das komplexe und dynamische Unterrichtsgeschehen (Oser et al. 2010, S. 6), stellt eine hohe Herausforderung vor allem für Noviz:innen im Lehrberuf dar (Seidel und Prenzel 2007, S. 211–213; Schäfer und Seidel 2015, S. 49; Seidel et al. 2010, S. 297–298). Lehrpersonen sind gefordert, die relevanten Situationen richtig wahrzunehmen, zu verarbeiten und adäquat Auswirkungen zu beurteilen (Seidel et al. 2010; Sherin 2009, S. 384–385). Die Fähigkeit zur systematischen Analyse sollte bereits im Studium gefördert werden (Santagata und Yeh 2014, S. 493). Wenn frühzeitig damit begonnen werden kann, die Studierenden auf diese Herausforderung vorzubereiten, kann sich dies positiv auf den Unterricht, die Motivation und den Lernerfolg der Schüler:innen auswirken.

Ziel der vorliegenden Arbeit soll es daher sein, ein geeignetes Instrument zu entwickeln, mit dem die Professionelle Unterrichtswahrnehmung hinsichtlich der Motivierung erfasst werden kann. Aus der Forschungslage heraus hat sich bestätigt, dass videobasierte Instrumente dafür besonders geeignet scheinen (u.a. Frommelt et al. 2016, S. 365; Gold und Hasselhorn 2013, S. 149; Hellermann et al. 2015, S. 104; Santagata und Guarino 2011, S. 142–143; Sunder et al. 2016, S. 9). Damit gelingt es auch, das komplexe Unterrichtsgeschehen zu erfassen und für den Einsatz im Studium anwendbar zu machen (van Es 2009, S. 101). Inwieweit sich die bisherigen Ergebnisse zur Forschung replizieren lassen, muss überprüft werden.

4 Einsatz von Videos in der Lehre

Im ersten Abschnitt dieses Kapitels (4.1. *Die fünf Heuristiken*) werden die von Blomberg et al. (2013) zusammengefassten fünf wesentlichen Aspekte für den Einsatz von Videos in der Lehre vorgestellt. Dabei wird unter anderem auf den Nutzen der Videos, verschiedene Arten von Vignetten sowie die Limitationen eingegangen. Zahlreiche Forschungsarbeiten belegen, dass die Professionelle Unterrichtswahrnehmung von Lehramtsstudierenden durch den Einsatz von Videos gemessen und gefördert werden kann (siehe Abschnitt 4.2 *Studienlage*). Krammer (2014, S. 172) stellt in ihrem Übersichtsartikel heraus, dass sich neben fachspezifischen Thematiken auch fächerübergreifende Qualitätsmerkmale von Unterricht für die Verwendung von Videos eignen. Dementsprechend eignet sich der Schwerpunkt Motivierung, wie im vorliegenden Beitrag gezeigt wird, gut für die Entwicklung von Videovignetten. Diese Vignetten werden in das bereits bestehende Videotool *ViviAn* der RPTU (siehe Abschnitt 4.3 *Das Videotool ViviAn*) eingebettet. Das Videotool basiert auf Unterrichtsvignetten, die kurze Ausschnitte aus dem Schulunterricht bzw. aus Gruppenarbeiten zeigen. Damit bilden sie den Unterrichtsalltag „weitgehend ganzheitlich und authentisch“ (Riegel 2013, S. 13) ab. In Kombination mit Analyseaufträgen wird den Studierenden ein videobasiertes Testinstrument angeboten, mit dem sie ihre Analysefähigkeiten an realen Unterrichtssituationen überprüfen können (Rehm und Bölsterli 2014, S. 224). Abschließend folgt ein Resümee des vierten Kapitels *Einsatz von Videos in der Lehre* mit abgeleiteten Konsequenzen für die (Weiter-)Entwicklung des vorliegenden Videoinstruments (siehe Abschnitt 4.4 *Zusammenfassung in Bezug auf die vorliegende Erhebung*).

4.1 Die fünf Heuristiken

Verschiedene Forschungsarbeiten zum Einsatz von Videos in der Lehramtsausbildung haben Blomberg et al. (2013, S. 92) zusammengetragen und fünf Heuristiken zur Verwendung von Videos in einer gut konzipierten Lernumgebung abgeleitet. Diese fünf Heuristiken sind (Blomberg et al. 2013, S. 95)

- 1) Lernziele identifizieren
- 2) Lehransatz wählen,
- 3) Videomaterial wählen,
- 4) Limitationen ansprechen,
- 5) Assessment an den Unterricht/ an die Zielen anpassen.

Sie spielen auch für die Entwicklung des Videotools dieses Beitrags eine Rolle. Daher werden sie im Folgenden näher betrachtet. Die Reihenfolge der fünf Heuristiken wird für eine stringente Argumentation angepasst.

Lehransatz

Mit dem Einsatz von Videos kann es gelingen, vermittelte Theorien und Konzepte praxisnah und anwendungsbezogen mit dem Alltag in den Schulen zu verknüpfen (Blomberg et al. 2013, S. 106; Reusser 2005, S. 10; Bartel und Roth 2017, S. 47). „Die Verbindung von Theorie und Praxis gilt als eine der zentralen Herausforderungen der Lehrerinnen- und Lehrerbildung, da die Gestaltung eines lernwirksamen Unterrichts eine situationsbezogene Anwendung von fachlichem, fachdidaktischem und pädagogisch-psychologischem Wissen und Können verlangt“ (Krammer 2014, S. 164). Bisherige Forschungen können eine hohe Verzahnung zwischen Theorie und Praxis mit Herstellung eines direkten Anwendungsbezugs durch den Einsatz von Videos belegen (Blomberg et al. 2013, S. 106). Dieser Lehransatz stellt eine innovative, empirisch abgesicherte und etablierte Form des videofallbasierten Lernens dar (Digel et al. 2014, S. 40). Die videografierten Fälle bilden dabei die Praxis ab, die es mit den theoretischen Grundlagen zu verknüpfen gilt (von Aufschnaiter et al. 2017, S. 88). Dies gelingt durch die hohe Authentizität und Realitätsnähe, die Videos vermitteln (Syring et al. 2015, S. 670; Yadav et al. 2011, S. 32). Durch die Ergänzung mit Analyseaufträgen können die Videofälle als Testinstrument verwendet werden (von Aufschnaiter et al. 2017, S. 86). Damit bleibt es nicht nur beim Beobachten einer Unterrichtssequenz, sondern es wird die Anwendung und Überprüfung des vorhandenen theoretischen Wissens und der Einbezug der Perspektive der Lernenden ermöglicht (Krammer und Reusser 2005, S. 47). Die videofallbasierte Arbeit kommt vielseitig in der Aus- und Weiterbildung für Lehrpersonen zur Anwendung. Sie trägt dazu bei, die Analyse- und Diagnosefähigkeiten bei Lehrkräften mit geringerer, aber auch mehr Berufserfahrung zu fördern (Digel 2013, S. 7). Von besonderem Vorteil für den Einsatz am Beginn der Lehramtsausbildung zählt die erhöhte wahrgenommene Relevanz der theoretischen Inhalte, die für das Begreifen von komplexen Unterrichtssituationen bedeutsam ist (von Aufschnaiter et al. 2017, S. 89). Es konnte gezeigt werden, dass die wahrgenommene Relevanz der Arbeit mit Videovignetten für das zukünftige Unterrichten von Lehramtsstudierenden im Vergleich zur Arbeit mit Transkriptvignetten signifikant höher beurteilt wird (Bartel und Roth 2020, S. 313). Dies fördert auch das Gefühl der Involviertheit und führt zur einer höheren Bereitschaft, Videos zu analysieren (Syring et al. 2015, S. 681). Zusätzlich ist eine positive, freudige Gefühlslage bei der Arbeit mit Videofällen zu verzeichnen, wohingegen die Arbeit mit Texttranskripten häufiger zu Ärger beim Lernen führt (Syring et al. 2015, S. 681). Während Yadav et al. (2011, S. 32) insgesamt eine höhere Motivation in der Arbeit mit Videos im Vergleich zu Texten belegen können, stellen Syring et al. (2015, S. 681) zwar einen Anstieg der Motivation bei der Arbeit mit Fällen fest, aber keinen Unterschied zwischen den beiden Formaten. Insgesamt können bisherige Forschungen eine hohe Akzeptanz bezüglich der Arbeit mit Videos in der Aus- und Weiterbildung belegen

(Seidel et al. 2010, S. 304–305; Digel 2013, S. 7; Altmann und Kändler 2015, S. 49). Damit stellt dieser Lehransatz ein geeignetes Medium in der Lehramtsausbildung dar. Aber auch im Weiterbildungssektor für Lehrende ist es ein Konzept, das „die Entwicklung grundlegender Fähigkeiten professionellen pädagogischen Handelns thematisiert und flexibel in bestehende oder neu zu konzipierende Weiterbildungen integrierbar ist“ (Digel et al. 2014, S. 40).

Lernziele

Der Vorteil von Videos ist es, komplexe Unterrichtssituationen mit vielen Facetten anschaulich abbilden zu können (Helmke 2015, S. 342; Reusser 2005, S. 10; van Es 2009, S. 101; Meschede und Steffensky 2018, S. 22). Sie tragen einen hohen Informationsgehalt (von Aufschnaiter et al. 2017, S. 95; Riegel und Macha 2013, S. 13) und können zudem „nonverbale Bearbeitungsprozesse, Erlebensäußerungen oder soziale Dynamiken“ (von Aufschnaiter et al. 2017, S. 95) festhalten. Die Videos bieten den Studierenden zum einen die Möglichkeit, die Beobachtungsfähigkeit zu schulen sowie Aufmerksamkeit zu fokussieren (von Aufschnaiter et al. 2017, S. 88), zum anderen realistische Sequenzen zu analysieren und Konsequenzen für das eigene Handeln abzuleiten (Reusser 2005, S. 10; Meschede und Steffensky 2018, S. 22). Dies ermöglicht das Medium Video durch die Distanz zum eigentlichen Unterrichtsgeschehen und die Verringerung des situationsabhängigen Drucks (Krammer und Reusser 2005, S. 36). Der Einsatz von Videos, beispielsweise im Rahmen von online basierten Videotools wie dem Observer (Seidel et al. 2010), ist im Vergleich zu schriftlichen Unterrichtsbeispielen zur Förderung der Professionellen Unterrichtswahrnehmung besser geeignet (Frommelt et al. 2016, S. 365). Neben Krammer (2014) haben auch Steffensky und Kleinknecht (2016) in ihrem Übersichtsartikel die Förderung der analytischen Fähigkeiten mittels Videotools aufzeigen können. In mehreren Interventionsstudien konnte ein positiver Effekt sowohl bei Noviz:innen als auch bei Lehrpersonen nachgewiesen werden (Steffensky und Kleinknecht 2016, S. 308). Auch Hellermann et al. (2015, S. 103) können belegen, dass die Professionelle Unterrichtswahrnehmung durch die Bearbeitung von Videos zunimmt. Ihre Forschung widmet sich dabei dem Aspekt der Klassenführung. Konkret bedeutet dies die Förderung der wissensbasierten Verarbeitung durch das Wahrnehmen relevanter Situationen im Unterricht, die Interpretation dieser bedeutsamen Lernaspekte und das Ableiten von Konsequenzen sowie Handlungsalternativen (Meschede und Steffensky 2018, S. 23). Entsprechend des theoretischen Hintergrunds und der Auswahl der Videosequenzen können diverse Lernziele formuliert werden. Ein videofallbasierter Lehransatz benötigt stets konkrete Lernziele, um die Videos adäquat und zielorientiert einsetzen zu können (Blomberg et al. 2013, S. 97).

Videomaterial

Bischoff et al. (2005, S. 385) setzen bei der Entwicklung ihres videobasierten Testinstruments auf authentische Unterrichtssequenzen, die durch Instruktionen erweitert werden, um vielseitige und vor allem auch problematische Handlungen zur Analyse bereitstellen zu können. Eine derartige „*modifizierte Vignette*“ (von Aufschnaiter et al. 2017, S. 96) wird in einer realen Situation aufgezeichnet, aber für die Erreichung der Lernziele angepasst (von Aufschnaiter et al. 2017, S. 96). Handelt es sich bei den Vignetten um Videomaterial, das speziell für die Situation oder das Lernziel geplant und entwickelt wird, werden diese als „*konstruierte Vignette*“ (von Aufschnaiter et al. 2017, S. 96) bezeichnet. Der Beweggrund für die Entwicklung modifizierter oder konstruierter Vignetten ist meist das seltene Beobachten solch konkret erwarteter Handlungen oder Situationen im Unterricht (Meschede und Steffensky 2018, S. 25). Ebenso bieten sie die Möglichkeit, kritische, aber auch ethisch bedenkliche Beispiele zu zeigen (Meschede und Steffensky 2018, S. 25). Im Gegensatz dazu hat eine „*reale [...] Vignette*“ authentische Szenen zum Gegenstand (von Aufschnaiter et al. 2017, S. 96). Videovignetten können im Speziellen ausschließlich best-practice Beispiele abbilden, die vor allem zur „Veranschaulichung und Modellierung qualitativ hochwertiger Praxis“ (Meschede und Steffensky 2018, S. 26) verwendet werden. Zudem können die Vignetten hinsichtlich der Herkunft zwischen eigenem und fremdem Material unterschieden werden (Blomberg et al. 2013, S. 100). Wenn die Beobachter:innen der Videos an den gefilmten Situationen nicht selbst teilnehmen, handelt es sich um fremde Vignetten (von Aufschnaiter et al. 2017, S. 96). Dem gegenüber stehen eigene Videos, mit denen der eigene Unterricht analysiert wird (Blomberg et al. 2013, S. 100). Es zeigt sich, dass sowohl eigene als auch fremde Videos in der Lehramtsbildung sinnvoll eingesetzt werden können (Krammer 2014, S. 167; Steffensky und Kleinknecht 2016, S. 313) und hinsichtlich der Förderung der Professionellen Unterrichtswahrnehmung die gleichen Effekte erzielen (Frommelt et al. 2016, S. 366). Dennoch können Unterschiede zwischen den Formaten eigene versus fremde Videos wahrgenommen werden. Krammer (2014, S. 167) sowie Steffensky und Kleinknecht (2016, S. 313) weisen darauf hin, dass laut Studienlage eigene Videos ein größeres Authentizitätsempfinden erzeugen und die Auseinandersetzung mit dem eigenen Videomaterial motivierender ist. Allerdings kann das Beobachten der eigenen Person oberflächlicher sein oder bei problematischen Handlungen zu einem negativen Selbstwertgefühl führen (Steffensky und Kleinknecht 2016, S. 313). Der Einsatz fremder Videos bewirkt eher eine kritische, distanziertere und tiefgründige Analyse von Unterrichtshandlungen (Krammer 2014, S. 168; Steffensky und Kleinknecht 2016, S. 313). Um genau diese Analysefähigkeit messen und fördern zu können, kann es hilfreich sein, Material einer unbekanntem Lehrperson zu verwenden. Zudem wird „die Reflexion des eigenen Handelns in Bezug auf

Qualitätsstandards unterstützt, im Falle kooperativer Unterrichtsreflexion auch in der Konfrontation mit anderen Sichtweisen auf das Geschehen“ (Pauli und Reusser 2006, S. 792). Speziell ausgewählte Unterrichtssequenzen können optimal in der Gruppe diskutiert werden (Meschede und Steffensky 2018, S. 26). Dabei erlaubt fremdes Filmmaterial eine effiziente Verwendung von Videos in der Lehre. Das Produzieren eigener Videos bedeutet einen enorm hohen zeitlichen und wenig ressourcenschonenden Aufwand. Nichtsdestotrotz wäre eine Kombination aus eigenen und fremden Videos denkbar und laut Hellermann et al. (2015, S. 103) noch effektiver. Ein weiterer Aspekt der Aufbereitung des Videomaterials ist die Bestimmung der Ausschnittlänge (Meschede und Steffensky 2018, S. 27). Das Aufzeichnen ganzer Unterrichtseinheiten bietet den Vorteil der Betrachtung des gesamten Kontextes sowie die Möglichkeit, eine Entwicklung aufzuzeigen (Meschede und Steffensky 2018, S. 26). Demgegenüber stehen kurze Sequenzen für eine Fokussierung auf ausgewählte Aspekte (Meschede und Steffensky 2018, S. 26).

Assessment an die Ziele anpassen

Der Einsatz von Videos in der Aus- und Weiterbildung von Lehrkräften hat den positiven Effekt, Situationen aus dem Unterrichtsalltag leicht und vielfach zugänglich zu machen (van Es 2009, S. 101). In Kombination mit einer digitalen Videoplattform, wie unter anderem Kersting et al. (2010) in ihrer Studie verwenden, können weitere Vorteile wie zum Beispiel die einmalige Aufbereitung der Videos mit Untertiteln, das Einbetten von zusätzlichen Informationen sowie das sichere Abspielen der Videos durch einen Passwortschutz genutzt werden (Kersting et al. 2010, S. 174). Das Bereitstellen von Zusatzinformationen, über die eine Lehrperson im Unterrichtsalltag zugleich verfügt, unterstützt „eine zufriedenstellende lernprozessbezogene Diagnose des beobachteten Prozesses“ (Bartel und Roth 2017, S. 45). Dadurch können fehlende Informationen des Kontextes wie das Alter der Lernenden oder eine Einordnung der Situation in den Gesamtprozess zur Verfügung gestellt werden (von Aufschnaiter et al. 2017, S. 95).

Zur Erreichung der Lernziele, können die Videos mit passenden Analyseaufträgen ergänzt werden. Grundsätzlich müssen bei den Aufträgen zwischen „die *Anregung des Kompetenzaufbaus* („Lernpotential“) und die *Erfassung vorliegender Kompetenzen* („Testpotential“)“ (von Aufschnaiter et al. 2017, S. 91) voneinander unterschieden werden. Bei der Konstruktion der Aufträge können verschiedene Aufgabenformate gewählt werden. Die Entscheidung für geschlossene, fokussierte oder offene Aufgaben (von Aufschnaiter et al. 2017, S. 93) sollte entsprechend der Lernziele sowie Vor- und Nachteilen der einzelnen Formate gut begründet sein. Geschlossene Aufgaben werden im Zusammenhang mit der Förderung der Professionellen Unterrichtswahrnehmung unter anderem beim OBSERVER (Seidel et al. 2010, S. 299–301), Star und

Strickland (2008, S. 115) oder Meschede et al. (2015, S. 324) verwendet. Da sie nur eine eindeutige Lösung (z.B. in Form von Rating-Items oder Multiple Choice Antworten) zulassen, strukturieren geschlossene Aufgaben stark vor und schränken die Information über das Testpotential der Teilnehmenden gegebenenfalls sehr ein (von Aufschnaiter et al. 2017, S. 93). Dennoch kommen sie aufgrund der hohen Auswertungsobjektivität (Eid und Schmidt 2014, S. 66) und Grad der Standardisierung häufig zum Einsatz. Offene Aufgaben wie „What did you notice?“ (Sherin 2009, S. 386) müssen zunächst mittels verschiedener Verfahren wie z.B. die Qualitative Inhaltsanalyse aufwendig codiert werden (Züll und Menold 2014, S. 715–716). Sie bieten allerdings den Vorteil verschiedene inhaltliche Aspekte oder Lösungswege der befragten Personen aufzuzeigen (von Aufschnaiter et al. 2017, S. 93). Zudem erfordern sie höhere kognitive Fähigkeiten (Züll und Menold 2014, S. 715) und geben dementsprechend eine bessere Auskunft über die tatsächlichen Kenntnisse. Eine Kombination aus beiden bisherigen Formaten stellen die fokussierten Aufgaben dar. Sie lenken die Aufmerksamkeit bereits auf einen spezifischen Aspekt, lassen dennoch verschiedene Lösungsmöglichkeiten zu (von Aufschnaiter et al. 2017, S. 93).

Limitationen

Der Einsatz von Videos in der Lehramtsausbildung geht mit einigen Limitationen einher. Dies betrifft zum einen das Videomaterial, zum anderen das methodische Vorgehen bei der Entwicklung und Validierung der Videos und Analyseaufträge.

Eine grundlegende Anforderung an Vignetten stellt die Bild- und Tonqualität dar (von Aufschnaiter et al. 2017, S. 97). Sobald es Einschränkungen hinsichtlich des Audio- oder Videomaterials gibt, kann die Analysefähigkeit der beobachtenden Personen eingeschränkt sein. Im Fall eines Tonproblems kann mit eingblendeten Untertiteln gearbeitet werden (von Aufschnaiter et al. 2017, S. 97). Der Kamerafokus kann so gewählt werden, dass ein bestimmter Ausschnitt (wie z.B. die Lehrperson oder Gruppenarbeitsprozesse) explizit betrachtet oder die Situation möglichst komplex (z.B. mit einer 360°Kamera) erfasst werden kann (Meschede und Steffensky 2018, S. 27). Dennoch können die Videos trotz einer Bild-in-Bild-Perspektive oder 360°-Perspektive, die sowohl die Lehrperson als auch die Schüler:innen zeigt, nicht alles abbilden. „Ferner definiert die Kameraperspektive die Erkenntnismöglichkeiten, denn sie zeichnet nur einen Ausschnitt des Unterrichts aus“ (Riegel 2013, S. 13). Die vorhandene Distanz zur Situation kann es schwierig machen, Kontextinformationen (Meschede und Steffensky 2018, S. 30) wie z.B. den Einfluss eines in der nächsten Stunde anstehenden Tests oder parallel auftretende Ereignisse (Meschede und Steffensky 2018, S. 22) wahrzunehmen. Dieser Limitation kann durch das Einbetten von Kontextinformationen (von Aufschnaiter et al. 2017, S. 95) und das mehrfache

Ansehen der Videos aus unterschiedlicher Perspektive (Meschede und Steffensky 2018, S. 22) zumindest teilweise entgegengewirkt werden.

Die bereits angesprochene Möglichkeit der komplexen und authentischen Darstellung der Situation im Video, kann zu einer Überforderung der Studierenden führen (von Aufschnaiter et al. 2017, S. 95; Syring et al. 2015, S. 670). Eine kognitive Belastung kann durch simultane Ereignisse im Video und die allgemein hohe Komplexität auftreten (Syring et al. 2015, S. 680). Die Arbeit mit Videos führt bei den analysierenden Personen aber auch zu einem besseren Eintauchen in die virtuelle Umgebung und zu mehr Freude, was den negativen Aspekt der höheren Belastung im Vergleich zu Textbeispielen mindern kann (Syring et al. 2015, S. 682). Textvignetten werden wegen der geringeren kognitiven Belastung eher für eine tiefere Analyse fokussierter (bzw. reduzierter) Probleme oder Handlungen eingesetzt (Syring et al. 2015, S. 670). Gegen Textvignetten spricht allerdings der geringere Realitätsbezug aufgrund der geringeren Komplexität (Syring et al. 2015, S. 682). Yadav et al. (2011, S. 32) können in ihrer Untersuchung zudem keinen Unterschied zwischen Text- und Videoformat hinsichtlich der kognitiven Verarbeitung feststellen.

Beim Betrachten von Videos muss man sich der Urteilsverzerrungen bewusst sein. Als Beispiel dienen zwei Effekte. Der Primacy-Effekt kann das Analysieren folgender Sequenzen durch den gewonnenen Ersteindruck beeinflussen (Meschede und Steffensky 2018, S. 30). Möglich ist auch die verzerrte Wahrnehmung der Situation aufgrund einer Sympathieentwicklung für die im Video handelnde Lehrperson (Halo-Effekt) (Meschede und Steffensky 2018, S. 30).

Unterrichtsvideos bieten neben den Limitationen eine Vielzahl von Vorteilen für den Einsatz in der Lehramtsausbildung. Es wird angenommen, dass den Studierenden eine bessere Transferleistung für das eigene Unterrichten durch die Analyse komplexer und authentischer Videos gelingt (Syring et al. 2015, S. 670). Wichtig bleibt zu berücksichtigen, dass ein Lerneffekt erst durch sinnvolle Analyse- oder Arbeitsaufträge einsetzen kann (Pauli und Reusser 2006, S. 793). Durch die Verwendung geschlossener Antwortformate wie in der vorliegenden Dissertationsarbeit wird der Aufmerksamkeitsfokus vorstrukturiert und kann deswegen die Fähigkeiten der Studierenden überschätzen (von Aufschnaiter et al. 2017, S. 93). Allerdings bleibt der Vorteil der hohen Standardisierung in der Auswertung. Zu einem hohen Standardisierungsgrad zählt ebenfalls die Verwendung von Expertenratings zur Erstellung einer Musterlösung (siehe Abschnitt 7.5 *Validierung der Analyseaufträge durch Expertenratings*). Die Validität dieser Expertenurteile bleibt kritisch zu hinterfragen (Altmann und Kändler 2015, S. 60). Ein Vergleich zwischen verschiedenen videobasierten Studien hinsichtlich des Standardisierungsgrades stellt Lindmeier (2013) her.

Eine Zusammenfassung der fünf Heuristiken in Bezug zur vorliegenden Forschungsarbeit wird im Abschnitt 4.4 dargelegt. Hier werden die einzelnen Aspekte und Entscheidungen für die Umsetzung des videobasierten Instruments zur

Förderung der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung begründet.

4.2 Studienlage

In diesem Abschnitt werden verschiedene Studien zum Einsatz von Videos in der Aus- und Weiterbildung von Lehrkräften vorgestellt. Das Videotool OBSERVER (Seidel et al. 2010) wird im deutschsprachigen Raum seit seiner Entwicklung vielfach in Studien eingesetzt. Es stellt damit ein bekanntes Tool zur Förderung der Professionellen Wahrnehmung dar. Der theoretische Bezugsrahmen zum Teilprozess Knowledge-based Reasoning mit den drei Bereichen Beschreiben, Erklären und Vorhersagen ist sowohl beim OBSERVER als auch beim vorliegendem Videotool ViviAn identisch. Hinsichtlich der Gestaltung der Analyseaufträge als vier-stufiges Rating und des methodischen Vorgehens, etwa durch den Einsatz von Expert:innenratings zeigen sich ebenfalls Parallelen mit der vorliegenden Arbeit. Um einen direkten Vergleich zwischen dem OBSERVER und ViviAn ziehen zu können bzw. Konsequenzen für die eigene Arbeit ableiten zu können, werden wesentliche Arbeiten zum OBSERVER mit ihren Ergebnissen separat im Abschnitt 4.2.1 *Studien zum OBSERVER* beschrieben. Weitere Forschungsarbeiten, die sich mit der Förderung der Professionellen Unterrichtswahrnehmung auseinandersetzen und wichtige Erkenntnisse zum Einsatz von Videos in der Lehre beitragen, folgen im Abschnitt 4.2.2 *weitere Studien zur Förderung der Professionellen Wahrnehmung*. Konsequenzen für die vorliegende Arbeit werden im Abschnitt 4.4 gezogen.

4.2.1 Studien zum OBSERVER

Das Videotool OBSERVER bietet den Nutzer:innen die Möglichkeit, verschiedene Vignetten aus dem Unterricht auf einer digitalen Videoplattform zu analysieren (Seidel et al. 2010, S. 300–301). Die verwendeten Rating-Items mit einer vier-stufigen Likert-Skala (von 1 = trifft zu bis 4 = trifft nicht zu) sind so konzipiert, dass sie die drei Facetten Beschreiben, Erklären und Vorhersagen des Knowledge-based Reasonings abbilden (Seidel et al. 2010, S. 301). Der Fokus der Professionellen Unterrichtswahrnehmung liegt beim OBSERVER auf den drei wahrzunehmenden (Noticing) lernrelevanten Komponenten Zielklarheit und -orientierung, Unterstützung und Führung durch die Lehrperson sowie Lernklima (Seidel und Stürmer 2014, S. 743). Der OBSERVER integriert damit die Facetten des Noticings und Reasonings in seinem Online-Tool (siehe Abbildung 12), wobei sich die Testung der Analysefähigkeit ausschließlich auf das Reasoning bezieht.

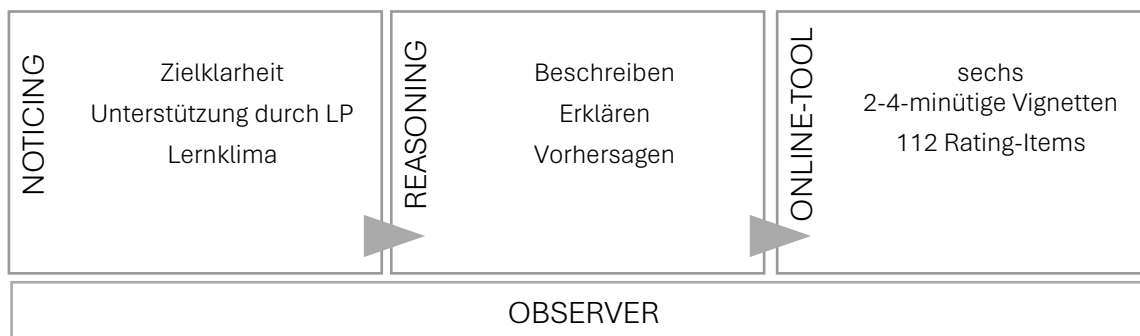


Abbildung 12 Das Videotool OBSERVER (Seidel et al. 2010) (eigene Darstellung)

Das Videotool besteht aus sechs zwei- bis vierminütigen Videoclips aus dem Unterricht der 8. und 9. Klassen und zusätzlichen Informationen zur Unterrichtssequenz, die vor der Betrachtung der Videos abgerufen werden können (Seidel und Stürmer 2014, S. 752). Zusätzlich erhalten die Nutzer:innen eine kurze theoretische Einführung zu den drei lernrelevanten Aspekten sowie zur Anwendung des Tools (Seidel et al. 2010, S. 300). Die Fragen zur Analyse der Professionellen Unterrichtswahrnehmung beziehen sich auf zwei der drei Noticing-Aspekte (siehe Tabelle 3, Beispielitems für das Lernklima) und erscheinen nach dem einmaligen Sehen des Videos (Seidel et al. 2010, S. 301).

Tabelle 3 Beispielitems aus dem OBSERVER (Seidel und Stürmer 2014, S. 751)

Beispielitems	
Beschreiben	„In the excerpt that you saw... the teacher is respectful of the students.“
Erklären	„In the excerpt that you saw... the students have the opportunity to feel that their teacher takes them seriously.“
Vorhersagen	„Based on what you saw... the teaching style will motivate the students.“

Die Auswahl der Videovignetten erfolgt im Rahmen einer Pilotierungsstudie mit sechs Expert:innen aus dem Bereich der Unterrichtsforschung (Seidel et al. 2010, S. 301). Als Eignung der Vignetten wird das Maß der Übereinstimmung zwischen den Expertiseurteilen ermittelt. Es kommt darauf an, inwiefern die Videoclips die entsprechenden Merkmale Zielklarheit, Unterstützung der Lehrperson und Lernklima repräsentieren (Seidel et al. 2010, S. 302). Die Expert:innen beurteilen die Passung der Vignetten für die Zielorientierung und Lernatmosphäre mit durchschnittlich 66,3%, die Vignetten der Lernbegleitung mit 48,8% (Seidel et al. 2010, S. 302).

In Folgestudien werden zwei Gruppen Lehramtsstudierende ($N_1 = 119$, $M_{Sem} = 4.63$, $SD_{Sem} = 3.98$; $N_2 = 152$, $M_{Sem} = 4.97$, $SD_{Sem} = 0.18$) (Seidel und Stürmer 2014, S. 754) ebenfalls zur Beurteilung der Vignetten hinsichtlich ihrer Passung herangezogen und

es wird die Übereinstimmung mit den Urteilen der Expert:innen aus der Pilotierung überprüft (Seidel und Stürmer 2014, S. 759). Die Ergebnisse zeigen aus beiden Studien eine hohe Übereinstimmung von 67% bis 80% (Seidel und Stürmer 2014, S. 759). Die Analysefragen beziehen sich wie bereits beschreiben auf die drei Teilprozesse des Reasonings. In einer Modelltestung wird anhand der Stichprobe N_2 diese Annahme der Dreidimensionalität getestet (Seidel und Stürmer 2014, S. 759). Anhand der Analyse der Infit- und Outfitwerte sowie eines Modellvergleichs mit einem ein- und zwei-dimensionalen Modell ergibt das dreidimensionale Modell mit finalen 112 dichotomen Items die beste Eignung (Seidel und Stürmer 2014, S. 759). Der Nachweis dieser dreidimensionalen Struktur kann anhand einer größeren Stichprobe ($N = 1029$, $M_{Sem} = 2.66$, $SD_{Sem} = 0.79$) repliziert werden (Jahn et al. 2014, S. 176). In einer weiteren Forschungsfrage wird die konsistente Messung der Professionellen Unterrichtswahrnehmung thematisiert und an einer Stichprobe von Lehramtsstudierenden aus dem sechsten und siebten Semester ($N = 20$) untersucht (Seidel und Stürmer 2014, S. 762). Anhand der kleinen Stichprobe lassen sich keine Interaktionseffekte hinsichtlich des wiederholten Einsatzes des Instruments nach 3, 7, 14 und 21 Tagen nachweisen ($F = 0.29$, $df = 2.36$, $p = 0.78$) (Seidel und Stürmer 2014, S. 762).

Die Arbeitsgruppe Jahn et al. (2014, S. 172) geht der Frage nach einer vergleichbaren Messung der Professionellen Unterrichtswahrnehmung unter Anwendung des OBSERVERS zwischen verschiedenen Lehramtsstudiengängen und universitären Standorten nach. Im Rahmen der DIF-Analyse in Bezug auf die Itemschwierigkeit deutet der Vergleich der drei Gruppen von Lehramtsstudiengängen ($N_1 = 166$ Studierende der Primarstufe, $N_2 = 671$ der Sekundarstufe, $N_3 = 171$ der Berufliche Schulen) auf Unterschiede hin ($\chi^2(206) = 6193.29$, $p = 0.00$) (Jahn et al. 2014, S. 176). Dabei weisen 25% der Items moderate bis kritische DIF auf (Jahn et al. 2014, S. 176). Mit allen drei Gruppen kann die dreidimensionale Modellstruktur unter Beachtung der Limitation bezüglich der Itemschwierigkeit vergleichbar gemessen werden (Jahn et al. 2014, S. 177).

Schäfer und Seidel (2015, S. 40) widmen sich der Förderung beider Facetten der Professionellen Unterrichtswahrnehmung, dem Noticing und dem Reasoning bei Lehramtsstudierenden. Dazu erhalten die Studierenden ($N = 109$, $M_{Sem} = 5$, $SD_{Sem} = 0.21$, ohne Vorerfahrung in der Beobachtung von Unterrichtsvideos) ein dreiminütiges Video einer 7. Klasse im Fach Physik aus dem OBSERVER mit einem offenen Arbeitsauftrag und 36 geschlossenen Rating-Items (Schäfer und Seidel 2015, S. 41–43). Die offene Fragestellung zur Beschreibung des Gesehenen (Noticing) ergibt nach qualitativer Analyse, dass die Studierenden in der Lage sind, relevante Aspekte zu erkennen (53%), jedoch die Aufmerksamkeit häufig auf irrelevante und oberflächliche Aspekte gelegt wird (Schäfer und Seidel 2015, S. 42). Die Auswertung der Rating-Items des Reasonings für die lernrelevanten Komponenten Zielklarheit und Lernklima deuten auf eher naive Annahmen (ca. 80%) als auf wissensbasierte

Urteile (ca. 20%) hin (Schäfer und Seidel 2015, S. 48). Ein Vergleich mit Urteilen von Expert:innen zeigt eine eher geringe inhaltliche Übereinstimmung (Zielklarheit 37%, Lernklima 26%) (Schäfer und Seidel 2015, S. 49). Bei Differenzierung der erreichten Übereinstimmung der Professionellen Wahrnehmung hinsichtlich des Beschreibens, Erklärens und Vorhersagens erreichen die Studierenden im Beschreiben das beste Ergebnis (40%), gefolgt vom Vorhersagen (29%) und Erklären (20%) (Schäfer und Seidel 2015, S. 50–51).

Eine angepasste Version des OBSERVERS zum OBSERVER Extended Research Tool stellen Stürmer und Seidel (2015, S. 56–57) vor. Die Neuauflage enthält eine Erweiterung der Vignettenanzahl von sechs auf zehn zwei- bis vierminütigen Videoclips (Stürmer und Seidel 2015, S. 56). Diese Vignetten decken nicht nur die drei bisherigen lernrelevanten Aspekte ab, sondern werden entsprechend des prozessorientierten Lehr-Lernmodells differenzierter betrachtet (siehe Abbildung 13) (Stürmer und Seidel 2015, S. 56–57).

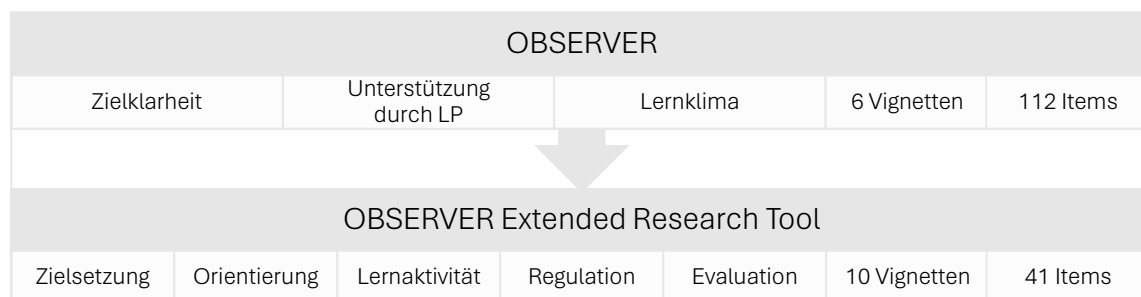


Abbildung 13 Erweiterung des OBSERVERS zum Extended Research Tool (Stürmer und Seidel, 2015) (eigene Darstellung)

Die Modellüberprüfung erfolgt jeweils eindimensional entsprechend der drei Teilprozesse Beschreiben, Erklären, Vorhersagen sowie als Gesamtskala anhand einer Stichprobe von Studierenden ($N_1 = 141$) und Referendar:innen ($N_2 = 176$) (Stürmer und Seidel 2015, S. 58–59). Im Vergleich zum OBSERVER zeigt die erweiterte Version eine geringere, noch akzeptable Reliabilität (siehe Tabelle 4), aber eine gute Varianz der Personenfähigkeitswerte (Stürmer und Seidel 2015, S. 60).

Tabelle 4 Vergleich der EAP-Reliabilitäten zwischen dem OBSERVER und der erweiterten Version (Vgl. Stürmer und Seidel 2015, S.60)

	OBSERVER	OBSERVER Extended Research Tool
	EAP-Reliabilität	EAP-Reliabilität
Beschreiben	0.90	0.60
Erklären	0.91	0.65
Vorhersagen	0.97	0.67
Reasoning	0.96	0.87

Eine DIF-Analyse zwischen den Studierenden und Referendar:innen indiziert eine vergleichbare Messung der Professionellen Unterrichtswahrnehmung (Stürmer und Seidel 2015, S. 60). Eine gezielte Betrachtung der Referendar:innen ($N = 173$) mit dem Extended Research Tool kommt zu dem Ergebnis, dass die höchste Fähigkeit im Beschreiben (im Durchschnitt 44% korrekte Antworten, $M = 0.44$, $SD = 0.20$) liegt, gefolgt vom Erklären ($M = 0.31$, $SD = 0.22$) und Vorhersagen ($M = 0.28$, $SD = 0.22$) (Stürmer et al. 2015, S. 352). Innerhalb der Gesamtstichprobe der Referendar:innen zeigen sich deutliche Unterschiede in der Fähigkeit, sodass diese entsprechend ihrer erreichten Punktzahl in vier Gruppen klassifiziert werden können (siehe Tabelle 5) (Stürmer et al. 2015, S. 352–353).

Tabelle 5 Klassifikation der Referendar:innen nach Fähigkeit ($N =$ Stichprobengröße, $M =$ Mittelwert, $SD =$ Standardabweichung) (Vgl. Stürmer et al. 2015, S.352-353)

	Gruppe 1 ($N = 34$)		Gruppe 2 ($N = 60$)		Gruppe 3 ($N = 47$)		Gruppe 4 ($N = 32$)	
	M	SD	M	SD	M	SD	M	SD
Beschreiben	0.20	0.11	0.38	0.11	0.53	0.12	0.69	0.13
Erklären	0.05	0.06	0.22	0.11	0.40	0.10	0.62	0.14
Vorhersagen	0.03	0.04	0.18	0.09	0.38	0.11	0.59	0.14
Einteilung nach Gesamtfähigkeit	geringe Fähigkeiten		niedrige-mittlere Fähigkeiten		mittlere-hohe Fähigkeiten		hohe Fähigkeiten	

Zusätzlich erheben Stürmer et al. (2015, S. 350) das Interesse am Thema Lehren und Lernen sowie die Einschätzung der theorie-praxisintegrierenden OTLs (opportunities to learn) im Referendariat. Die Gruppe mit den geringen Fähigkeiten weist dabei die niedrigsten Werte in den beiden erhobenen Skalen auf und zeigt einen signifikanten Unterschied zur Gruppe mit mittleren bis hohen Fähigkeiten im Interesse ($F(3, 168) = 3.00$, $p = 0.03$, $\eta^2 = 0.05$) sowie in einem der drei erfassten Merkmale der Einschätzung der theorie-praxisintegrierenden OTLs „Kommunikation mit ihren Kollegen“ ($F(3, 169) = 2.90$, $p = 0.04$, $\eta^2 = 0.05$) (Stürmer et al. 2015, S. 354–355). Die zwei weiteren Merkmale der OTLs „Informationelle[n] Unterstützung durch den Seminarleiter“ und „Teilen von Erfahrungen im Seminar“ scheinen nicht als Indikatoren für eine geringe Fähigkeit in der Professionellen Wahrnehmung zu dienen (Stürmer et al. 2015, S. 355). Zum einen zeigt diese Erhebung, dass ein Großteil der Referendar:innen in der Lage ist, Unterrichtssituationen mittelmäßig bis gut wissensbasiert verarbeiten zu können (Reasoning) (Stürmer et al. 2015, S. 356). Zum anderen wird deutlich, dass eine Gruppe mit besonders geringer Ausprägung der Professionellen Wahrnehmung, möglicherweise durch ein niedriges Interesse am Lehren und Lernen sowie einer geringeren Beurteilung der Kommunikation mit Kolleg:innen, frühzeitig

identifiziert werden kann (Stürmer et al. 2015, S. 356). Dadurch ist es möglich, eine entsprechende Förderung zu initiieren.

Zusammenfassend sprechen die Forschungsergebnisse zur Anwendung des OBSERVERs für ein valides Instrument, dass hinsichtlich der Konstruktvalidität, Reliabilität und Objektivität für die Erfassung der Professionellen Unterrichtswahrnehmung hinsichtlich der drei lernrelevanten Aspekte geeignet scheint. Die wesentlichen Erkenntnisse aus den Projekten sind:

- 1) Zur Überprüfung der inhaltlichen Eignung der Vignetten belegen die Urteile der Expert:innen (49% bis 66%) und der Studierenden (67% bis 80%) eine akzeptable bis gute Passung (Seidel et al. 2010, S. 302; Seidel und Stürmer 2014, S. 759).
- 2) Die Konstruktvalidität hinsichtlich der Modellstruktur ergibt eine dreidimensionale Struktur (Seidel und Stürmer 2014, S. 759; Jahn 2014, S. 176). Es kann zwischen Beschreiben, Erklären und Vorhersagen unterschieden werden. Die Reliabilität der drei Dimensionen ist hoch ($\alpha = 0.91-0.97$) (Stürmer und Seidel 2015, S. 60). Die erweiterte Version OBSERVER Extended Research Tool erreicht eine akzeptable bis gute Reliabilität ($\alpha = 0.65-0.87$) (Stürmer und Seidel 2015, S. 60).
- 3) An einer kleinen Stichprobe kann eine konsistente Messung der Professionellen Unterrichtswahrnehmung nachgewiesen werden (Seidel und Stürmer 2014, S. 762).
- 4) Der OBSERVER kann zwischen den drei Lehramtsstudiengängen Primar-, Sekundarstufe und Berufliche Schule an verschiedenen Standorten keine vergleichbare Messung der Professionellen Wahrnehmung erzielen (Jahn 2014, S. 176). Es zeigen sich Unterschiede hinsichtlich der Itemschwierigkeit (Jahn et al. 2014, S. 176). Dahingegen können Stürmer und Seidel (2015, S. 60) eine vergleichbare Messung zwischen Studierenden und Referendar:innen belegen.
- 5) Studierende können relevante Aspekte erkennen (53%), fokussieren aber häufig irrelevante und oberflächliche Aspekte (Schäfer und Seidel 2015, S. 42). Hinsichtlich des Reasonings weisen die Ergebnisse eher auf naive Annahmen (80%) hin (Schäfer und Seidel 2015, S. 48). Die Diskrepanz zu den Urteilen von Expert:innen ist eher hoch (63-74%) (Schäfer und Seidel 2015, S. 49). Sowohl den Studierenden als auch den Referendar:innen fällt das Beschreiben der Situation am leichtesten (40% bzw. 44% korrekte Antworten), weniger leicht fällt ihnen das Erklären und am wenigsten das Vorhersagen (Schäfer und Seidel 2015, S. 50–51; Stürmer et al. 2015, S. 352). Innerhalb der Gruppe der Referendar:innen sind deutliche Fähigkeitsunterschiede sichtbar (Stürmer et al. 2015, S. 352–353).

4.2.2 Weitere Studien zur Förderung der Professionellen Wahrnehmung

Die Förderung der Professionellen Unterrichtswahrnehmung hinsichtlich der Klassenführung überprüft die Arbeitsgruppe Gold et al. (2013) in einem Vergleich zwischen einer Interventions- und Kontrollgruppe mit einer Prä-Posttestung (Gold et al. 2013, S. 145). Die Interventionsgruppe ($N = 57$ Grundschullehramtsstudierende, $M_{Sem} = 3.46$, $SD_{Sem} = 1.98$) erhält ein videobasiertes Trainingsprogramm aus sieben ein- bis dreiminütigen Vignetten aus dem Grundschulbereich mit einem Fragenkatalog aus geschlossenen Items zur Klassenführung (Gold et al. 2013, S. 148). Die Kontrollgruppe ($N = 63$ Studierende) absolviert ausschließlich die Prä-Posttestungen und ist hinsichtlich des Studiengangs und der Semesteranzahl vergleichbar (Gold et al. 2013, S. 148). Im Vergleich der mittleren Lösungsscores zwischen beiden Gruppen zeigt sich bei der Interventionsgruppe ein signifikanter Zuwachs bei zwei der drei getesteten Facetten der Klassenführung (Strukturierung des Unterrichtsverlaufs: $F(1, 117) = 3.80$, $p = 0.05$, $\eta^2 = 0.03$, Gruppenfokus: $F(1, 117) = 13.41$, $p < 0.001$, $\eta^2 = 0.10$) (Gold et al. 2013, S. 149). Die Facette der Allgegenwärtigkeit bringt keinen signifikanten Gruppenunterschied hervor, zeichnet sich allerdings im Prätest als sehr leichtes Konstrukt ab und kann möglicherweise deswegen keinen Effekt erzielen (Gold et al. 2013, S. 151).

Den Fokus auf die Klassenführung setzen ebenfalls Hellermann et al. (2015) in ihrem Interventions-Kontrollgruppendesign. Dabei untersuchen sie zum einen die Auswirkung der videobasierten Intervention ($N = 68$ Studierende des Grundschullehramts) im Vergleich zur Kontrollgruppe ohne Training ($N = 28$, $M_{Sem} = 4.96$, $SD_{Sem} = 1.71$), zum anderen den Effekt des Einsatzes fremder ($N = 36$, $M_{Sem} = 4.66$, $SD_{Sem} = 1.27$) versus eigener und fremder Videos ($N = 32$, $M_{Sem} = 4.66$, $SD_{Sem} = 1.18$) als Analyseinstrument (Hellermann et al. 2015, S. 99–100). Die Testung zur Überprüfung der Professionellen Unterrichtswahrnehmung hinsichtlich der drei Facetten der Klassenführung Monitoring, prozessuale Strukturierung sowie Regeln und Routinen setzt sich aus vier zwei- bis fünfminütigen Vignetten und 47 geschlossenen Rating-Items zusammen (Hellermann et al. 2015, S. 100). Hellermann et al. (2015, S. 103–104) können ebenfalls eine signifikante Verbesserung der Professionellen Unterrichtswahrnehmung hinsichtlich der Klassenführung bei den Interventionsgruppen im Gegensatz zur Kontrollgruppe ermitteln (Hellermann et al. 2015, S. 104). Damit decken sich die Ergebnisse mit der Erhebung von Gold et al. (2013). Zusätzlich können Hellermann et al. (2015, S. 104) Unterschiede in der Wirksamkeit der Analyse fremder Videos versus eigener und fremder Videos zu Gunsten der Kombinationsvariante feststellen ($p = 0.031$; $d = 0.39$). Zusammenfassend kann aus dieser Studie festgehalten werden, dass die Professionelle Unterrichtswahrnehmung hinsichtlich der Klassenführung im Gesamten sowie in den drei Facetten mit der Analyse eigener

und fremder Videos am besten gefördert werden kann (Hellermann et al. 2015, S. 103–104).

Eine weitere Arbeit aus dem Grundschulbereich stellen Meschede et al. (2015) vor. Ziel ist es, die Förderung der Professionellen Unterrichtswahrnehmung der Lernunterstützung im naturwissenschaftlichen Bereich sowie mögliche Unterschiede zwischen Expert:innen und Noviz:innen aufzuzeigen (Meschede et al. 2015, S. 323). Zum Einsatz kommt ein videobasiertes Instrument, bestehend aus sechs Vignetten und insgesamt 68 Rating-Items (Meschede et al. 2015, S. 327). Die Konstruktvalidierung des Tools mit einer Pilotierungsstichprobe von Bachelor- und Masterstudierenden sowie Lehrpersonen ($N_{gesamt} = 235$) ergibt eine eindimensionale Struktur des Knowledge-based Reasonings als untersuchtem Teilprozess der Professionellen Unterrichtswahrnehmung (Meschede et al. 2015, S. 326). Damit unterscheidet sich die Modellstruktur von den Ergebnissen des OBSERVERS (Seidel et al. 2010), der eine dreidimensionale Struktur des Knowledge-based Reasonings identifizieren kann (siehe Abschnitt 4.2.1 *Studien zum OBSERVER*). Ein weiterer Unterschied liegt in der Auswertung, die im Gegensatz zum OBSERVER nicht über eine dichotome (Seidel und Stürmer 2014, S. 759), sondern über eine Partial-Credit-Übereinstimmung erfolgt (Meschede et al. 2015, S. 327). Trotz der Eindimensionalität betrachten Meschede et al. (2015, S. 328) die Schwierigkeit der Teilprozesse des Beschreibens ($M = -0.15$, $SD = 0.47$) und Interpretierens ($M = 0.19$, $SD = 0.53$) mit dem Ergebnis, dass signifikante Unterschiede ($t(66) = 2.60$, $p = 0.01$) zwischen den Prozessen auftreten. In einem nächsten Schritt wird an einer weiteren Stichprobe überprüft, ob Gruppenunterschiede zwischen Bachelorstudierenden ($N = 193$, $M_{Sem} = 1.03$, $SD_{Sem} = 0.16$), Masterstudierenden ($N = 116$, $M_{Sem} = 7.86$, $SD_{Sem} = 1.24$) und erfahrenen Lehrpersonen ($N = 125$, $M_{Berufsjahe} = 9.61$, $SD = 7.73$) hinsichtlich der Professionellen Unterrichtswahrnehmung in der Lernunterstützung auftreten (Meschede et al. 2015, S. 326). Diese Annahme kann mit den gewonnenen Daten unterstützt werden. Es zeigen sich signifikante Unterschiede der Mittelwerte zwischen

- 1) den Bachelorstudierenden und den Masterstudierenden ($p < 0.001$, $d = 0.77$) sowie den Lehrpersonen ($p < 0.001$, $d = 1.10$)
- 2) den Masterstudierenden und Lehrpersonen ($p < 0.05$, $d = 0.33$) (Meschede et al. 2015, S. 330).

Damit kann die Schlussfolgerung abgeleitet werden, dass sich mit zunehmender Erfahrung die Professionelle Unterrichtswahrnehmung hinsichtlich der Lernunterstützung verbessert. Der Effekt zwischen dem Beginn des Studiums und der fortschreitenden Studierdauer bzw. vorhandener Lehrerfahrung ist deutlich höher als zwischen den Masterstudierenden und Lehrpersonen.

Das vorgestellte Instrument in der Erhebung von Meschede et al. (2015) kommt ebenfalls bei Sunder et al. (2016) als Testinstrument zum Einsatz. In ihrer Arbeit untersuchen sie den Effekt einer Intervention ($N = 21$ Grundschullehramtsstudierende) im Rahmen zweier Wahlpflichtseminare im Fach Sachkunde in Gegenüberstellung zu einer Kontrollgruppe ohne Intervention ($N = 16$ Grundschullehramtsstudierende ohne Sachkunde, physik- oder chemiebezogenen Veranstaltungen) (Sunder et al. 2016, S. 6–7). Die Intervention bezieht sich auf die theoriebasierte Analyse fremder und eigener Videos zu den Unterrichtsthemen Schwimmen und Sinken sowie Aggregatzustände hinsichtlich lernunterstützender Maßnahmen wie zum Beispiel des Scaffoldings (Sunder et al. 2016, S. 4). Die Studierenden dieser Intervention erreichen im Vergleich zur Kontrollgruppe eine signifikante Verbesserung der Professionellen Unterrichtswahrnehmung hinsichtlich der Lernunterstützung bei den Vignetten zum Thema Schwimmen und Sinken ($F(1, 35) = 18.86, p < 0.001, \eta^2 = 0.35$), aber nicht zum Thema Aggregatzustände ($F(1, 35) = 0.00, p = 0.99, \eta^2 = 0.00$) (Sunder et al. 2016, S. 8). Damit zeigt sich eine themenspezifische Verbesserung der Professionellen Wahrnehmung im Bereich Sachkunde (Sunder et al. 2016, S. 9). Zusätzlich werten Sunder et al. (2016, S. 8–9) die Zufriedenheit und Relevanz der Arbeit mit Videos im Allgemeinen und differenziert nach ihrer Herkunft aus. Insgesamt bewerten die Studierenden die Videoanalyse als positives Mittel (85,7%), dabei erzielt die Arbeit mit eigenen Videos (76,2%) mehr positive Resonanz als die Arbeit mit fremden Videomaterial (33,3%) (Sunder et al. 2016, S. 8–9). Fremdes Unterrichtsmaterial wird beispielsweise als sinnvolles Medium gesehen, basierend auf der Theorie des Modelllernens (Sunder et al. 2016, S. 9). Negative Aspekte beziehen sich ausschließlich auf allgemeiner Ebene wie des ökonomischen Einsatzes (Sunder et al. 2016, S. 9).

Den Vergleich zwischen eigenen und fremden Videos betrachten Frommelt et al. (2016, S. 362) in einem Interventions-Kontrollgruppendesign als Teil der Interventionsstudie VideA (Videos in der Ausbildung von Lehrkräften). Dabei arbeiten Studierende im 1. Ausbildungsjahr mit eigenen ($N = 46$) oder fremden Videos ($N = 40$) bzw. mit schriftlichen Lehr-Lernmaterialien ($N = 43$) (Frommelt et al. 2016, S. 363). Im Prä-Post-Vergleich erreicht die Videogruppe (eigene und fremde Videos) im Vergleich zur Kontrollgruppe mit den schriftlichen Dokumenten eine signifikante Verbesserung der Professionellen Unterrichtswahrnehmung ($F(1,127) = 5.67, p = 0.20, \eta^2 = 0.40$) (Frommelt et al. 2016, S. 365). Die Analyse eigener versus fremder Videos erzielt dabei keinen Effekt ($F(1,84) = 0.26, p = 0.62, \eta^2 = 0.00$) (Frommelt et al. 2016, S. 366). Hinsichtlich der höchsten Fähigkeit innerhalb des Knowledge-based Reasonings decken sich bei Frommelt et al. (2016, S. 369) die Ergebnisse mit anderen Studien, bei denen ebenfalls das Beschreiben am besten gelingt (Stürmer et al. 2015, S. 352; Schäfer und Seidel 2015, S. 50–51; Vgl. Meschede et al. 2015, S. 328). In einer Regressionsanalyse wird der Einfluss des Interesses am Lehren und Lernen sowie des Wissens über Unterrichtsqualitätsmerkmale überprüft (Frommelt et al. 2016,

S. 367). Der Zuwachs der Professionellen Unterrichtswahrnehmung kann damit nicht erklärt werden, sondern nur über die Interventionszugehörigkeit (Frommelt et al. 2016, S. 369).

In den meisten Forschungsarbeiten liegt der Fokus auf dem Teilprozess des Knowledge-based Reasonings der Professionellen Unterrichtswahrnehmung. Santagata und Guarino (2011) beziehen sich mit ihrer qualitativen Studie zur Förderung der Unterrichtsanalyse auf beide interagierenden Prozesse, das Noticing und das Knowledge-based Reasoning. Dabei untersuchen sie die Fähigkeit, auf wichtige Dinge im Unterricht eingehen, den Unterricht wissensbasiert beurteilen und alternative Strategien vorschlagen zu können (Santagata und Guarino 2011, S. 140). In einem Prä-Post-Design analysieren angehende Lehrkräfte ($N = 27$ Grundschulstudierende im fünften Jahr) eine 16 minütige, zusammengeschnittene Videosequenz aus der 1.Klasse zum Thema Zahlenwerte, indem sie zum einen die Unterrichtseinheit beschreiben, zum anderen lernrelevante Inhalte und Lehrstrategien ausführlich beschreiben, kommentieren und diskutieren sollen (Santagata und Guarino 2011, S. 140). Bereits zu Beginn können die Studierenden die Unterrichtsaktivitäten gründlich beschreiben und zeigen demnach keine Verbesserung in der Posttestung ($t(26) = 1.546, p = 0.134$) (Santagata und Guarino 2011, S. 142). Eine signifikante Verbesserung der wissensbasierten Beschreibung der Handlungen der Schüler:innen ($t(26) = -5.355, p = 0.000$) und der Lehrpersonen ($t(26) = -3.31, p = 0.003$) geht mit detaillierteren und spezifischeren Formulierungen einher (Santagata und Guarino 2011, S. 142). Ebenfalls werden die Kommentare ausführlicher und integrierter ($t(26) = -3.923, p = 0.000$) (Santagata und Guarino 2011, S. 142–143). Zu Beginn der Erhebung werden von den Studierenden weniger alternative Lehrstrategien ($N = 6$) vorgeschlagen als am Ende ($N = 14$) (Santagata und Guarino 2011, S. 143). Den signifikanten Unterschied ($t(269) = -2.487, p = 0.020$) begründen Santagata und Guarino (2011, S. 143) mit der Voraussetzung, ein lernrelevantes Problem wahrzunehmen und dem fehlenden Wissen für effektivere Strategien.

Eine Studie, die den Fokus auf das Noticing setzt, stammt von Star und Strickland (2008). Lehramtsstudierende der Sekundarstufe mit Schwerpunkt Mathematik ($N = 28$) erhalten einen einsemestrigen Methodenkurs als Intervention (Star und Strickland 2008, S. 111). Während im Pretest eine 50-minütige Videosequenz einer 8. Klasse zum Thema Exponenten zum Einsatz kommt, wird im Posttest eine 45-minütige Sequenz zum Thema Winkel und Kreis verwendet (Star und Strickland 2008, S. 115). Nach der Betrachtung der Videos erhalten die Studierenden Analyseaufträge mit unterschiedlichen Aufgabenformaten (Richtig/Falsch, Multiple Choice und offenen Kurzantworten) zu folgenden Beobachtungskriterien: Klassenraumumgebung, Klassenführung, mathematische Inhalte, Aufgaben und Kommunikation (Star und Strickland 2008, S. 115). Anhand der Auswertung wird deutlich, dass die Studierenden nach dem Methodenkurs eine signifikante Zunahme der Noticing-

Fähigkeit erreichen ($t(27) = 10.727, p < 0.001$) (Star und Strickland 2008, S. 121). Beim Wahrnehmen der mathematischen Inhalte fokussieren sich die Studierenden zum einen auf die Beurteilung der Richtigkeit der Aussagen, zum anderen auf die Beurteilung des Tafelbilds, ob eine chronologische Beschreibung der Inhalte erkennbar ist (Star und Strickland 2008, S. 122). Dies ist vergleichbar mit den Ergebnissen von Schäfer und Seidel (2015, S. 42), die eine Häufung der Fokussierung auf irrelevante und oberflächliche Aspekte festgestellt haben. Kritisch zu hinterfragen bleibt, ob es sich in der Forschungsarbeit von Star und Strickland (2008) um die Erfassung der Fähigkeit des Noticings handelt oder um das wissensbasierte Beschreiben als Teilprozess des Reasonings.

Eine Arbeit, die sich im Gegensatz zu den bisher vorgestellten Arbeiten nicht mit Studierenden oder Lehramtsanwärter:innen auseinandersetzt, sondern den Zusammenhang zwischen der Analysefähigkeit ausgebildeter Lehrpersonen und dem Lernen der Schüler:innen auseinandersetzt, stellen Kersting et al. (2010) vor. Als Messinstrument dient eine Videoplattform mit 13 drei- bis fünfminütigen authentischen Vignetten und Zusatzinformationen aus der Klassenstufe fünf und sechs mit dem Schwerpunkt der Unterstützung durch die Lehrperson (Kersting et al. 2010, S. 174). Die Lehrpersonen ($N = 223, M_{\text{Erfahrung}} = 7.7, SD = 7.5$) sollen die Interaktion der Lehrperson und Lernenden im Video in Bezug zum mathematischen Inhalt schriftlich diskutieren (Kersting et al. 2010, S. 174–175). Die Bewertung der Antworten erfolgt entsprechend den vier Kategorien: mathematischer Inhalt, Schüler:innendenken, Verbesserungsvorschläge und Interpretationstiefe (Kersting et al. 2010, S. 174). Um eine Aussage über den Zusammenhang zwischen der Analysefähigkeit der Lehrperson ($N = 19$) und dem Lernen der Schüler:innen ($N = 317$) treffen zu können, absolvierten die Lernenden ein Lernquiz zum Thema Brüche, das jeweils vor und nach der Unterrichtslektion eingesetzt wurde (Kersting et al. 2010, S. 175). Dabei zeigt sich, dass ausschließlich die Subskala Verbesserungsvorschläge einen Zuwachs beim Lernen der Schüler:innen vorhersagen kann ($r = 0.521, p < 0.05, R^2 = 0.27$) (Kersting et al. 2010, S. 177). Die Studie von Kersting et al. (2010) stellt eine Ausnahme in der aktuellen Forschung zur Professionellen Unterrichtswahrnehmung dar. Bisher lassen sich kaum Arbeiten zur Erforschung der Auswirkungen der Professionellen Wahrnehmung auf den tatsächlichen Handlungsprozess in den Schulen sowie die Auswirkungen bei den Lernenden finden. Hier bedarf es noch weiterer Forschung.

Zusammengefasst können aus den hier beschriebenen Forschungsarbeiten folgende zentrale Ergebnisse abgeleitet werden:

- 1) Die Professionelle Wahrnehmung im Teilprozess des Knowledge-based Reasonings kann themenspezifisch (u.a. im Bereich der Klassenführung oder Lernunterstützung) durch den Einsatz eines videobasierten Trainings bereits im Studium gefördert werden (Frommelt et al. 2016, S. 365; Gold und

Hasselhorn 2013, S. 149; Hellermann et al. 2015, S. 104; Santagata und Guarino 2011, S. 142–143; Sunder et al. 2016, S. 9). Ebenfalls lassen sich Fortschritte im Noticing belegen (Star und Strickland 2008, S. 121).

- 2) Die Kombination aus eigenem einschließlich fremdem Videomaterial erzielt bessere Erfolge hinsichtlich der Förderung der Professionellen Wahrnehmung als die ausschließliche Verwendung von fremden Videos (Hellermann et al. 2015, S. 104).
- 3) Hinsichtlich der Ausprägung der Professionellen Wahrnehmung können Unterschiede zwischen Noviz:innen und Expert:innen theoriekonform aufgedeckt werden (Meschede et al. 2015, S. 330). Bei Noviz:innen zeigt sich, dass das wissensbasierte Beschreiben von Situationen im Vergleich zu den anderen Teilprozessen des Reasonings am besten gelingt (Meschede et al. 2015, S. 328; Frommelt et al. 2016, S. 369).
- 4) Ein Vergleich zwischen der Förderung der Professionellen Wahrnehmung mittels Videos oder Transkripten zeigt eine positivere Beurteilung bei der Verwendung von Videomaterial (Sunder et al. 2016, S. 8–9).

Daraus abgeleitete Konsequenzen für die eigene Forschungsarbeit werden im Abschnitt 4.4 zusammengetragen.

4.3 Das Videotool ViviAn

Das Videotool ViviAn (Videovignetten zur Analyse von Unterrichtsprozessen, abrufbar über die Homepage <https://vivian-training.de>) wurde im Rahmen der Arbeitsgruppe Bartel und Roth (2017) entwickelt. Der Lehransatz und die Lernziele, die mit ViviAn verfolgt werden, sind im Abschnitt 4.3.1 beschrieben. Zur Erreichung der Lernziele ist ein entsprechender Aufbau der Lernumgebung notwendig. Um einen Einblick in die Gestaltung der Oberfläche und die Funktionen von ViviAn zu erhalten, werden die Charakteristika des Tools im Abschnitt 4.3.2 *Beschreibung der Lernumgebung* dargestellt. Seit der Entwicklung von ViviAn wird es stets erweitert und um neue Vignetten zu verschiedenen Themenbereichen aus der Mathematik ergänzt und evaluiert (siehe Abschnitt 4.3.3 *Bisherige Forschungsarbeiten mit ViviAn*). Die Videovignetten der vorliegenden Arbeit mit den entsprechenden Analyseaufträgen zum Thema Motivierung werden in das bestehende Video-Tool ViviAn mit entsprechenden Anpassungen implementiert (siehe Abschnitt 7.6 *Implementierung in ViviAn*).

4.3.1 Lehransatz und Lernziele

Mit dem Einsatz des Videotools soll Lehramtsstudierenden die Möglichkeit gegeben werden, diagnostische Kompetenzen hinsichtlich der Erstellung von Fähigkeitsprofilen von Schüler:innen während Gruppenarbeitsprozesse zu mathematischen

Themenkomplexen zu fördern (Bartel und Roth 2017, S. 43–44). Die dafür verwendeten Videosequenzen stammen aus dem Mathematik-Labor „Mathe ist mehr“, einem außerschulischen Lernort am Campus Landau der RPTU, den Schulklassen besuchen, um vor Ort Laborstationen zu ausgewählten mathematischen Problemen zu bearbeiten (Bartel und Roth 2017, S. 46). Der Fokus liegt dabei auf Gruppenarbeitsprozessen, die von den Studierenden analysiert werden (Bartel und Roth 2017, S. 46–47). Um besonders eine theoriebasierte Beantwortung zu unterstützen, wird der Schwerpunkt auf ein spezifisches fachdidaktisches Thema gelegt, das zuvor in einer universitären Veranstaltung differenziert betrachtet wird (Roth 2020, S. 73–74). Die Studierenden beantworten Diagnoseaufträge mit geschlossenen und offenen Antwortformaten und erhalten am Ende ein Feedback mit Lösungen einer Gruppe von Expert:innen (Bartel und Roth 2017, S. 50). Durch einen selbstständigen Abgleich der eigenen Antworten mit den Antworten der Expert:innen wird zudem die Reflexionsfähigkeit der Studierenden gefördert (Roth 2020, S. 75). Das Videotool wird im Rahmen universitärer Veranstaltungen unterstützend eingesetzt, um eine Theorie-Praxis Verknüpfung zu ermöglichen, indem die Studierenden ihr theoretisches Wissen aus den Seminaren und Vorlesungen in realistischen Unterrichtssituationen anwenden (Bartel und Roth 2017, S. 43–44).

4.3.2 Beschreibung der Lernumgebung

Die Homepage (<https://vivian-training.de>), über die das Videotool abgerufen werden kann, bietet mit ihrem klaren und übersichtlichen Design eine gute Orientierung (siehe Abbildung 14). Bereits über die Startseite können erste Einblicke vom Videotool gewonnen, zahlreiche Informationen und Hilfestellungen zur Anmeldung und Durchführung abgerufen sowie Beschreibungen zu den Forschungsarbeiten und beteiligten Personen gelesen werden.



Abbildung 14 Startseite der Homepage zum Videotool Vivian (Stand 2021) (<https://vivian-training.de>)

Zudem erfolgt der Log-in zum Tool und der Zugang zu den Vignetten über diese Homepage. Zur Wahrung des Schutzes der gefilmten Personen greifen vor der Betrachtung der Vignetten mehrere Kontrollmechanismen (Roth 2020, S. 70):

- 1) Die Nutzer:innen müssen sich registrieren und eine Datenschutzerklärung unterzeichnen.
- 2) Es muss eine Anmeldung in einem Kurs, der von Dozierenden eingerichtet und verwaltet wird, erfolgen. Die Zulassung zum Kurs muss durch die Dozierenden genehmigt werden.
- 3) Die Dozierenden sollen die Anmeldungen und Datenschutzerklärungen zunächst überprüfen und können dann die Nutzer:innen für die Betrachtung und Analyse der Vignetten freischalten.

Die besondere Gestaltung der Lernumgebung (siehe Abbildung 15) ermöglicht den Studierenden ein übersichtliches Abrufen zusätzlicher Informationen zur gezeigten Szene durch Verwendung entsprechender Buttons (Bartel und Roth 2017, S. 47). Es können Informationen zum Thema und Ziel sowie zeitliche und inhaltliche Einordnung der videografierten Situation, verwendete Materialien, Arbeitsaufträge, Arbeitsergebnisse und schulrelevante Daten der Lernenden (Klassenstufe, Alter, Schulart) abgerufen werden (Bartel und Roth 2017, S. 48–49).

The screenshot displays the Vivian interface. On the left is a sidebar with a 'Material' header and a back arrow. Below it are sections for 'Lernumgebung' (containing 'Thema und Ziele'), 'Schülerebene' (containing 'Arbeitsauftrag', 'Material', 'Schülerdokumente'), and 'Meine Ebene' (containing 'Schülerprofile' and a grid of student profiles labeled S1, S2, S3, S4). The main content area features the Vivian logo, a 'Vignette beenden' button, and buttons for 'Video' and 'Diagnoseauftrag'. The selected vignette is titled 'Thema und Ziele' and includes the following text:

Thema und Ziele

Lernumgebung: Thema und Ziele

Die Station zum „Funktionalen Denken“ behandelt verschiedene Aufgaben zum Funktionalen Denken, innerhalb derer die Arbeit mit Funktionsgraphen im Mittelpunkt stehen. Hierbei wird sowohl die Konstruktion als auch die Interpretation von Graphen thematisiert.

Ziel: Die Schülerinnen und Schüler sollen in dieser Station ihr Wissen über folgende Inhalte wiederholen bzw. vertiefen:

- Graphische Darstellung von Zusammenhängen (Konstruktion von Funktionsgraphen)
- Informationsgewinnung aus Funktionsgraphen
- Interpretation von Funktionsgraphen
- Herstellung von Zusammenhängen zwischen verschiedenen Funktionsgraphen

Um die Station erfolgreich bearbeiten zu können, ist es notwendig, dass die Schülerinnen und Schüler ein Vorwissen zu Funktionen mitbringen. Außerdem sollten sie in der Lage sein, ein kartesisches Koordinatensystem zu zeichnen sowie Punkte abzulesen und einzuzichnen. Weiterhin sollten die Schülerinnen und Schüler im Koordinatensystem dargestellte Zusammenhänge interpretieren können.

A 'Schließen' button is located at the bottom right of the main content area.

Abbildung 15 Abrufen von Zusatzinformationen am Beispiel "Thema und Ziele"

ViviAn stellt ein innovatives Instrument zur Analyse und Förderung der Diagnosefähigkeit dar, da es folgende Vorteile miteinander vereint (Roth 2020, S. 70):

- 1) Alle Nutzer:innen haben bei der Bearbeitung der Vignetten die Möglichkeit, wesentliche Informationen zum gezeigten Gruppenarbeitsprozess abzurufen. Damit soll gewährleistet werden, dass sie möglichst den gleichen Erkenntnisstand wie die Lehrperson der Schüler:innen vor Ort haben.
- 2) Jede Antwort kann direkt im Online-Tool beantwortet werden. Es findet eine automatische Speicherung der Angaben statt.
- 3) Nach Fertigstellung der Analyseaufträge können die Nutzer:innen ihre Antworten direkt mit den Antworten einer Gruppe von Expert:innen abgleichen.
- 4) Aus ökonomischer Sicht bietet ViviAn den Nutzer:innen den Vorteil, die Bearbeitung der Vignetten unabhängig von Ort und Zeit durchzuführen (lediglich ein vorgegebenes Zeitfenster seitens des Kursleitenden ist zu beachten).
- 5) Über die Homepage gelangt man zu einem übersichtlichen Menü, mit dem die Vignetten je nach Themenfeld ausgewählt werden können. Eine Beschreibung der Vignetten gibt einen inhaltlichen Einblick der Szenen.
- 6) Dozierende erhalten einen separaten Zugang, der die Verwaltung ihres Kurses ermöglicht. Alle angemeldeten Personen im Kurs werden aufgelistet und können individuell für die Nutzung vorgesehener Vignetten freigeschaltet werden. Zudem wird den Dozierenden der Fortschritt der Vignettenbearbeitung der einzelnen Studierenden angezeigt.
- 7) Die Verwaltung der hinterlegten Daten, der Videomaterialien sowie der zugangsberechtigten Personen obliegt der Verantwortung von zwei Administrator:innen. Der Schutz und die Sicherheit der Daten sollen über die Verwendung des universitären Servers gewährleistet werden.

4.3.3 Bisherige Forschungsarbeiten mit ViviAn

Das Videotool ViviAn wird fortführend durch neue Vignetten ergänzt und in verschiedenen Forschungsarbeiten validiert. Der Schwerpunkt der bisherigen Projekte liegt aktuell auf Themen aus dem Fachbereich Mathematik. Ziel der Arbeiten ist es, wie bereits beschrieben, die Lehramtsstudierenden schon im Studium mit den herausfordernden Aufgaben in der Schule vertraut zu machen und vor allem die diagnostischen Fähigkeiten frühzeitig zu fördern. An dieser Stelle wird ein Einblick zu ausgewählten, bisherigen Arbeiten gegeben.

In einem ihrer Forschungsprojekte untersuchen Bartel und Roth (2020) die Wahrnehmung der Lernumgebung ViviAn von Studierenden. Hierzu wird ein Vergleich zwischen der Analyse von Transkriptvignetten ($N_1 = 75$ Studierende) und Videovignetten ($N_2 = 71$ Studierende) herangezogen, die die Studierenden in ViviAn mit den dazugehörigen Diagnoseaufträgen und bereitgestellten Zusatzinformationen bearbeiten

(Bartel und Roth 2020, S. 304–306). Zur Beurteilung der wahrgenommenen Relevanz und des Interesses an der Arbeit mit Videos oder Transkripten wird ein selbst entwickelter Fragebogen, der bereits in einer Vorstudie pilotiert und mit der oben genannten Stichprobe weiter validiert wird, zu zwei Messzeitpunkten eingesetzt (Bartel und Roth 2020, S. 305). Die hier beschriebenen Ergebnisse beziehen sich auf den ersten Messzeitpunkt nach der Bearbeitung der ersten Vignette. Hinsichtlich der wahrgenommenen Relevanz zeigt sich ein signifikanter Unterschied ($t(138) = 3.475$, $p < 0.001$, $d = 0.59$) zwischen der Videogruppe ($M = 12.41$, $SD = 2.08$) und der Transkriptgruppe ($M = 11.16$, $SD = 2.15$) (Bartel und Roth 2020, S. 311). Das Interesse wird in beiden Gruppen ähnlich beurteilt ($M_{Video} = 10.95$, $M_{Transkript} = 10.55$), ein bedeutsamer Unterschied kann nicht nachgewiesen werden (Bartel und Roth 2020, S. 311). Dagegen beurteilen die Studierenden in der Videogruppe ($M = 4.87$, $SD = 1.04$) die Realitätsnähe signifikant höher ($t(138) = 7.509$, $p = 0.000$, $d = 1.26$) als in der Transkriptgruppe ($M = 3.47$, $SD = 1.19$) (Bartel und Roth 2020, S. 312). Die Ergebnisse deuten darauf hin, dass die Studierenden grundsätzlich an der Arbeit mit ViviAn (Videos oder Transkripte) Interesse zeigen und eine Relevanz für sich wahrnehmen (Bartel und Roth 2020, S. 313). Allerdings beurteilen die Studierenden die Arbeit mit Videovignetten in ViviAn als relevanter und realitätsnäher (Bartel und Roth 2020, S. 313). Dieser Befund unterstützt die Arbeit mit Videovignetten in ViviAn und scheint damit ein geeignetes Instrument zu sein, um die Theorie-Praxis-Verknüpfung zu fördern.

Im Projekt von Hofmann und Roth (2020) wird der Fokus auf die Diagnose von Fehlkonzepten von Schüler:innen in der Arbeit mit Funktionsgraphen aus dem Bereich des Funktionalen Denkens gesetzt. Hierzu werden Schüler:innen bei der Bearbeitung von speziell ausgewählten Aufgaben zu Funktionsgraphen im Rahmen des außerschulischen Lernorts „Mathe ist mehr“ am Universitätscampus Landau gefilmt (Hofmann und Roth 2020, S. 5). Eine kriteriengeleitete Auswahl des Videomaterials wird in einer Gruppe aus Expert:innen diskutiert und mit offenen Diagnoseaufträgen ergänzt (Hofmann und Roth 2020, S. 8–9). Lehramtsstudierende der Mathematik ($N = 112$) des ersten und zweiten Semesters nehmen im Rahmen einer fachdidaktischen Vorlesung an der Interventionsstudie im Pre-Post-Design teil (Hofmann und Roth 2020, S. 10). Zwischen der Analyse der beiden Testvignetten haben die Studierenden die Möglichkeit, Übungsvignetten mit Feedback in Form einer Musterlösung zu bearbeiten (Hofmann und Roth 2020, S. 10). Die Ergebnisse zeigen einen signifikanten Zuwachs beim Diagnostizieren von Fehlern und Schwierigkeiten von Schüler:innen ($M_{Pre} = 2.02$, $SD_{Pre} = 1.50$, $M_{Post} = 2.46$, $SD_{Post} = 1.51$, $p = 0.009$, $d = 0.30$) (Hofmann und Roth 2020, S. 11). Im Vergleich zu den Expert:innen ($N = 12$, $M = 4.25$, $SD = 1.79$) erreichen die Noviz:innen einen deutlich geringeren Punktwert (Hofmann und Roth 2020, S. 11). Zusammenfassend bedeuten die Ergebnisse, dass die Diagnosefähigkeit im fachspezifischen Bereich hinsichtlich der Fehlkonzepte mit der Bearbeitung von Videovignetten in ViviAn gefördert werden kann. Zudem stützen die

Ergebnisse bisherige Forschungserkenntnisse zur unterschiedlichen Diagnosefähigkeit von Expert:innen und Noviz:innen.

In einer weiteren Forschungsarbeit untersuchen Walz und Roth (2019) den Zusammenhang zwischen den diagnostischen Fähigkeiten und der Interventions- und Reflexionsfähigkeit von Studierenden. Zur Beurteilung der Ausprägung der diagnostischen Fähigkeiten der Studierenden ($N = 48$) hinsichtlich der Gruppenarbeitsprozesse von Schüler:innen zum Themenkomplex Terme werden drei Videovignetten mit 14 Diagnoseaufträgen im Videotool ViviAn verwendet (Walz und Roth 2019, S. 1100). Nach einer qualitativen Analyse zeigt sich, dass die Studierenden unterschiedlich stark ausgeprägte diagnostische Fähigkeiten im „Deuten“ (Vgl. Beretz et al. 2017) besitzen. Je nachdem wie viele ihrer Antworten mit denen der Expert:innen übereinstimmen, können sie in drei Gruppen eingeteilt werden (Walz 2020, S. 105–111):

- 1) verhältnismäßig hohe diagnostische Fähigkeiten (mindestens zehn Übereinstimmungen)
- 2) mittlere diagnostische Fähigkeiten (sieben bis neun Übereinstimmungen)
- 3) geringe diagnostische Fähigkeiten (keine bis sechs Übereinstimmungen)

Die maximale Übereinstimmung richtig beantworteter Diagnoseaufträge von teilnehmenden Studierenden mit Expert:innen entspricht 11 (Walz 2020, S. 105). Ein Vergleich mit Lehrpersonen ($N = 15$) bestätigt, dass Noviz:innen geringere diagnostische Fähigkeiten besitzen als erfahrene Lehrkräfte (Walz 2020, S. 110–111). Zur Beurteilung der Ausprägung der Interventionsfähigkeit beobachten Studierende über einen Monitor Schüler:innen in Kleingruppen im außerschulischen Lernort „Mathe ist mehr“ bei der Bearbeitung einer Aufgabe und müssen selbstständig entscheiden, wann eine Intervention notwendig ist und diese selbstständig ausführen (Walz und Roth 2019, S. 1101). Als ein Ergebnis der qualitativen Auswertung durch Expert:innen kann festgehalten werden, dass Studierende mit geringen diagnostischen Fähigkeiten weniger angemessene Interventionen durchführen als Studierende mit höheren diagnostischen Fähigkeiten (Walz 2020, S. 209). Auch mit dieser Arbeit kann der Unterschied in der Ausprägung der Diagnosefähigkeit zwischen Expert:innen und Noviz:innen belegt werden. Zudem macht sie deutlich, dass eine Differenzierung hinsichtlich der diagnostischen Fähigkeit innerhalb der Gruppe der Studierenden sinnvoll erscheint. Eine Klassifizierung entsprechend der Ausprägung der Fähigkeiten nehmen auch Stürmer et al. (2015, S. 352–353) bei den Referendar:innen vor, da sie ebenfalls große Unterschiede zwischen den einzelnen teilnehmenden Personen feststellen.

Welchen Einfluss das Feedback nach der Bearbeitung von Vignetten auf die Förderung diagnostischer Fähigkeiten hat, betrachten Enenkiel und Roth (2019) in ihrer Arbeit näher. Die Videovignetten aus dem Projekt beziehen sich auf

Gruppenarbeitsprozesse von Schüler:innen zum Thema Längen, Flächen und Rauminhalten (Enekiel und Roth 2019, S. 1089). Es zeigt sich, dass Studierende mit der Analyse von Trainingsvignetten ihre diagnostischen Fähigkeiten (Vgl. Beretz et al. 2017) im Beschreiben ($t(103) = 9.21, p < 0.05, r = 0.67$), Deuten ($t(103) = 10, p < 0.05, r = 0.7$) und Ursache finden/Intervenieren ($t(103) = 8.48, p < 0.05, r = 0.64$) fördern können (Enekiel und Roth 2019, S. 1089). Ein Vergleich zwischen den unterschiedlichen Zeitpunkten des Feedbacks a) verzögertes Feedback (zusammengefasste Rückmeldung am Ende jeder Vignette) und b) sofortiges Feedback (nach jedem Analyseauftrag) zeigt keinen Einfluss auf die Zunahme der diagnostischen Fähigkeiten (Enekiel und Roth 2019, S. 1089–1090). Währenddessen kann die Kontrollgruppe, die kein Training durchläuft und damit kein Feedback erhält, ihre diagnostischen Fähigkeiten zwischen der Pre- und Posttestung nicht verbessern (Enekiel und Roth 2019, S. 1090). Nicht nur Hofmann und Roth (2020), sondern auch Enekiel und Roth (2019) können für ihren spezifischen Themenbereich im Fach Mathematik nachweisen, dass die Analyse von Videos in ViviAn einen Zuwachs der diagnostischen Fähigkeiten bewirkt. Dabei scheint die Rückmeldung korrekter Antworten für den Erfolg unabdingbar.

Die vorgestellten Arbeiten unterstützen den Einsatz eines videobasierten Instruments in der Lehramtsausbildung zur Förderung diagnostischer Fähigkeiten. Dabei stellt ViviAn mit den vorläufigen Ergebnissen ein vielversprechendes Instrument dar, das vor allem durch die Verdeutlichung der Relevanz und der Realitätsnähe eine gute Verzahnung zwischen der universitären Lehre und der Schulpraxis herstellen kann.

4.4 Zusammenfassung in Bezug auf die vorliegende Erhebung

Der Einsatz von Videos in der Lehre bietet den großen Vorteil komplexe Unterrichtssituationen festzuhalten und aus unterschiedlichen Perspektiven analysieren zu können (van Es 2009, S. 101). Wenn dies bereits während der Lehramtsausbildung erfolgt, erhalten die Studierenden nicht nur einen praktischen Bezug, sondern können gezielt ihr theoretisch erworbenes Wissen anwenden und überprüfen (Vgl. Bartel und Roth 2017, S. 43–44; Blomberg et al. 2013, S. 106). Dieses Verständnis von Lehre soll auch mit der vorliegenden Erweiterung des Videotools ViviAn mit Vignetten und Analyseaufträgen zum Bereich der Motivierung verfolgt werden. Konkret bedeutet dies im Hinblick auf die Betrachtung der fünf Heuristiken nach Blomberg et al. (2013) (siehe Abschnitt 4.1) zur Verwirklichung des dargestellten Lehransatzes, dass es zunächst authentische und realitätsnahe Videovignetten zu entwickeln gilt. Diese sollen motivationsförderliche oder auch -hemmende Maßnahmen der Lehrperson abbilden. Hierbei ist die Auswahl geeigneter Filmsequenzen bedeutsam, die in der vorliegenden Arbeit nicht wie in der Literatur gängig beschrieben über die Urteile von Expert:innen erfolgt (u.a. OBSERVER, Seidel et al. 2010, S. 301–302), sondern über

einen Abgleich mit der tatsächlichen Motivation der Lernenden (siehe Abschnitt 7.2.2 *Validierung der Videovignetten*). Da die Motivation als innerer Prozess nicht direkt erfasst werden kann (Krapp und Hascher 2014a, S. 234), steht am Anfang ganz bewusst eine Selbsteinschätzung. Dies bedeutet, dass sich die Schüler:innen über motivierende und weniger motivierende Phasen im Unterricht durch die Beantwortung von Fragen über ihre aktuelle Motivation (Prenzel et al. 1996) bewusst werden. Diese Selbsteinschätzung dient als Ausgangspunkt für die Auswahl der Vignetten. Bei der Bearbeitung der Filmsequenzen scheint eine mehrperspektivische Darstellung der Situation sinnvoll. Zum einen soll das Handeln der Lehrperson dargestellt werden können, zum anderen sollen Mimik und Gestik der Schüler:innen als Reaktion eingefangen werden. Dies muss bereits bei der Aufnahme im Klassenzimmer berücksichtigt werden. Demnach braucht es mindestens zwei Kameras, um eine Mehrperspektivität umsetzen zu können.

Nach dem Einsatz des Videotools *ViviAn* sollen die Anwender:innen mit Hilfe eines Fragebogens (siehe Kapitel 9 *Die Akzeptanz des Tools ViviAn*) Authentizität und Relevanz der Vignetten beurteilen. Diese Aspekte sind für die Arbeit mit Videos in der Lehre bedeutsam (von Aufschnaiter et al. 2017, S. 89).

Das Videotool *ViviAn*, wie im Kapitel 4.3. beschrieben, kann unter Berücksichtigung der vorgestellten Forschungsarbeiten als sinnvolles Instrument in der Lehramtsausbildung gesehen werden. Damit es als Testinstrument dienen kann, werden passend zu den entwickelten Vignetten Analyseaufträge ergänzt (siehe Abschnitt 8.1 *Entwicklung der Analyseaufträge*) (von Aufschnaiter et al. 2017, S. 86). Dazu wird ein geschlossenes Aufgabenformat gewählt, um eine hohe Objektivität der Auswertung (von Aufschnaiter et al. 2017, S. 93) und einen ökonomischen Einsatz (siehe Abschnitt 5.4 *Nebengütekriterien*) zu gewährleisten. Im Hinblick auf die Verfolgung möglicher Lernziele (siehe Abschnitt 4.1) durch den Einsatz von Vignetten und Analyseaufträgen zur Motivierung, können folgende Ziele formuliert werden:

- 1) Die Nutzer:innen des Videotools können die gezeigten Unterrichtssituationen mit der Beantwortung von Fragen entsprechend der vierstufigen Likert-Skala wissensbasiert beschreiben, erklären und vorhersagen.
- 2) Sie können selbstständig das Tool zur strukturierten Analyse verwenden (Videos aufrufen, Unterrichtsmaterialien ansehen sowie Fragen beantworten).
- 3) Mit dem Abgleich der Urteile von Expert:innen können die Nutzer:innen ihre eigenen Analysefähigkeiten einschätzen.

Weiterführende Zielsetzungen sind:

- 4) Den eigenen Unterricht hinsichtlich der Motivierung reflektieren können.
- 5) Ideen zur Förderung der Motivation der Schüler:innen für den eigenen Unterricht entwickeln können.

Das Videomaterial (siehe Abschnitt 4.1) für die Entwicklung der vorliegenden Vignetten soll aus mehreren Gründen bei einer fremden Lehrperson mit ihren Klassen erstellt werden: Videoszenen einer fremden Lehrperson eignen sich besonders für eine kritisch durchdachte Analyse (Krammer 2014, S. 168; Steffensky und Kleinknecht 2016, S. 313) und verzerren die Auseinandersetzung mit dem Material nicht durch die Beobachtung der eigenen Person. Außerdem können mit einer fremden Lehrperson auch problematische Unterrichtshandlungen besser thematisiert werden. Zudem ist die Entwicklung ökonomischer, da nicht jedes Mal neu gefilmt werden muss und die Datenschutzrichtlinien in den gefilmten Unterrichtsklassen nur einmalig umzusetzen sind. Dennoch wäre eine Kombination aus eigenen und fremden Videos eine Möglichkeit zur Optimierung des vorliegenden Tools. Unter dem Aspekt des ökonomischen Einsatzes in der Lehramtsausbildung stellt dies aktuell keine Option dar. Zudem bringt eine hohe Standardisierung des Tools in Form fremder Videos genau diesen Vorteil der Effizienz mit sich. Um sowohl motivationsförderliche als auch motivationshemmende Facetten gezielt in den Videoszenen aufzeigen zu können (Vgl. Bischoff et al. 2005, S. 385; Meschede und Steffensky 2018, S. 25), sollen authentische, aber modifizierte Vignetten (von Aufschnaiter et al. 2017, S. 96) verwendet werden. Dies bedeutet, dass motivationsrelevante Aspekte durch die Instruktion der Lehrperson hervorgehoben werden (siehe Abschnitt 7.2.1 *Validierung mit Unterrichtsadaption*) und ein Zusammenschnitt einzelner Szenen vorgenommen werden soll. Damit kann eine stärkere inhaltliche Fokussierung stattfinden (Meschede und Steffensky 2018, S. 26). Zur Vermeidung einer Überbelastung beim Analysieren der Videos werden einzelne Videoausschnitte von maximal zwei Minuten Länge statt ganzer Unterrichtsstunden eingesetzt. Zusätzlich werden Untertitel eingefügt, um zu gewährleisten, dass alle Aussagen gut verstanden werden können. Das erstellte Videomaterial wird auf der Online-Plattform ViviAn „Videovignetten zur Analyse von Unterrichtsprozessen“ der RPTU (siehe Abschnitt 4.3 *Das Videotool ViviAn*) implementiert. Damit gehen unter anderem ökonomische Vorteile und das Einbetten von Zusatzinformationen einher (Kersting et al. 2010, S. 174). Solche Zusatzinformationen erweisen sich als sinnvoll (Bartel und Roth 2017, S. 45; von Aufschnaiter et al. 2017, S. 95) und werden auch in ViviAn integriert. Sie geben den Studierenden die Möglichkeit, alle relevanten Materialien und Informationen über die Unterrichtseinheit und die Klasse zu erfahren.

Bisher eingesetzte Videotools in der Lehre sind sehr vielversprechend (siehe Abschnitte 4.2.1, 4.2.2 und 4.3.3). Sie können als Instrument zur Analyse und Förderung der Professionellen Unterrichtswahrnehmung eingesetzt werden (u.a. Frommelt et al. 2016, S. 365; Gold und Hasselhorn 2013, S. 149; Hellermann et al. 2015, S. 104; Santagata und Guarino 2011, S. 142–143; Sunder et al. 2016, S. 9). Ziel der vorliegenden Arbeit ist die Überprüfung und Validierung des Videotools ViviAn mit den Vignetten und Analyseaufträgen zur Förderung der Professionellen Unterrichts-

wahrnehmung hinsichtlich der Motivierung, um auch hier ein geeignetes Analyseinstrument vorweisen zu können. Dies ist vor allem notwendig, da die Förderung der Professionellen Wahrnehmung in einigen Studien nur für themenspezifische Bereiche nachgewiesen werden kann (Gold et al. 2013, S. 149; Sunder et al. 2016, S. 9).

Der OBSERVER (Seidel et al. 2010) als ein bekannter Vertreter eines videobasierenden Test- und Trainingsinstruments in Deutschland, stellt ein reliables und valides Instrument zur Förderung der Professionellen Wahrnehmung hinsichtlich der lernrelevanten Aspekte Zielklarheit und -orientierung, Unterstützung und Führung durch die Lehrperson sowie Lernklima dar (Seidel und Stürmer 2014, S. 743). Im Gegensatz dazu werden in ViviAn Vignetten zu motivationsrelevanten Aspekten wie z.B. Relevanz, Fehlerklima und Leistungsdifferenzierung (siehe Abschnitt 7.5 *Beschreibung der Vignetten*) implementiert. Diese Themenbereiche deckt der OBSERVER nicht ab. Die Forschungsergebnisse zum OBSERVER zeigen, dass der Aufbau mit den geschlossenen Antwortformaten und einer vierstufigen Likert-Skala sowie die Differenzierung der Analyseaufträge hinsichtlich des Beschreibens, Erklärens und Vorhersagens (Seidel et al. 2010, S. 301) statistisch abgesichert werden können (Seidel und Stürmer 2014, S. 759; Jahn 2014, S. 176; Stürmer und Seidel 2015, S. 60). Dies spricht für die Übertragung der Struktur auf die Analyseaufträge der vorliegenden Arbeit zum Thema Motivierung. Kritisch betrachtet werden die Beispielitems des OBSERVERS zum Erklären der Situation (siehe Abschnitt 4.2.1 Studien zum OBSERVER, Tabelle 3). Es muss der Frage nachgegangen werden, ob die Formulierung “the students have the opportunity to feel that their teacher takes them seriously“ (Seidel und Stürmer 2014, S. 751) tatsächlich auf eine wissensbasierte Erklärung der Situation abzielt. Bei der Entwicklung der Analyseaufträge zu den motivationsrelevanten Aspekten in dieser Arbeit wird daher eine Abgrenzung der Fragenstruktur zum Erklären, wie beim OBSERVER erfolgt, vorgenommen (siehe Abschnitt 8.1 *Entwicklung der Analyseaufträge*). Bei der Auswertung der Items werden die Antworten der teilnehmenden Personen am OBSERVER dichotomisiert. Um ein differenzierteres Abbild der Fähigkeiten mit dem vorliegenden ViviAn-Tool zum Bereich Motivierung darstellen zu können, wird bei der hier geplanten Auswertung auf Partial-Credit-Items zurückgegriffen (siehe Abschnitt 8.2 *Validierung der Analyseaufträge*).

Mit der begründeten Anlehnung am OBSERVER, aber auch der Abgrenzung zu ihm, wird eine Anpassung des Videotools ViviAn für den vorliegenden Bereich der Motivierung vorgenommen. Mit dieser Arbeit werden die ersten Schritte der Validierung des adaptierten Videotools ViviAn zur Analyse der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung im Unterricht durchgeführt.

5 Gütekriterien der Testentwicklung

Die Validierung des vorliegenden videobasierten Instruments ist zentrales Anliegen dieser Arbeit. Bevor das Videotool als Instrument zur Testung und Förderung der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung eingesetzt werden kann, muss es hinsichtlich der Erfüllung der Gütekriterien überprüft werden. Zur Beurteilung der Güte eines Tests gelten die Gütekriterien als Norm wissenschaftlicher Erhebungen. Neben den drei Hauptgütekriterien (siehe Abschnitt 5.1 – 5.3) *Objektivität*, *Reliabilität* und *Validität* (Rost 2004, S. 33; Lienert und Raatz 1998, S. 7) gibt es zahlreiche andere Kriterien, die zur Beurteilung der Qualität einer Erhebung herangezogen werden können bzw. sollen. Solche *Nebenkriterien* (siehe Abschnitt 5.4) sind laut Lienert und Raatz (1998, S. 7) Normierung, Vergleichbarkeit, Ökonomie und Nützlichkeit. Diese können durch Unverfälschbarkeit und Fairness ergänzt werden (Gold und Hasselhorn 2013, S. 404). Darüber hinaus nimmt die Skalierung (Kubinger 2009, S. 82) vor allem bei Verwendung von Messmodellen eine wesentliche Rolle ein. Im Anschluss der allgemeinen Beschreibung der Gütekriterien folgt eine *Zusammenfassung in Bezug auf die vorliegende Erhebung* (siehe Abschnitt 5.5), in der explizit auf das im Rahmen der Arbeit entwickelte Videotool eingegangen wird.

5.1 Objektivität

Die Objektivität eines Tests wird gewährleistet, indem das Testergebnis innerhalb der Durchführung, Auswertung und Interpretation nicht von außenstehenden Faktoren beeinflusst wird (Rost 2004, S. 39; Lienert und Raatz 1998, S. 8).

Durchführungsobjektivität

Unter Durchführungsobjektivität versteht man die Unabhängigkeit von Einflussfaktoren auf die untersuchenden Personen während der Bearbeitung des Tests (Lienert und Raatz 1998, S. 8). Als Einflussfaktoren können beispielsweise Rahmenbedingungen wie Zeit und Ort oder kommunikative Aspekte wie Hilfestellungen und Formulierungen eruiert werden (Bühner 2021, S. 569). Mit einem hohen Standardisierungsgrad kann die Durchführungsobjektivität erhöht werden (Lienert und Raatz 1998, S. 8). Dies umfasst beispielsweise eine genaue, schriftliche Instruktion zur Durchführung oder eine geringe Interaktion mit den Proband:innen (Lienert und Raatz 1998, S. 8). Vor allem in Anbetracht der Versuchsleiter:inneneffekte sollte der Austausch mit den Proband:innen möglichst gering gehalten werden (Kubinger 2009, S. 40). Bei computergestützten Testverfahren müssen zusätzlich die Bedingungen vergleichbar sein bzw. die technischen Voraussetzungen festgelegt werden (Bühner 2021, S. 569).

Auswertungsobjektivität

Mithilfe „genauer Auswertungsvorschriften“ (Bühner 2011, S. 59) können die Antworten aller Teilnehmer:innen im Test gleichermaßen ausgewertet, also ein gleicher Bewertungsmaßstab zugrunde gelegt werden. Ziel ist es, dass unabhängig von der Person, die auswertet, die gleichen Ergebnisse pro Testperson festgestellt werden (Bühner 2021, S. 569). Als Kriterium zur Beurteilung der Übereinstimmung mehrerer Auswerter:innen können verschiedene Koeffizienten ermittelt werden, z.B. Cohens Kappa, Intra-Klassen-Korrelation oder Fleiss Kappa (Döring und Bortz 2016, S. 346–347). Eine hohe Übereinstimmung bedeutet, dass die auswertenden Personen zu gleichen oder ähnlichen Ergebnissen kommen und damit zu einer höheren Auswertungsobjektivität beitragen. Einen weiteren Aspekt stellt der Standardisierungsgrad des Antwortsystems dar. Durch die Verwendung geschlossener Antworten wird ein höheres Objektivitätsmaß erreicht (Eid und Schmidt 2014, S. 66). Im Vergleich dazu sind offene Antworten weniger objektiv, da eine Person über die Angemessenheit der Antwort urteilen muss (Lienert und Raatz 1998, S. 8; Kubinger 2009, S. 44). Computerbasierte Auswertungsverfahren haben grundsätzlich eine hohe Auswertungsobjektivität (Kubinger 2009, S. 46), weil sie nicht von personellen Faktoren beeinflusst werden.

Interpretationsobjektivität

Die Interpretation der Testergebnisse erfordert eine größtmögliche Unabhängigkeit seitens der auswertenden Person (Lienert und Raatz 1998, S. 8). Es sollten möglichst objektive Aussagen darüber getroffen werden, wie hoch die Personenfähigkeit ausgeprägt ist: beispielsweise von unterdurchschnittlich bis überdurchschnittlich (Bühner 2011, S. 60). Die Vorgehensweise bei der Interpretation der Ergebnisse kann normorientiert oder kriteriumsorientiert erfolgen (Moosbrugger und Kelava 2020a, S. 21). Bei der normorientierten Interpretation wird der Vergleich der individuellen Testleistung zu einer repräsentativen Stichprobe hergestellt (Moosbrugger und Kelava 2020a, S. 22). Damit ist eine Beurteilung zum Beispiel über Prozentränge möglich (Kubinger 2009, S. 46). Die kriteriumsorientierte Interpretation eines Testwerts erfolgt durch Zuordnung zuvor festgelegter inhaltlich begründeter Merkmalsausprägungen (Moosbrugger und Kelava 2020a, S. 23). Die Verwendung eines IRT-Modells (siehe Abschnitt 8.2 *Item-Response-Theorie*) ermöglicht eine differenzierte Betrachtung der kriteriumsorientierten Interpretation von Testwerten (Rauch und Hartig 2020, S. 412–416). Analysiert wird dazu die Wahrscheinlichkeit, mit der eine Person ein Item mit einer bestimmten Schwierigkeit lösen kann (Rauch und Hartig 2020, S. 416). Dies wird in Verbindung mit sogenannten Kompetenzniveaus gebracht. Es können verschiedene Kompetenzstufen durch „Post-hoc-Analysen der Items“ oder

„Verwendung von A-priori-Aufgabenmerkmalen“ aufgedeckt werden (Rauch und Hartig 2020, S. 422). Um ein hohes Maß an Interpretationsobjektivität erreichen zu können, wäre es im Nachgang (durch eine Post-hoc-Analyse) möglich, Ankerpunkte für die Kompetenzniveaus mit einer Itemanalyse von Expert:innen durchzuführen (Rauch und Hartig 2020, S. 422).

5.2 Reliabilität

Die Reliabilität eines Tests gibt Auskunft über den Grad der Genauigkeit der Messung eines bestimmten Merkmals (Bühner 2011, S. 60). Diese kann über verschiedene Wege überprüft werden.

Interne (innere) Konsistenz

Die interne Konsistenz wird in der klassischen Testtheorie über die Homogenität der Items mittels Cronbachs Alpha berechnet (Gold und Hasselhorn 2013, S. 403). Dabei wird jedes Item einzeln über eine Korrelation mit der gesamten Skala analysiert (Bühner 2021, S. 598). Bei IRT-Modell wird stattdessen der EAP-Reliabilitätskoeffizient als Kriterium für die Reliabilität zu Grunde gelegt (Rost 2004, S. 382). In der Literatur findet man hinsichtlich der Interpretation des Cronbachs Alpha unterschiedliche Angaben. Häufig werden in Studien Angaben mit einem $\alpha = 0.8$ als gut und $\alpha = 0.7$ als akzeptabel bezeichnet (Vgl. Pospeschill 2010, S. 21). Witte (2019, S. 148) gibt eine optimale Spannweite von $\alpha = 0.65 - 0.95$ an. Für Fähigkeitstest sprechen Werte ab $\alpha = 0.7$ für eine gute Reliabilität (Field et al. 2012, S. 806). Laut Schecker (2014, 5) können auch Tests mit einem $\alpha < 0.7$ akzeptiert werden, sofern im Rahmen der Fragebogenkonstruktion der Test als sinnvoll und die Items als inhaltlich relevant beurteilt werden können.

Alternativ kann zur Überprüfung der inneren Konsistenz die Split-Half-Reliabilität berechnet werden. Dazu wird ein Test „in möglichst ‚gleiche‘ Testhälften unterteilt und die Messwerte dieser Testhälften werden miteinander korreliert“ (Bühner 2011, S. 60). Dieses Vorgehen wirft die kritische Frage von Teilungskriterien auf (Rost 2004, S. 379), die begründet entschieden werden sollten.

Retest-Reliabilität

Die Retest-Reliabilität geht von einer wiederholten Messung mit demselben Instrument aus (Rost 2004, S. 378). Wenn der Test zu verschiedenen Messzeitpunkten durchgeführt wird, kann eine Korrelationen zwischen den Testergebnissen ermittelt werden (Bühner 2011, S. 61). Dabei wird der Test in der gleichen Stichprobe in einem bestimmten Zeitintervall wiederholt (Gold und Hasselhorn 2013, S. 402). Es besteht die Gefahr, dass das Instrument bei der wiederholten Messung nicht mehr das

gleiche misst (Rost 2004, S. 378). Dies kann beispielsweise durch Erinnerungseffekte, persönliche Faktoren oder Übungseffekte bedingt sein (Bühner 2021, S. 598). Mit der Ermittlung der Retest-Reliabilität soll eine Auskunft über die Merkmalsstabilität getroffen werden (Bühner 2021, S. 598).

Paralleltestreliabilität

Der Einsatz eines alternativen Testinstruments, das streng genommen das gleiche misst, wird in Korrelation mit dem vorhandenen Testinstrument gesetzt (Bühner 2011, S. 61; Lienert und Raatz 1998, S. 9). Die Herausforderung der Überprüfung einer solchen Paralleltestreliabilität liegt zum einen in der Entwicklung eines parallelen Testinstruments und dem Einsatz des zusätzlichen Tests in derselben Stichprobe (Rost 2004, S. 377). Zum anderen können Einflussfaktoren wie zum Beispiel eine abfallende Motivation, verminderte Konzentration oder Wiederholungseffekte, die bei einem zweiten Test auftreten können, schwerer kontrolliert werden (Rost 2004, S. 377).

5.3 Validität

Die Validität eines Tests trifft Aussagen über die Gültigkeit der Messung eines bestimmten Merkmals (Lienert und Raatz 1998, S. 10). Es wird bei der Testkonstruktion als wichtigstes Gütekriterium betrachtet (Kubinger 2009, S. 55; Moosbrugger und Kelava 2020a, S. 30).

Augenscheinvalidität

Die Augenscheinvalidität bezeichnet das „augenscheinliche“ Erkennen einer Testperson, auf welche Merkmalsausprägung der Test zielt (Bühner 2011, S. 62). Dies kann für die Akzeptanz des Tests entscheidende Auswirkungen haben (Moosbrugger und Kelava 2020a, S. 31).

Inhaltliche Validität

„Von Inhaltsvalidität spricht man, wenn ein Test (bzw. seine Testitems im Gesamten) und auch jedes einzelne Item das zu messende Merkmal wirklich bzw. hinreichend präzise erfasst“ (Bühner 2011, S. 61–62). Die Überprüfung der Inhaltsvalidität setzt bereits bei der Planung und Itemkonstruktion ein (Moosbrugger und Kelava 2020a, S. 32). Im Anschluss an die Testentwicklung werden zur Überprüfung, ob die Items bzw. der Test inhaltlich valide sind, in der Regel Expertenratings durchgeführt (Kubinger 2009, S. 56; Lienert und Raatz 1998, S. 11; Moosbrugger und Kelava 2020a, S. 32).

Konstruktvalidität

„Unter dem Aspekt der Konstruktvalidität beschäftigt man sich mit der theoretischen Fundierung (vor allem mit der Dimensionalität und der Struktur) des mit dem Test gemessenen Merkmals“ (Moosbrugger und Kelava 2020a, S. 33). Es kann zwischen konvergenter, divergenter Validierung und faktorieller Validierung unterschieden werden (Moosbrugger und Kelava 2020a, S. 34–35). Die konvergente Validität ergibt sich durch eine hohe Korrelation der Messwerte des entwickelten Tests mit denen eines bestehenden Tests, der eine hohe Ähnlichkeit hinsichtlich des Konstrukts aufweist (Bühner 2021, S. 604). Im Gegensatz dazu wird bei der divergenten Validierung ein Test, der eine geringe Ähnlichkeit hinsichtlich des zu messenden Konstrukts aufweist, herangezogen und auf eine niedrige Korrelation getestet (Bühner 2021, S. 604). Bei der faktoriellen Validierung wird die theoretisch angenommene faktorielle Struktur mittels verschiedener Verfahren überprüft, klassischerweise mittels exploratorischer und konfirmatorischer Faktorenanalyse (Moosbrugger und Kelava 2020a, S. 34–35).

Kriteriumsvalidität

Sofern ein positiver Zusammenhang zwischen dem Testwert und einem externen Kriterium aufgezeigt werden kann, besteht eine kriterienbezogene Validität (Eid und Schmidt 2014, S. 62). Döring und Bortz (2016, S. 447) differenzieren entsprechend des Zeitpunkts der Erhebung des im Zusammenhang mit dem gemessenen Merkmal stehenden externen Kriteriums zwischen der retrospektiven, konkurrenten und prognostischen Validität:

- Retrospektiv bedeutet, dass eine Kriteriumsvalidierung des Tests mit einem zuvor gemessenen Merkmal erfolgt.
- Bei der konkurrenten Validierung wird die Korrelation mit einem gleichzeitig erfassten, externen Merkmal ermittelt.
- Entsprechend wird bei der Prognostischen Validierung das externe Kriterium nach der Erhebung des Testwerts erfasst.

Kreuzvalidität

Zur Überprüfung eines modifizierten Modells werden die Indizes der Modellgüte mittels einer Kreuzvalidierung mit einer zweiten Stichprobe verglichen (Wirtz und Nachtigall Christof 2008, S. 217). Dadurch ist es möglich, eine Auskunft über die Replizierbarkeit der Ergebnisse zu geben. Eine ressourcensparende Variante ist die Teilung der Stichprobe in zwei Hälften, um mit der ersten Teilstichprobe das Modell zu testen und mit der zweiten Stichprobenhälfte zu überprüfen (Bühner 2021, S. 502).

5.4 Nebenkriterien

Skalierung

„Ein Test erfüllt das Gütekriterium *Skalierung*, wenn die laut Verrechnungsvorschriften resultierenden Testwerte die empirischen Verhaltensrelationen adäquat abbilden (Kubinger 2009, S. 82)“. Das bedeutet, dass der Test zwischen Personen mit einer hohen Fähigkeit und einer geringen Fähigkeit statistisch trennen kann (Moosbrugger und Kelava 2020a, S. 20). Der Vorteil der Verwendung von Rasch-Modellen (siehe Abschnitt 8.2 und 8.2.1) ist die Möglichkeit, den Aspekt der Skalierung empirisch überprüfen zu können (Moosbrugger und Kelava 2020a, S. 21).

Fairness

Die Fairness eines Tests wird erfüllt, sofern die Beantwortung der Items nicht von testrelevanten systematischen Faktoren wie zum Beispiel Geschlecht, Kultur oder Sprache abhängig ist (Döring und Bortz 2016, S. 449; Moosbrugger und Kelava 2020a, S. 25; Gold und Hasselhorn 2013, S. 405). Die Items eines Tests sollen von allen Testpersonen sprachlich verstanden werden, ebenso wie die Instruktionen (Moosbrugger und Kelava 2020a, S. 26). Zudem sollte hinsichtlich der Durchführungsfairness, insbesondere bei einem digitalen Test, darauf geachtet werden, dass allen Personen der Umgang mit dem Testmedium vertraut ist (Moosbrugger und Kelava 2020a, S. 26). Ein Vergleich zwischen verschiedenen Gruppen kann mittels einer DIF-Analyse (differential item functioning) eines Rasch-Modells hergestellt werden (Bühner 2011, S. 73). Damit kann erfasst werden, ob die Bearbeitung des Tests von spezifischen Merkmalen beeinflusst wird.

Ökonomie

Die Ökonomie eines Tests wird anhand zeitlicher, materieller und finanzieller Faktoren in Bezug zum Erkenntnisgewinn betrachtet (Lienert und Raatz 1998, S. 12; Moosbrugger und Kelava 2020a, S. 24). Dabei können die Durchführungszeit, der Materialverbrauch, die Handhabbarkeit, die Art der Durchführung und die Auswertung hinsichtlich der Ökonomie analysiert werden (Döring und Bortz 2016, S. 449).

Nützlichkeit

Ein Test erfüllt das Gütekriterium der Nützlichkeit, sofern eine Relevanz für die Praxis besteht und es keinen weiteren Test gibt, der dieses Merkmal in gleicher Weise und Güte misst (Döring und Bortz 2016, S. 449; Lienert und Raatz 1998, S. 13; Gold und Hasselhorn 2013, S. 405). Sofern es einen ähnlichen Test gibt, sollte genau verglichen und begründet werden, warum die Entwicklung eines neuen Test sinnvoll

erscheint (Bühner 2021, S. 634). Dies kann Aspekte der Kriteriumsvalidität, Skalierbarkeit (Bühner 2021, S. 634) oder weiterer Gütekriterien beinhalten.

Zumutbarkeit

Die Zumutbarkeit eines Tests wird in der Regel hinsichtlich der zeitlichen, physischen oder psychischen Belastung einer Testperson in Bezug zur Nützlichkeit beurteilt (Moosbrugger und Kelava 2020a, S. 25).

Unverfälschbarkeit

Ein Test gilt als unverfälschbar, wenn es einer Testperson nicht, bzw. kaum ermöglicht wird, das Testergebnis bewusst zu manipulieren (Kubinger 2009, S. 120), zumindest nicht in dem Maße, dass es eindeutig als Testbeeinflussung ersichtlich wird (Döring und Bortz 2016, S. 449). Bei Leistungstests ist es unwahrscheinlich, dass eine absichtliche Verzerrung der Leistung ins Positive erfolgen kann (Kubinger 2009, S. 120). Wohl eher eine Ausnahme wird eine absichtliche Falschbeantwortung bleiben, da eine Person selten ein schlechtes Testergebnis provozieren möchte (Kubinger 2009, S. 120).

Normierung

Eine Testnormierung oder Testeichung stellt ein Gütekriterium dar, das auf Repräsentativität hinsichtlich der Auswertung zielt (Moosbrugger und Kelava 2020a, S. 22; Döring und Bortz 2016, S. 449). Damit soll es möglich sein, individuelle Ergebnisse in Beziehung zu einer Vergleichsstichprobe zu setzen (Lienert und Raatz 1998, S. 11; Moosbrugger und Kelava 2020a, S. 22; Döring und Bortz 2016, S. 449). Für eine Normierung wird eine repräsentative Stichprobe von mehr als 300 Proband:innen benötigt (Döring und Bortz 2016, S. 449).

5.5 Zusammenfassung in Bezug auf die vorliegende Erhebung

An dieser Stelle erfolgt die Dokumentation hinsichtlich der Erfüllung bzw. Nicht-Erfüllung der beschriebenen Haupt- und Nebengütekriterien. Als zentrale Kriterien gelten die Objektivität, Reliabilität, Validität und Skalierung (Bühner 2021, S. 614). Ein Test, der das Kriterium der Skalierung nicht erfüllt, kann kein gutes Testinstrument sein, da es die Merkmalsausprägungen nicht adäquat abbildet (Bühner 2021, S. 614). Daher wird mit dem Kriterium der Skalierung begonnen.

Skalierung

Um die Passung des theoretisch angenommenen Modells mit den vorliegenden Daten überprüfen zu können, werden die Itemfitwerte (Infit und Outfit) aus der Rasch-Skalierung herangezogen (Bond und Fox 2012, S. 238). Zu hohe oder zu niedrige Outfit-Werte ($\text{Outfit}_i = 0.8 - 1.2$, Wu et al. 2016, S. 148) deuten darauf hin, dass die Personenfähigkeit und die Itemschwierigkeit nicht zusammenpassen (Bühner 2021, S. 274). Bei den Infit-Werten kann eine Aussage darüber getroffen werden, ob das Antwortverhalten der Fähigkeit entspricht (Bühner 2021, S. 274). Genauere Angaben zu den Fit-Werten, zur Interpretation sowie zu den Ergebnissen sind im Kapitel 10 zur *Validierung des Videotools Teil I* beschrieben.

Objektivität

Da die Objektivität als Voraussetzung für die Reliabilität gesehen werden kann (Bühner 2011, S. 71), wird im nächsten Schritt die Objektivität als Gütekriterium betrachtet. Der vorliegende Test erhält ein ausführliches Testmanual mit allen raschkonformen Items. Die Durchführungsobjektivität soll in der vorliegenden Erhebung durch folgende Maßnahmen gesichert sein: Zum einen erhalten alle Teilnehmer:innen die gleiche schriftliche Instruktion, in der schrittweise die Durchführung mit dem Videotool beschrieben wird. Zum anderen findet aufgrund der computerbasierten Testung während der Durchführung kein Kontakt zu den Personen statt, um somit eine Beeinflussung durch die Testleitung zu verhindern. Alle technischen Voraussetzungen werden im Vorfeld erklärt sowie der Zugang zum Programm freigeschaltet. Da die teilnehmenden Personen auf Grund der Corona Pandemie während der Erhebungsphase teilweise von zu Hause aus das Videotool anwenden, sind die Bedingungen nicht identisch. Dank des Supports können alle auftretenden technischen Probleme und Schwierigkeiten schnell gelöst werden.

Die Auswertung der Testergebnisse basiert auf einem Expertenrating zur Beurteilung der richtigen, teilrichtigen oder falschen Antworten. Eine hohe Übereinstimmung der Expert:innen bedeutet eine hohe Objektivität. Als Kriterium zur Beurteilung der Übereinstimmung wird Fleiss Kappa herangezogen (siehe Abschnitt 8.2.2 *Methodisches Vorgehen* zur Validierung der Analyseaufträge). Ziel ist es, alle Proband:innen hinsichtlich der Testwerte gleich auswerten zu können. Anhand der Ergebnisse des Expertenratings kann ein Auswertungsbogen erstellt werden, anhand dessen die Daten der Testpersonen objektiv bewertet werden. Die Darstellung der Testergebnisse erfolgt aktuell über eine Musterlösung am Ende der Bearbeitung einer Vignette innerhalb des digitalen Videotools. Damit können die Teilnehmenden selbstständig überprüfen, ob sie die Fragen richtig beantwortet haben. Perspektivisch ist eine automatisierte, programmierte Punktevergabe im Tool denkbar. Computerbasierte

Auswertungsverfahren haben grundsätzlich eine hohe Auswertungsobjektivität (Kubinger 2009, S. 46). Weiterhin wird durch das geschlossene Antwortsystem im vorliegenden Videotool die Objektivität hinsichtlich der Auswertung erhöht (Eid und Schmidt 2014, S. 66). Da keine offenen Antworten aufwendig und individuell beurteilt werden müssen, ist die Vergleichbarkeit der Ergebnisse höher.

Zur Interpretation der Ergebnisse kann aktuell nur ein Vergleich mit den Ergebnissen des Expertenratings herangezogen werden. Inwieweit die Testwerte der teilnehmenden Personen als „durchschnittlich oder über- bzw. unterdurchschnittlich“ (Bühner 2011, S. 60) beurteilt werden können, kann mit dem aktuellem Stand des Tools nicht erfolgen. Aufgrund des Fehlens einer zweiten repräsentativen Stichprobe kann keine Aussage zur normorientierten Interpretationsobjektivität getroffen werden. Um ein hohes Maß an einer kriteriumsorientierten Interpretationsobjektivität erreichen zu können, wäre es im Nachgang (durch eine Post-hoc-Analyse) möglich, Ankerpunkte für die Kompetenzniveaus mit einer Itemanalyse von Expert:innen durchzuführen (Rauch und Hartig 2020, S. 422). Aufgrund des hohen Aufwands bleibt zunächst der Vergleich mit der maximal möglichen Punktzahl (auf Grundlage des Expertenratings) für die Interpretation der Ergebnisse ausschlaggebend.

Reliabilität

Zur Beurteilung der Genauigkeit der vorliegenden Messung können verschiedene Reliabilitätskriterien betrachtet werden. In der vorliegenden Validierung des Videotools wird ein IRT-Modell zu Grunde gelegt, daher wird der EAP-Reliabilitätskoeffizient als Kriterium der internen Konsistenz herangezogen. Das Testinstrument zur Beurteilung der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung erzielt eine gute Reliabilität (siehe Abschnitt 10.3 *Ergebnisse* der Validierung des Videotools Teil I). Das vorliegende Videotool wird mit einem Teil der Stichprobe im Abstand von vier Wochen erneut eingesetzt. Dadurch ist es möglich, die Retest-Reliabilität durch eine erneute Raschskalierung und einen Vergleich der Modellparameter zu überprüfen. Möglich ist es, dass durch die erste Anwendung des Tests Lerneffekte auftreten und dadurch das Instrument bei der wiederholten Messung nicht mehr das gleiche misst (Rost 2004, S. 378). In der vorliegenden Erhebung muss auf die Paralleltestmethode zur Bestimmung der Paralleltestreliabilität verzichtet werden, da kein passendes Instrument zur Verfügung steht. Zudem ist das entwickelte Videotool bereits sehr umfangreich, sodass ein zusätzlicher Test erhebliche Auswirkungen auf die Motivation und Konzentration der Nutzer:innen hätte (Rost 2004, S. 377) sowie der zeitliche Rahmen einer weiteren Testung schwer umzusetzen wäre.

Validierung

Die Validierung des vorliegenden videobasierten Tools wird hinsichtlich der Augenscheinvalidität, inhaltlichen Validität, Konstrukt- und Kreuzvalidität sowie der Änderungssensitivität beurteilt. Hinsichtlich der Augenscheinvalidität wird bei der Testkonstruktion darauf geachtet, dass den Studierenden bewusst wird, welche Merkmale mit dem Test erfasst werden sollen. Die Fragenstämme des vorliegenden Videotools sind so konzipiert (siehe Abschnitt 8.1 *Entwicklung der Analyseaufträge*), dass den teilnehmenden Personen der Zusammenhang mit dem zu messenden Konstrukt der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung eindeutig auffallen soll. Ein Item der Dimension des Erklärens wird beispielsweise folgendermaßen eingeleitet „Beurteilen Sie, welche der folgenden Erklärungen bezüglich der Motivationsförderung theoretisch angemessen sind. Die Lehrperson verdeutlicht, dass Fehler Teil des Lernprozesses sind“. Die Studierenden erhalten bei der Instruktion zum Test eine genaue Beschreibung (siehe Anhang F) der drei Fragestammkategorien (Beschreiben, Erklären, Vorhersagen). Diese wiederholen sich pro Vignette. Inwieweit die Studierenden diese Form der Augenscheinvalidität des Tests wahrnehmen, wird nicht explizit überprüft.

Die *Entwicklung der Analyseaufträge* erfolgt theoriegeleitet (siehe Abschnitt 8.1). Zur inhaltlichen Validierung findet ein Austausch mit zwei Fachexpert:innen statt. Hinsichtlich der Auswahl inhaltlich passender Videovignetten werden in der Regel ebenfalls Expertenratings durchgeführt (Oser et al. 2010; u.a. Meschede et al. 2015; Seidel et al. 2010; Star und Strickland 2008). Damit soll festgestellt werden, ob die Vignetten tatsächlich das zu untersuchende Merkmal bzw. Thema abbilden. In der vorliegenden Arbeit wird ein anderes Verfahren angewendet, da es sich bei der Motivation um ein hoch-inferentes Konstrukt handelt. Dies bedeutet, dass die tatsächliche Motivation der Schüler:innen von außen schwer zu beobachten ist und am besten durch eine individuelle Rückmeldung der einzelnen Schüler:innen erfasst werden kann (Krapp und Hascher 2014a, S. 234). Dazu wird ein Fragebogen eingesetzt und die tatsächliche Motivation der Schüler:innen ausgewertet (siehe Abschnitt 7.2 *Methodisches Vorgehen* der Entwicklung und Validierung der Videovignetten). Es können somit einzelne Unterrichtssequenzen ausgewählt werden, die für die Schüler:innen tatsächlich motivationsförderlich oder -hemmend sind. Dies dient der Sicherstellung der inhaltlichen Validität der Vignetten über Fragebogenerhebungen.

Wie bereits beschrieben, basiert das vorliegende Videoinstrument auf der theoretischen Annahme einer dreigliedrigen Dimensionalität. Es wird bezüglich der faktoriellen Validität erwartet, dass sich die Professionelle Unterrichtswahrnehmung hinsichtlich des Beschreibens, Erklärens und Vorhersagens trennen lässt. Aus diesem Grund wird ein strukturprüfender Ansatz gewählt. Unter Verwendung von IRT-Modellen „können die Beziehungen zwischen den Testitems und quantitativen latenten

Konstrukten statistisch überprüft werden“ (Moosbrugger und Kelava 2020a, S. 35). Die Vorgehensweise der Überprüfung der Konstruktvalidität kann im Kapitel 10 *Validierung des Videotools – Haupterhebung Teil I* nachvollzogen werden. Ein Vergleich zwischen verschiedenen Modellen mit jeweils unterschiedlicher Dimensionalitätsannahme soll die anzunehmende dimensionale Struktur hervorbringen. Die konvergente und divergente Validierung wird aufgrund fehlender Vergleichstests und aus ökonomischen Gründen nicht überprüft. Ebenfalls können im Rahmen der vorliegenden Arbeit keine Aussagen zur Kriteriumsvalidität getroffen werden. Zusätzliche Variablen wie Erfahrungen im Lehren und Lernen mit Schüler:innen, in der Bearbeitung eines Videotools, im Themenkomplex Motivation werden erhoben, aber aus Gründen der Komplexität an dieser Stelle nicht ausgewertet. Stattdessen wird das Semester, in dem sich die Studierenden zum Testzeitpunkt befinden, in die Auswertung einbezogen. Forschungsarbeiten zur Professionellen Unterrichtswahrnehmung zeigen, dass Studierenden das Beschreiben von Situationen besser gelingt als das Erklären oder Vorhersagen von Auswirkungen (Stürmer et al. 2015, S. 352; Schäfer und Seidel 2015, S. 50–51; Meschede et al. 2015, S. 328; Vgl. Frommelt et al. 2016, S. 369) (siehe Kapitel 11 *Validierung des Videotools Teil II*, theoriegeleitete Konstruktvalidität). Ebenfalls soll überprüft werden, ob sich die Studierenden im Verlauf des Studiums hinsichtlich der Professionellen Unterrichtswahrnehmung verbessern (siehe Kapitel 12 *Validierung des Videotools Teil III*, gruppendifferenzierte Konstruktvalidität). In diesem Zusammenhang wäre die zusätzliche Erhebung des pädagogisch-psychologischen Wissens interessant, um feststellen zu können, ob die Studierenden mit einem hohen Wissen über die Motivation eine höhere Fähigkeit der professionellen Wahrnehmung hinsichtlich der Motivierung im Unterricht aufweisen. Ob die Ergebnisse des raschskalierten Modells replizierbar sind, soll mit einer Kreuzvalidierung festgestellt werden. Dazu wird die vorhandene Stichprobe in zwei gleichgroße Testhälften geteilt. Die erste Stichprobe wird zur Modifizierung des Modells genutzt. Anhand der Ergebnisse der erneuten Raschskalierung mit der zweiten Stichprobe soll ein Vergleich der Modellgüte Aufschluss über die Kreuzvalidität geben (siehe Abschnitt 10.3 *Ergebnisse der Validierung des Videotools Teil I*).

Nebenkriterien

Inwieweit testrelevante Faktoren die Fairness des Tests beeinflussen, kann nicht geprüft werden, da im Rahmen der vorliegenden Validierungsschritte des Videotools keine personenspezifischen Merkmale für eine DIF-Analyse erhoben werden. Möglichen Einschränkungen in der Durchführungsfairness bei der Anwendung des Videotools soll sowohl durch eine schriftliche als auch videobasierte Anleitung begegnet werden. Ein weiterer Einflussfaktor kann eine Testroutine sein (Moosbrugger und Kelava 2020a, S. 26). Da es sich beim vorliegenden Videotool um eine Adaption des

bisherigen ViviAn-Tools, bezogen auf Thema, Theoriekonzept, Fragenkonstruktion oder Perspektivebene handelt, wird nicht von einem Wiederholungseffekt ausgegangen.

Die Nützlichkeit des Tests wird mit einer hohen Relevanz des Inhalts begründet. Diese basiert auf einer hohen Theorie-Praxis-Verzahnung im Lehramtsstudium sowie der Förderung und Analyse der Professionellen Unterrichtswahrnehmung hinsichtlich bedeutsamer motivationsrelevanter Aspekte im Unterricht. Zusätzlich zur theoretischen Bedeutsamkeit wird die wahrgenommene Relevanz des Instruments durch die Studierenden untersucht (siehe Kapitel 9 *Akzeptanz des Tools ViviAn*). Zur Erfüllung des Gütekriteriums der Nützlichkeit ist neben der Relevanz auch das Alleinstellungsmerkmal des Instruments bzw. die Abgrenzung zu einem bestehenden Instrument wichtig. Wie bereits erwähnt, stellt der OBSERVER (Seidel et al. 2010) ein ähnliches Messinstrument dar. Die Parallelitäten sowie die klare Abgrenzung bzw. Erweiterung zum OBSERVER werden im Abschnitt 4.4 in der *Zusammenfassung in Bezug auf die vorliegende Arbeit* zum Einsatz von Videos in der Lehre beschrieben. Zur Beurteilung der Gütekriterien wird der OBSERVER an entsprechender Stelle als Vergleichsinstrument herangezogen. Der Fokus beim OBSERVER liegt pro Vignette auf den drei Schwerpunkten Zielklarheit, Unterstützung durch die Lehrperson und Lernklima (Seidel und Stürmer 2014, S. 743). Diese beinhalten Aspekte zu Lehr-/Lernzielen, Strukturierung, Fragen und Feedback der Lehrperson sowie Wertschätzung der Schüler:innen und Humor der Lehrperson (Seidel und Stürmer 2014, S. 743). Das vorliegende Videoinstrument fokussiert explizit einzelne motivationsrelevante Aspekte in jeder Vignette. Das bedeutet, die Professionelle Unterrichtswahrnehmung kann im ViviAn-Tool hinsichtlich der Merkmale Leistungsdifferenzierung, individuelle Bezugsnormorientierung, Relevanz, Unterstützung der Freiräume, adäquate Strukturiertheit und Fehlerklima spezifischer erfasst werden. Zudem können im Videotool ViviAn vor und während der Betrachtung der Vignette relevante Zusatzinformationen abgerufen werden. Hinsichtlich der methodischen Auswertung kann aufgrund der konsequenten Überprüfung unter Verwendung eines Rasch-Modells (siehe Kapitel 10 *Validierung des Videotools Teil I*) das Gütekriterium der Skalierung abgebildet werden.

Die Verwendung eines digitalen Videotools bietet hinsichtlich der Ökonomie mehrere Vorteile. Da alle Dokumente digital zur Verfügung stehen, verbraucht das ViviAn-Tool kein zusätzliches Testmaterial. Die Anwendungsoberfläche wird stets optimiert, sodass die Bedienung intuitiv und einfach ist (siehe Abschnitt 4.3 *Das Videotool ViviAn*). Eine aufwendige Einarbeitung des Instruments ist daher nicht erforderlich. Da das Tool selbstständig bedient und angewendet werden kann, ist keine zusätzliche Person für die Durchführung notwendig, lediglich für die Freischaltung des Zugangs (Roth 2020, S. 70). Die Reflexion der Ergebnisse erfolgt ebenfalls eigenständig durch die Nutzer:innen auf Basis einer automatisch zur Verfügung gestellten Lösung am Ende einer Vignette (Roth 2020, S. 70). Der Zeitaufwand der Bearbeitung des

Videotools ist im Vergleich zum OBSERVER etwas geringer. Die Länge der Videosequenzen beim OBSERVER betragen zwischen zwei und vier Minuten (Seidel und Stürmer 2014, S. 749), wohingegen die Videoausschnitte im vorliegenden Tool alle kürzer als zwei Minuten sind. Die Testlänge des OBSERVERs ist mit sechs Vignetten und insgesamt 112 Items (Seidel und Stürmer 2014, S. 752) ähnlich. Wobei zu berücksichtigen gilt, dass nicht die gleichen Merkmale hinsichtlich der Professionellen Unterrichtswahrnehmung betrachtet werden (siehe Gütekriterium Nützlichkeit). Beide Testverfahren verwenden ein geschlossenes computerbasiertes Antwortverfahren. Diese beanspruchen im Vergleich zu Tests mit offenen Antwortformaten weniger Zeit (Bühner 2011, S. 115). Der finanzielle Aufwand für folgende Nutzungen des Videotools ist gering. Es kann nach entsprechender Anmeldung von interessierten Personen aktuell kostenlos genutzt werden. Allerdings muss berücksichtigt werden, dass die Wartung, Instandhaltung und Betreuung der ViviAn-Website selbstverständlich personelle Ressourcen beansprucht, die entsprechend vergütet werden müssen.

Da das vorliegende Instrument so konzipiert ist, dass die einzelnen zu analysierenden Vignetten unterschiedliche Lernsequenzen abbilden, ist die Bearbeitung der Vignetten voneinander unabhängig. Dies hat den Vorteil, dass die Nutzer:innen nach jeder Vignette pausieren können, sofern sie eine kognitive Belastung wahrnehmen.

Zur abschließenden Beurteilung der Nebenkriterien wird auf die Normierung eingegangen. Eine Eichung mit der vorliegenden Stichprobe (N = 239) ist nicht empfohlen. Ebenfalls liegt keine Vergleichsstichprobe vor, um die gewonnenen Daten anhand dieser Testnormierung interpretieren zu können.

Der Schwerpunkt der vorliegenden Arbeit liegt in der Entwicklung und Validierung des Videotools ViviAn zur Testung und Förderung der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung. Anhand der Gütekriterien werden im nächsten Kapitel die Forschungsfragen der vorliegenden Arbeit formuliert.

Teil 2

Eigene empirische Erhebung

6 Zielsetzung und Fragestellung

Im Unterrichtsalltag stehen die Lehrpersonen vor einer Vielzahl an Herausforderungen. Ein wesentlicher Aspekt ist die Wahrnehmung, Beschreibung und Erklärung relevanter Situationen im komplexen Unterrichtsgeschehen auf der Grundlage professionellen Wissens, mit dem Ziel Vorhersagen für künftige Entscheidungen treffen bzw. den aktuellen Unterricht adaptieren zu können (Seidel et al. 2010, S. 297; Sherin und van Es 2008, S. 22). Bei diesem Prozess handelt es sich um die Professionelle Unterrichtswahrnehmung (siehe Kapitel 3), die in Bezug auf diverse Aspekte der Prozessqualität erforderlich ist. Dazu gehören z.B. die Klassenführung (Vgl. Hellermann et al. 2015), lernunterstützende Maßnahmen (Seidel und Stürmer 2014; Vgl. Meschede et al. 2015) oder Gruppenarbeitsprozesse (Vgl. Bartel und Roth 2017) (siehe Kapitel 4.2 *Studienlage* zum Einsatz von Videos und 4.3 *Das Videotool ViviAn*). Der spezielle Fokus der Professionellen Unterrichtswahrnehmung in der vorliegenden Arbeit liegt auf dem Bereich motivationsrelevanter Aspekte. Die *Motivation und Motivierung im Unterricht* (siehe Kapitel 2) ist aus vielerlei Hinsicht für das Lernen bedeutsam. Die Lehrperson hat mit der Gestaltung des Unterrichts einen Einfluss auf die Motivation der Lernenden und kann damit den Lernprozess und die Lernentwicklung initiieren (Kunter et al. 2011, S. 59). Als Prädiktor für die Schulleistung kommt der Lernmotivation eine hohe Bedeutung zu (Steinmayr und Spinath, 2007). Das Herausfordernde für Lehrpersonen ist, die tatsächliche Motivation der Schüler:innen wahrnehmen zu können, da es sich um ein hoch-inferentes Konstrukt handelt (Büttner 2008, S. 288). Auf der Grundlage einer solchen Wahrnehmung kann die Situation kognitiv verarbeitet und eine adäquate Reaktion darauf geplant und umgesetzt werden. Dies bedeutet, dass nicht nur das professionelle Wissen über den Einsatz motivationsfördernder Maßnahmen für einen gelingenden Lernprozess entscheidend ist, sondern auch die professionelle Wahrnehmung der Situation. Bereits im Studium ist die Verwendung von videobasierten Instrumenten zur Analyse und Förderung der Professionellen Unterrichtswahrnehmung sinnvoll (siehe Kapitel 4 *Einsatz von Videos in der Lehre*). Bei der Entwicklung eines solchen Instruments muss zunächst sichergestellt werden, dass es sich um eine geeignete Lernumgebung handelt. Diese muss in der Lage sein, einen Lernzuwachs verzeichnen zu können. Dafür benötigt man ein Testinstrument. Die vorliegende Arbeit nimmt sich zum Ziel, ein valides und standardisiertes videobasiertes Testinstrument zur Analyse und Förderung der Professionellen Wahrnehmung zu entwickeln. Dazu wird das bestehende Video-Tool *ViviAn* („Videovignetten zur Analyse von Unterrichtsprozessen“) an der RPTU (siehe Abschnitt 4.3. *Das Videotool ViviAn*) erweitert. Die Ergänzung und Anpassung von *ViviAn* umfasst Vignetten und Analyseaufträge zu motivationsrelevanten Aspekten. Im Rahmen dieser Arbeit wird überprüft, ob *ViviAn* die Anforderung an ein Testinstrument erfüllen kann. Als Kriterien dienen verschiedene *Gütekriterien der Testentwicklung* (siehe Kapitel 5). Diese werden kurz beschrieben und mit entsprechender Forschungsfrage für die anschließende empirische Überprüfung skizziert.

Inhaltliche Validität:

Die Videovignetten sollen inhaltlich geeignete Situationen darstellen, die motivationsförderliches oder -hemmendes Handeln der Lehrperson im Unterricht abbilden. Inwieweit dieses Handeln in den gefilmten Unterrichtssequenzen tatsächlich zur Steigerung oder Verminderung der Motivation beiträgt, soll eine Befragung der Schüler:innen zu ihrer aktuellen Motivation in dieser Unterrichtseinheit klären (siehe Abschnitt 7.2.2 *Validierung der Videovignetten*). Dabei muss sich die Motivation der Schüler:innen mit dem motivationsförderlichen Handeln der Lehrperson decken.

1. *Stellen die Vignetten repräsentative motivationsförderliche oder -hemmende Situationen dar?*

Auswertungsobjektivität:

Die Analyseaufträge in Form von Rating-Items werden zur Beurteilung der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung ausgewertet. Die Beurteilung, ob eine Antwort richtig, tendenziell richtig oder falsch ist, basiert auf der Grundlage eines Expert:innenratings. Kommen die Rater:innen bei der Beantwortung der Items auf eine hohe Übereinstimmung, kann von einer hohen Auswertungsobjektivität ausgegangen werden. Dabei dient der Richtwert für ein Fleiß-Kappa > 0.61 (Landis und Koch 1977, S. 164–165) als akzeptables Kriterium.

2. *Ist die Übereinstimmung der Expert:innen bei der Auswertung der Rating-Items ausreichend hoch, um von einer hohen Auswertungsobjektivität ausgehen zu können?*

Nützlichkeit:

Das Instrument erweist sich als nützlicher Test, wenn es eine hohe Relevanz für die Praxis aufzeigen kann (Döring und Bortz 2016, S. 449; Lienert und Raatz 1998, S. 13; Gold und Hasselhorn 2013, S. 405). Die wahrgenommene Relevanz für das vorliegende ViviAn-Tool wird durch eine Befragung der Teilnehmer:innen erhoben (siehe Kapitel 9. Neben der Relevanz spielt das situationsbezogene Interesse für die intrinsische Motivation eine wichtige Rolle (Wild und Möller 2015, S. 162). Daher werden beide Aspekte zur Beantwortung der folgenden Forschungsfrage herangezogen.

3. *Können die Vignetten zum Themenbereich Motivierung bei den Studierenden das Interesse an der Arbeit mit Videos wecken sowie eine hohe Relevanz wahrnehmen lassen?*

Interne Konsistenz:

Das neu entwickelte Testinstrument soll mit seinen Items reliabel die Professionelle Unterrichtswahrnehmung hinsichtlich der Motivierung messen. Dazu wird die interne Konsistenz als Kriterium der Reliabilität ermittelt. Diese soll Auskunft über die Homogenität der Items geben (Gold und Hasselhorn 2013, S. 403). Die Auswertung des vorliegenden Instruments basiert auf der Item-Response-Theorie. Als Kriterium für die Interne Konsistenz werden hierbei die EAP-Reliabilitätskoeffizienten betrachtet (siehe Abschnitt 10.2.2 Raschskalierung), die bei einem Testinstrument mit $R_{EAP} > 0.7$ für eine reliable Messung sprechen (Vgl. Field et al. 2012, S. 806).

4. *Bilden die Items die Skalen des Knowledge-based Reasonings reliabel ab?*

Skalierung:

Um die Passung des theoretisch angenommenen Modells mit den vorliegenden Daten überprüfen zu können, werden die Itemfitwerte (Infit und Outfit) aus der Rasch-Skalierung (siehe Abschnitt 10.2.2) herangezogen (Bond und Fox 2012, S. 238). Als Richtwert für eine gute Passung gelten Itemfitwerte $\text{Infit}_i / \text{Outfit}_i = 0.7 - 1.3$ (Bond und Fox 2012, S. 310).

5. *Zeigen die Items modellkonforme Itemfitwerte?*

Konstruktvalidität des Instruments:

Als ein Kriterium der Konstruktvalidität wird die theoretisch angenommene Struktur und Dimensionalität des Tests überprüft (Moosbrugger und Kelava 2020a, S. 33). In bisherigen Forschungsarbeiten zur Förderung der Professionellen Unterrichtswahrnehmung durch den Einsatz von Videos kann bezüglich des Nachweises der Struktur des Reasonings sowohl eine drei-dimensionale (u.a. Seidel und Stürmer 2014, S. 759) als auch eine eindimensionale Modellannahme (Meschede et al. 2015, S. 326) bestätigt werden. Modellvergleiche zwischen verschiedenen Dimensionalitätsannahmen sollen die bestpassende Struktur aufdecken (siehe Abschnitt 10.3).

6. *Differenziert das vorliegende Instrument zwischen den drei Phasen des Knowledge-based Reasonings: Beschreiben, Erklären und Vorhersagen?*

Als weiteres Kriterium der Validität soll die nachgewiesene Modellstruktur des Instruments an einer weiteren Stichprobe beobachtbar sein.

7. *Lassen sich die Ergebnisse an einer zweiten Stichprobe replizieren?*

Bisherige Forschungsarbeiten (Stürmer et al. 2015, S. 352; Schäfer und Seidel 2015, S. 50–51; Meschede et al. 2015, S. 328; Vgl. Frommelt et al. 2016, S. 369) belegen, dass sowohl Studierende als auch Referendar:innen die höchste Fähigkeit innerhalb des Reasonings im Beschreiben vorweisen können, gefolgt vom Erklären und Vorhersagen. Das vorliegende Instrument soll für den Nachweis einer theoriegeleiteten Konstruktvalidität in der Lage sein, die angenommenen Unterschiede in den Analysefähigkeiten innerhalb des Reasonings abzubilden (siehe Kapitel 11 *Validierung des Videotools Teil II*).

8. *Sind die angenommenen Schwierigkeitsgrade innerhalb des Reasonings mit dem Instrument überprüfbar?*

Zudem soll es als valides Instrument zwischen Studierenden in unterschiedlichen Phasen des Studiums differenzieren können (gruppendifferenzielle Konstruktvalidität). Studienbasierte Vergleiche zwischen Bachelor- und Masterstudierenden indizieren, dass die Fähigkeit der Professionellen Unterrichtswahrnehmung zu Beginn des Studiums geringer als am Ende des Studiums ausgeprägt ist (u.a. Meschede et al. 2015, S. 330). Deswegen wird in der vorliegenden Arbeit von einer Zunahme der Fähigkeit im Verlauf des Studiums ausgegangen (siehe Kapitel 11).

9. *Kann mit ViviAn eine Entwicklung in der Ausprägung der Professionellen Unterrichtswahrnehmung zu Beginn des Studiums im Vergleich zum Ende des Studiums erfasst werden?*

Änderungssensitivität

Mit der Durchführung einer Intervention zwischen einer Pre- und Posttestung sollen Rückschlüsse auf die Änderungssensitivität des Instruments gezogen werden (siehe Kapitel 12 *Validierung des Videotools Teil III*). Zwischen den beiden Messzeitpunkten erfolgt ein online-basiertes Seminar zum Thema „Lernumgebungen gestalten“, das den Fokus auf die Vermittlung professionellen Wissens motivationsrelevanter Aspekte im Unterricht anhand von praktischen text- und videobasierten Beispielen mit Analyse- und Reflektionsaufträgen legt. Diese Intervention soll sich in den Testergebnissen positiv widerspiegeln.

10. *Erweist sich ViviAn als geeignetes Instrument, um eine Entwicklung der Professionellen Wahrnehmung hinsichtlich der Motivierung durch eine Intervention aufzuzeigen?*

7 Entwicklung und Validierung der Videovignetten

Zentrale Elemente des videobasierten Tools zur Analyse und Förderung der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung stellen die Videovignetten dar. Die *Entwicklung und Validierung der Videovignetten* entstehen in Zusammenarbeit mit einer Lehrperson und ihren Schüler:innen aus der Sekundarstufe I (siehe Abschnitt 7.1 *Stichprobe*). Das Ziel ist die Entwicklung und Auswahl von Videovignetten für das videobasierte Instrument. Um geeignete Vignetten aus den videografierten Unterrichtseinheiten auswählen zu können, werden in der vorliegenden Studie die Urteile der Schüler:innen zu ihrer aktuellen Motivation während der Filmeinheiten analysiert und zur Validierung herangezogen. Damit soll die erste Forschungsfrage zur Überprüfung der inhaltlichen Validität beantwortet werden:

Forschungsfrage

- 1 *Stellen die Vignetten repräsentative motivationsförderliche oder -hemmende Situationen dar?*

Das *methodische Vorgehen* und die Auswertung werden in den Abschnitten 7.2 und 7.3 beschrieben. Die daraus resultierenden Vignetten sind zum Überblick im Abschnitt 7.5 *Beschreibung der Vignetten* zusammengefasst. Um aus den Videovignetten ein Analysetool generieren zu können, bedarf es im nächsten Schritt der *Entwicklung und Validierung der Analyseaufträge* (siehe Kapitel 8).

7.1 Stichprobe

Im Rahmen des Projektes werden zwei Schulklassen einer Realschule plus in Rheinland-Pfalz im Fach Erdkunde begleitet. Der Zeitraum der Erhebung erstreckt sich über sieben Monate im ersten Schulhalbjahr 2017/2018. Am Projekt beteiligen sich in der 7.Klasse 14 Schülerinnen und Schüler (davon 4 weiblich), in der 8.Klasse 20 Schülerinnen und Schüler (davon 8 weiblich). Der Fachunterricht wird in beiden Klassen von derselben Lehrperson durchgeführt.

7.2 Methodisches Vorgehen

Das methodische Vorgehen wird in zwei Teilabschnitten beschrieben: die Entwicklung (siehe Abschnitt 7.2.1) und die Validierung (siehe Abschnitt 7.2.2) der Videovignetten. Im ersten Schritt erfolgt die Beschreibung des Prozesses zur Videografie im Unterricht inklusive der Adaption in Bezug auf die Unterrichtsplanung. Die Validierung der so entstandenen Videoausschnitte wird im zweiten Schritt betrachtet.

7.2.1 Videografie mit Unterrichtsadaption

Zur Erstellung eines videobasierten Instruments zur Förderung der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung im Unterricht sollen authentische Filmaufnahmen aus dem Schulunterricht verwendet werden. Dadurch kann die reale Praxis abgebildet werden (von Aufschnaiter et al. 2017, S. 88; Syring et al. 2015, S. 670). In beiden beschriebenen Klassen wird der Unterricht im Fach Erdkunde videografiert. Insgesamt werden 34 Unterrichtseinheiten von jeweils 45 Minuten (davon 16 in der 8.Klasse, 18 in der 7.Klasse) im Zeitraum von November 2017 bis Februar 2018 mit insgesamt drei Kameras aufgezeichnet. „Damit wird es möglich, das jeweilige Unterrichtsgeschehen mit den entsprechenden Interaktionen zwischen der Lehrperson und den Lernenden möglichst realitätsnah und mitsamt seiner Komplexität und Substanz in verkürzter Zeit abzubilden“ (Oser et al. 2010, S. 12). Im Vorfeld plant die Lehrperson ihren Unterricht entsprechend den Lehrplanvorgaben. Zusätzlich verschriftlicht sie ihre Unterrichtskonzeptionen für die Erhebungen.

Ziel ist die Entwicklung eines Instruments, das den Nutzer:innen eine möglichst große Varianz an motivationsförderlichen und -hemmenden Aspekten im Unterricht aufzeigt. Die Videosequenzen sollen zunächst im Rahmen eines Seminareinsatzes an der Universität und im Referendariat analysiert und interpretiert werden. Aufgrund der ökonomisch begrenzten Ressourcen und zur Erzielung eines höheren Gewinns an optimal nutzbarem Videomaterial werden der Lehrperson zusätzlich zu ihrer eigentlichen Unterrichtsplanung unterschiedliche Impulse zum Handeln hinsichtlich der Motivierung gegeben. Damit handelt es sich dennoch um authentische, aber modifizierte Vignetten (von Aufschnaiter et al. 2017, S. 96) (siehe Abschnitt 4.1 *Die fünf Heuristiken*). Um dabei einem Authentizitätsverlust entgegenzuwirken, wird die schriftliche Planung der Lehrperson als Ausgangspunkt verwendet. Die Instruktionen zur Hervorhebung bestimmter motivationsrelevanter Aspekte erfolgt auf Grundlage dieser Unterrichtsplanungen (Beispiel siehe Abbildung 16). Dadurch können entsprechend des geplanten Inhalts und der Methode für diese Unterrichtseinheit passende motivationsrelevante Kriterien ausgewählt werden. Als theoretisches Modell wird dazu das IBAS-Modell (siehe Abschnitt 2.2. *Motivierung durch Zielorientierung*) herangezogen. Dieses Modell wird der Lehrperson im Vorfeld ausführlich erklärt. Somit können einzelne Facetten motivationalen Handelns (als Subdimensionen bezeichnet) fokussiert werden. Entsprechend der Unterrichtsplanung der Lehrperson mit den Inhalten und der ausgearbeiteten Methode werden geeignete Subdimensionen aus dem IBAS-Modell ausgewählt und Umsetzungsvarianten vorgeschlagen. Zusätzlich zu den schriftlich festgehaltenen Impulsen zur Anpassung ihres motivierenden Handelns werden mit der Lehrperson vor der Durchführung des Unterrichts einzelne Punkte ausführlich besprochen. Dabei werden noch einmal die Subdimensionen mit den aufgelisteten Beispielen beschrieben, Unklarheiten geklärt und weitere Beispiele erläutert. In jeder Phase des Unterrichts werden ein bis vier Subdimensionen instruiert.

Exogene Kräfte (verschiedene Beispiele)

TEIL 1: Einstieg und Erarbeitung

POSITIVES, KONSTRUKTIVES FEHLERKLIMA ↑

- Fehler sind erlaubt und gehören zum Lernprozess, Fehler sind nicht schlimm, niemand wird dafür ausgelacht oder genervt angeguckt, keine Wertung
- Bei Fehlern unterstützen (z.B. Verknüpfung mit vorhandenem Wissen unterstützen, an Beispielen erklären)
- Die Fehler sollen genau untersucht werden, Denkprozesse beschreiben lassen (wie ist jemand zu der Lösung gekommen), Fehler reflektieren lassen
- Fehler sollen genutzt werden, um daraus zu lernen oder den Stoff wirklich zu verstehen

Einstieg: Wir haben uns als Beispiel zu den exogenen Kräften die Auswirkung des Windes näher betrachtet. Beschreibt bitte nochmals mit eigenen Worten, was man eigentlich unter exogenen Kräften versteht.

Ein Tafelbild wird von den Schülerinnen und Schülern ins Heft übernommen, das die exogenen Kräfte erläutert.

Erarbeitung: Die Lerngruppe wird in Gruppen eingeteilt. Jede Gruppe erstellt ein Plakat zum Thema, das ihr zugeteilt wird. Bevor sich die Gruppen zusammensetzen, werden die Plakatregeln mit ihnen besprochen und ins Heft übernommen.

Abbildung 16 Unterrichtsplanung der Lehrperson mit ergänzter Instruktion (oberes blaues Textfeld)

In diesem Beispiel wird der Fokus auf die Subdimension Positives, konstruktives Fehlerklima (aus der Bewertungs-Dimension) gelegt. Die Lehrperson erhält Hinweise, inwiefern sie diesen Bereich in der ersten Unterrichtsphase motivationsförderlich gestalten kann.

Eine Anpassung des Unterrichts kann einen Authentizitätsverlust bewirken. Die Schüler:innen sollen nicht den Eindruck erhalten, dass die Lehrperson im Rahmen des Projektes ihr Vorgehen im Unterricht sehr verändert und der Situation anpasst. Um dem entgegenzuwirken, werden für die Instruktionen zum Hervorheben bestimmter motivationsrelevanter Aspekte im Vorfeld typische Handlungsmuster der Lehrperson hinsichtlich motivationsförderlicher und eher motivationshemmender (Anmerkung im Textverlauf beachten) Maßnahmen aufgedeckt. Dazu wird vor der Erhebung der Unterricht der Lehrperson ausführlich beobachtet, mittels Hospitationsprotokollen dokumentiert und ausgewertet (siehe Anhang C, D). Auf Grundlage der Analysen werden die motivationsförderlichen und -hemmenden Aspekte, die von der Lehrperson in dieser Zeit eingesetzt werden, herausgestellt und für die Anpassung der Unterrichte verwendet.

Zur Erläuterung ein Beispiel:

Die Lehrperson verwendet in ihrem Unterricht gelegentlich Fahrpläne für die Stunde, die sie an der Tafel notiert. Während des Unterrichts nimmt sie Bezug zum Fahrplan, verdeutlicht der Klasse, welche Punkte bereits abgearbeitet wurden und wo sie sich derzeit befinden. Diese Gestaltung des Unterrichts wird als positives Beispiel zur Förderung der Strukturierung des Kontextes (aus der Autonomie-Dimension) für die Anpassung verwendet.

Wichtige Anmerkung: Im Rahmen der motivationshemmenden Maßnahmen handelt es sich zu jedem Zeitpunkt um ethisch vertretbares Handeln der Lehrperson. Dabei stellen hemmende Maßnahmen insbesondere Maßnahmen dar, die entsprechend der Theorie und Forschung als weniger förderliche Maßnahmen gelten. Es handelt sich eher um das Weglassen förderlicher Aspekte. Grundsätzlich wird darauf geachtet, dass der Umgang nicht-diskriminierend und frei von Gewalt ist. Das Lernen soll nicht nachträglich negativ gehemmt werden.

Zur Erläuterung ein Beispiel:

Die Einführung zum Thema „Exogene Kräfte am Beispiel eines Tornados“ soll hinsichtlich der Relevanz motivationshemmend gestaltet werden. Die Lehrperson soll während der ersten Filmsequenz keinen Alltagsbezug herstellen, nicht nach eigenen Erfahrungen fragen, die Bedeutsamkeit des Themas nicht hervorheben und auch nicht begründen, warum sich die Schüler:innen mit dem Thema beschäftigen sollen. Um den Schüler:innen keinen Nachteil im Lernen zu verschaffen, wird der Aspekt der Relevanz des Themas im Anschluss noch verdeutlicht und der Alltagsbezug und die Bedeutsamkeit mit den Schüler:innen besprochen.

Im Anschluss an die komplette Erhebung werden die Schüler:innen über den Einsatz der gezielten motivationsförderlichen und „motivationshemmenden“ Situationen aufgeklärt. Dies kann im Vorfeld nicht erfolgen, da davon ausgegangen werden muss, dass damit die aktuelle Lernmotivation beeinflusst wird. Diese ist relevant für die Validierung der Videovignetten (siehe folgender Abschnitt)

7.2.2 Validierung der Videovignetten

Zur Validierung von inhaltlich passenden Videoausschnitten werden in Forschungsarbeiten meist Expert:innenratings durchgeführt (Oser et al. 2010; u.a. Meschede et al. 2015; Seidel et al. 2010; Star und Strickland 2008). Anhand der Motivationstheorien und -modelle kann der Einfluss der Unterrichtsgestaltung auf die Motivation der Schüler:innen differenziert beschrieben werden. Ein entsprechend entwickeltes Kategoriensystem

könnte zur Validierung der Vignetten durch Expert:innen herangezogen werden. Da es sich bei der Motivation jedoch um ein hoch-inferentes Konstrukt handelt (Büttner 2008, S. 288), kann die tatsächliche Motivation der Schüler:innen nur durch eine Selbsteinschätzung ermittelt werden. Daher wird im Projekt nicht auf ein Expert:innenrating gesetzt, sondern ein reliables, valides und standardisiertes Instrument zur Erfassung der aktuellen Motivation (Prenzel et al. 1996) der Schüler:innen verwendet (siehe Anhang B). Die Auswahl dieses Instruments basiert auf der Analyse verschiedener Instrumente zur *Messung der Motivation im Lernkontext* (siehe Abschnitt 2.3 und Anhang A). Das ausgewählte Instrument muss bei wenigen Items für die Erhebung in der Schule angepasst werden, da es ursprünglich für den Ausbildungsbereich konzipiert wurde. Daher wird zu Beginn eine Analyse zur Überprüfung der internen Konsistenz der Skalen durch die Ermittlung des Cronbachs Alpha durchgeführt (Gold und Hasselhorn 2013, S. 403). Erfasst werden die fünf Faktoren amotiviert, external, introjiziert, identifiziert und intrinsisch motiviert (siehe Selbstbestimmungstheorie im Abschnitt 2.1 *Motivation durch Bedürfnisbefriedigung*) mit jeweils drei Items. Die Schüler:innen beantworten die Items online auf einem Tablet durch Ankreuzen auf einer fünfstufigen Likert-Skala (0 - stimmt gar nicht, 1 - stimmt eher nicht, 2 - teils teils, 3 - stimmt eher, 4 - stimmt genau) (Beispiel siehe Abbildung 17). Pro Unterrichtseinheit von 90 Minuten wird zweimal die Motivation der Lernenden erfasst. Die Entscheidung für zwei Erhebungen wird folgendermaßen begründet:

- 1) Um die Motivation im Unterricht auf Basis des motivationsfördernden oder -hemmenden Handelns der Lehrperson erfassen zu können, muss zunächst eine gewisse Unterrichtsdauer stattfinden. Zudem soll eine möglichst hohe Authentizität und Realität erreicht werden (Syring et al. 2015, S. 670; Yadav et al. 2011, S. 32). Dies bedeutet, dass die Lernenden den Unterricht als alltäglich wahrnehmen sollen. Eine große Anzahl von Erhebungen kann dies voraussichtlich negativ beeinflussen.
- 2) Je länger der Zeitraum ist, desto schwieriger wird es, die tatsächliche Motivation auf das forcierte Handeln der Lehrperson beziehen zu können. Daher werden keine zusätzlichen Messzeitpunkte gewählt.

Mit Abzug der Vorbereitungszeit und der Erhebungen beträgt die Dauer einer Unterrichtsphase circa 20 Minuten. Insgesamt liegen 34 Datenerhebungen für die Auswertung vor. Die Bearbeitungsdauer des Fragebogens liegt bei durchschnittlich drei Minuten. Im Vorfeld kamen weitere Fragebögen zur Erfassung soziodemografischer Daten, des Fachinteresses sowie der Lern- und Leistungsmotivation (SELLMO) zum Einsatz. Zur Validierung der Vignetten werden diese nicht herangezogen. Bei der *Diskussion* (Kapitel 14) werden einzelne Ergebnisse daraus herangezogen.

Aktuelle Lernmotivation
Später fortfahren Umfrage beenden und löschen

50%

Beim Lernen/Arbeiten in dieser Unterrichtsphase...

	stimmt gar nicht	stimmt eher nicht	teils teils	stimmt eher	stimmt genau
... machte das Arbeiten/Lernen richtig Spaß.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... war ich neugierig/wissbegierig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...verging die Zeit wie im Flug.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...wollte ich selbst den Stoff verstehen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...war mir klar, dass ich das später brauchen kann/können muss.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 17 Auszug aus dem Fragebogen zur Erfassung der aktuellen Motivation

Die Datenauswertung zur aktuellen Motivation erfolgt zunächst über eine deskriptive Statistik. Für jeden Messzeitpunkt werden getrennt nach den fünf Faktoren die Mittelwerte bestimmt. Für jede Klasse wird zudem der Gesamtmittelwert pro Faktor, ermittelt über alle Messzeitpunkte, berechnet. Dieser dient als Kriterium zur Feststellung von Abweichungen der Mittelwerte des jeweiligen Faktors für einen Messzeitpunkt. Somit soll bestimmt werden, ob beispielsweise die intrinsische Motivation zum Messzeitpunkt fünf höher oder niedriger ausgeprägt ist als im Durchschnitt über alle Messzeitpunkte hinweg. Um deutliche Unterschiede aufzeigen zu können, wird eine Abweichung von 0.3 als Mindestabweichung festgelegt. Dabei kann die Abweichung vom Gesamtmittelwert sowohl positiv als negativ ausfallen. Für die weiteren Stufen gilt eine zusätzliche Abweichung von 0.1. In der Tabelle 6 ist die Einteilung der Abweichungsstufen mit Zuordnung einer Farbcodierung für eine bessere Veranschaulichung der Ergebnisse dargestellt. Zu berücksichtigen ist, dass bei der extrinsischen Motivation eine gegenpolige Auswertung erfolgen muss. Die Farbcodierungen sollen einen ersten Überblick über Unterschiede in der aktuellen Lernmotivation der Schüler:innen geben. Im weiteren Verlauf sollen die Mittelwertsunterschiede zwischen den einzelnen Messzeitpunkten mittels paarweisen t-Tests statistisch abgesichert werden (Bortz und Schuster 2010, S. 117). Sofern keine Normalverteilung der Daten vorliegt (Bortz und Schuster 2010, S. 119), wird äquivalent der Wilcoxon-Test zur Auswertung herangezogen. Da nicht alle Schüler:innen immer an jeder Erhebung teilnehmen konnten, liegt nicht dieselbe Personenstruktur vor. Daher kann keine Varianzanalyse mit Messwiederholung vorgenommen werden. Um der Gefahr einer alpha-Fehler Kumulation entgegen zu wirken (Rasch et al. 2014a, S. 3), wird eine False Discovery Rate-Korrektur (auch Benjamini-Hochberg-Korrektur, kurz BH) vorgenommen. Diese Methode eignet sich bei multiplen Signifikanztests (Benjamini und Hochberg 1995,

S. 289). Es sollen im Vergleich zur Bonferroni-Methode bei der BH-Korrektur durch eine höhere Teststärke möglichst viele echte Effekte identifiziert und der Anteil der Fehler unter den signifikanten Ergebnissen reduziert werden (Benjamini und Hochberg 1995, S. 298). Bei der Auswertung der Daten wird zwischen den fünf Faktoren amotiviert, external, introjiziert, identifiziert und intrinsisch differenziert. Von besonderem Interesse ist die intrinsische Motivation, da sie für den Lernprozess als besonders relevant gilt (siehe Abschnitt 2.1 *Motivation durch Bedürfnisbefriedigung*). Die anderen Regulationsformen werden zur Auswahl der Vignetten zum aktuellen Zeitpunkt nicht berücksichtigt.

Tabelle 6 Legende zur Abweichung der Mittelwerte pro Faktor vom Gesamtmittelwert

Farbcodierung	Abweichung vom Mittelwert	
	intrinsisch, identifiziert, amotiviert, introjiziert	extrinsisch
	-0.5	0.5
	-0.4	0.4
	-0.3	0.3
	0	0
	0.3	-0.3
	0.4	-0.4
	0.5	-0.5

Nach Auswertung der erhobenen aktuellen Lernmotivation zu den unterschiedlichen Messzeitpunkten werden die Ergebnisse in einen Zusammenhang mit der geplanten Instruktionenanpassung (z.B. besonderes Augenmerk auf die Freiräume legen) gebracht. Dies bedeutet, dass eine motivationsförderliche Anpassung, mit einer höheren intrinsischen Motivation der Schüler:innen einhergehen soll als bei einer motivationshemmenden Maßnahme. Der Abgleich mit der Instruktion zur Motivierung wird aus ökonomischen Gründen vorgenommen und dient der Eingrenzung des Videomaterials. Da die Instruktion möglicherweise nicht so wie geplant während der Unterrichtseinheit umgesetzt werden kann, findet zusätzlich ein Abgleich mit der Dokumentation (siehe Beispiel aus Tabelle 7 und Anhang E) zu der videografierten Unterrichtseinheit statt. In dieser Dokumentation wird festgehalten, ob die Lehrperson z.B. das Fehlerklima wie geplant besonders motivationsförderlich gestaltet hat oder nicht.

Tabelle 7 Dokumentation der Videoanalyse

Messzeitpunkt	Videozeiten	Kategorie	Inhalt
1 – Klasse 7	01:25 – 01:48	Fehlerklima, wertschätzender fürsorglicher Umgang	Fehler werden nicht besprochen, Lösungen werden durch Lehrkraft vorgegeben, der Umgangston ist eher direktiv
	08:20 – 08:28	Fehlerklima	Schüler:in kann Frage nicht beantworten, Lehrperson fordert nacharbeiten, keine Unterstützung oder Anregung der Denkprozesse

Entsprechend der Auswertung der Übereinstimmung der tatsächlichen Motivation der Lernenden mit dem Handeln der Lehrperson erfolgt die Vorauswahl geeigneter Vignetten. Diese werden im nächsten Schritt auf statistisch geprüfte Mittelwertsunterschiede getestet. Die Voraussetzung für die Verwendung eines t-Tests ist die Normalverteilung der Daten (Bortz und Schuster 2010, S. 119). Diese wird mittels Shapiro-Wilk-Test durchgeführt, der besonders für kleine Stichproben geeignet ist (Witte 2019, S. 34). Sofern sich statistisch signifikante Unterschiede zwischen den Messzeitpunkten aufzeigen lassen, wird deren Effektstärke mittels der Effektgröße Cohens d bestimmt (Döring und Bortz 2016, S. 816). Dabei können die Werte entsprechend der Einteilung nach Cohen als geringer (ab $d = 0.2$), mittlerer (ab $d = 0.5$) oder starker Effekt (ab $d = 0.8$) interpretiert werden (Cohen 1977, S. 25–26).

Die ausgewählten Videosequenzen werden unter Berücksichtigung zuvor festgelegter Regeln transkribiert. Auf eine Volltranskription u.a. mit wörtlichen Zitaten, Dialekten und Pausen kann in Bezug auf die geplanten Analysezwecke verzichtet werden. Anhand der Transkripte können relevante Schlüsselszenen für die Auswahl der Videoabschnitte herausgestellt werden. Die motivationsrelevanten Aspekte aus dem IBAS-Modell dienen zur Kategorisierung der Situationen und Aussagen. Zudem dient dieses Vorgehen dem erneuten Abgleich des wirklichen motivationsförderlichen oder -hemmenden Handelns der Lehrperson mit der tatsächlichen aktuellen Lernmotivation der Schüler:innen. Somit lassen sich am Ende geeignete Videosequenzen für das Videotool herausstellen.

Diese Sequenzen werden mit Camtasia 8 zu einzelnen themenbezogenen Videovignetten geschnitten (entsprechend der Dokumentation Videoanalyse, siehe Anhang E). Bei der Erstellung dieser Vignetten soll sowohl die Klassenperspektive als auch die Lehrperson dargestellt werden, daher wird mit einer Bild-in-Bild-Perspektive gearbeitet (siehe Abbildung 18).



Abbildung 18 Standbild aus einer Videovignette

7.3 Ergebnisse

Entsprechend der deskriptiven Auswertung des Fragebogens zur Erfassung der aktuellen Motivation der Schüler:innen zeigt sich, dass Item 7 (zugehörig zur extrinsischen Motivation) eine negative Trennschärfe aufweist. Dieses Item wird aus den weiteren Analysen ausgeschlossen, da es die Skala nicht abbilden kann (Luhmann 2015, S. 270). In die Auswertung gehen damit 14 Items ein. Die Überprüfung der Reliabilität (Cronbachs Alpha) ergibt eine akzeptable bis gute Passung (siehe Tabelle 8). Die Werte liegen, abgesehen von der Skala amotiviert, leicht oberhalb der Angaben von Prenzel et al. (1996).

Tabelle 8 Reliabilität der Skalen (α = Cronbachs alpha, M = Mittelwert, SD = Standardabweichung, : die Items zur Skala amotiviert weisen eine negative Polung auf, die Angaben beziehen sich auf die Gesamtstichprobe)

	intrinsisch	identifiziert	extrinsisch	amotiviert	introjiert
α	0.79	0.75	0.73	0.71	0.84
(α bei Prenzel (1996))	0.77	0.72	0.70	0.77	0.68)
M	2.2	2.1	2.3	2.6	2.5
SD	0.98	0.92	0.99	0.98	0.97

Auswertung Klassenstufe 7

Die Auswertung der aktuellen Lernmotivation der 7.Klasse (siehe Tabelle 9) hinsichtlich der intrinsischen Motivation findet zur Auswahl der Vignetten statt. Anhand der Mittelwerte mit entsprechender Farbcodierung (siehe Tabelle 6) wird deutlich, dass die intrinsische Motivation der Schüler:innen zu den Messzeitpunkten (MZP) 2, 7 und 15 ($M = 2.5$) höher ausfällt als im Durchschnitt über alle Messzeitpunkte ($M = 2.2$). Besonders hoch ist die intrinsische Motivation im Vergleich zum Durchschnitt beim MZP 8 ($M = 2.9$) und 9 ($M = 3.0$). Die intrinsische Motivation fällt am geringsten beim MZP 17 aus ($M = 1.5$).

Tabelle 9 Auswertung der aktuellen Lernmotivation der Klassenstufe 7

Mittelwerte Aktuelle Motivation (Klasse 7)					
Messzeitpunkt	intrinsisch	identifiziert	extrinsisch	amotiviert	introjiert
1	2	2.6	2.2	0.7	2.7
2	2.5	2.4	2.3	1	2.9
3	2.1	2.5	1.9	0.9	2.4
4	2.1	2.5	2.3	1	2.6
5	2.2	2.4	2.4	0.8	2.8
6	2.1	2.2	2.6	1	2.9
7	2.5	2.6	2.3	1.2	3
8	2.9	2.5	2.1	1.2	2.4
9	3	2.9	2.4	0.8	2.7
10	2.1	2.1	1.9	1.3	2.2
11	2	2.3	2.2	1.4	2.3
12	2	2.2	2.4	1.3	2.4
13	2.4	1.8	2	1.4	2.6
14	1.9	2.1	2.1	1.5	2.3
15	2.5	2.3	2	1.9	2.2
16	2	2.3	2.8	1.1	2.7
17	1.5	1.5	1.8	1.5	2.2
gesamt	2.2	2.3	2.2	1.2	2.6

Entsprechend der Mittelwertvergleiche innerhalb der intrinsischen Motivation können sieben Messzeitpunkte ausgewählt werden, die positiv im Vergleich zur durchschnittlichen Motivation auffallen. Diese sieben Messzeitpunkte werden im nächsten Schritt hinsichtlich der Übereinstimmung mit den motivationsförderlichen und -hemmenden Instruktionen (siehe Tabelle 10) überprüft und ausgewertet.

Tabelle 10 Instruktionen der Lehrperson, Schwerpunkte motivationsfördernder und -hemmender Aspekte für die Unterrichtseinheit (↑ = motivationsfördernd einsetzen, ↓ = motivationshemmend einsetzen)

Messzeitpunkt	Anpassungen
2	kognitive Aktivierung ↑, adäquate Freiräume ↓
7	keine Anpassung
8	adäquate Strukturiertheit ↑, individuelle Bezugsnorm ↑, positive Schüler-Schüler-Interaktion ↑
9	adäquate Freiräume ↑, Relevanz ↑, Verständlichkeit ↑, wertschätzender fürsorglicher Umgang ↑
14	anstrengungsbezogenes Feedback ↓, Leistungsdifferenzierung ↓
15	adäquate Freiräume ↑, Leistungsdifferenzierung ↑
17	demotivierend

Dies bedeutet für die einzelnen Messzeitpunkte im Hinblick auf die Auswahl der Vignetten Folgendes:

- MZP 1, 3, 4, 5, 10, 11, 12, 13 und 16 werden nicht für die Auswahl motivationsförderlicher bzw. -hemmender Situationen herangezogen, da sie hinsichtlich der intrinsischen Motivation keine relevante Ausprägung aufzeigen.
- MZP 7 wird nicht für die Vignettenauswahl herangezogen, da in dieser Unterrichtseinheit keine explizite Anpassung bzw. Fokussierung eines motivationsrelevanten Aspekts vorgenommen wurde. Eine Analyse der Unterrichtsstunde kann Aufschluss darüber geben, welche motivationsförderlichen oder -hemmenden Aspekte die Lehrperson eingesetzt hat. Da aber die eindeutige Fokussierung auf einzelne Aspekte in dieser Unterrichtseinheit gefehlt hat, werden diese Einheiten aktuell nicht verwendet. Für die Erstellung weiterer Vignetten ist es sinnvoll, die Analyse des MZP 7 vorzunehmen, da diese Unterrichtseinheit hinsichtlich der aktuellen Motivation geeignet scheint.
- Während des MZP 2 hat die Lehrperson den Aspekt adäquate Freiräume nicht wie theoretisch geplant und besprochen motivationshemmend umsetzen können. Die positive intrinsische Motivation ist passend zur motivationsförderlichen Umsetzung der kognitiven Aktivierung erhöht. Dementsprechend wird der MZP 2 für die Vignettenauswahl berücksichtigt.
- MZP 8, 9, 14 und 15 zeigen eine Übereinstimmung zwischen der Ausprägung der tatsächlichen Motivation und der Fokussierung motivationsförderlicher Aspekte.
- MZP 17 entspricht hinsichtlich der Ausprägung der tatsächlichen Motivation und der allgemein demotivierenden Fokussierung der Erwartung. Für die Auswahl geeigneter Vignetten ist dieser Messzeitpunkt anhand der Transkription und Kategorisierung der Unterrichtseinheit geeignet.
- Zusammenfassend werden folgende MZP vorausgewählt: 2, 8, 9, 14, 15 und 17.

Anhand dieser Auswahl an geeigneten videografierten Unterrichtseinheiten erfolgt die Überprüfung der Mittelwertsunterschiede mittels statistischem Testverfahren. Die Überprüfung der Daten auf Normalverteilung wird mittels Shapiro-Wilk-Test durchgeführt. Über alle relevanten Messzeitpunkte der 7.Klasse hinweg kann hinsichtlich der intrinsischen Motivation von einer Normalverteilung ausgegangen werden (siehe Tabelle 11). Daher erfolgt der Mittelwertvergleich mittels des t-Tests. Zusätzlich wird der Messzeitpunkt 3 ($M = 2.1$) herangezogen, um auszuschließen, dass bereits bei diesem Wert signifikante Unterschiede in der intrinsischen Motivation auftreten. Gegebenenfalls müssen die Grenzwerte angepasst werden, um besonders motivationsförderliche oder -hemmende Unterrichtseinheiten herausstellen zu können.

Tabelle 11 Testung auf Normalverteilung Klassenstufe 7 (MZP = Messzeitpunkt, W = Testgröße des Shapiro Wilk-Tests, p = Signifikanz)

MZP	W	p	MZP	W	p	MZP	W	p
2	0.877	0.120	8	0.904	0.181	14	0.957	0.738
3	0.886	0.153	9	0.883	0.095	15	0.901	0.203
						17	0.908	0.268

Es erfolgt die statistische Überprüfung der Mittelwertunterschiede zwischen verschiedenen Messzeitpunkten hinsichtlich der intrinsischen Motivation (siehe Tabelle 12). Die zentralen Ergebnisse zur intrinsischen Motivation sind an dieser Stelle zusammengefasst.

- 1) Es lassen sich signifikante Unterschiede zwischen folgenden Messzeitpunkten nachweisen:
 - MZP 2 unterscheidet sich signifikant von MZP 17 ($t = 2.57$, $p_{\text{korr}} = 0.03$, $d = 1.15$)
 - MZP 8 unterscheidet sich signifikant von drei Messzeitpunkten: MZP 3 ($t = 2.57$, $p_{\text{korr}} = 0.03$, $d = 1.15$), MZP 14 ($t = 2.75$, $p_{\text{korr}} = 0.03$, $d = 1.13$) und MZP 17 ($t = 3.86$, $p_{\text{korr}} = 0.01$, $d = 1.72$)
 - MZP 9 unterscheidet sich signifikant von drei Messzeitpunkten: MZP 3 ($t = 2.75$, $p_{\text{korr}} = 0.03$, $d = 1.19$), MZP 14 ($t = 2.92$, $p_{\text{korr}} = 0.01$, $d = 1.20$) und MZP 17 ($t = 3.98$, $p_{\text{korr}} = 0.01$, $d = 1.75$)
 - Es zeigen sich bei allen signifikanten Mittelwertsunterschieden eine hohe Effektstärke ($d > 0.8$).
- 2) Zwischen folgenden Messzeitpunkten ist der Unterschied in der intrinsischen Motivation der Lernenden statistisch nicht signifikant ($p_{\text{korr}} > 0.05$):
 - MZP 2 und MZP 14, MZP 14 und MZP 15, MZP 2 und MZP 8, MZP 2 und MZP 9, MZP 15 und MZP 17

Tabelle 12 Mittelwertvergleich zwischen der intrinsischen Motivation zu verschiedenen Messzeitpunkten – Klassenstufe 7 (MZP = Messzeitpunkt, M = Mittelwert, N = Anzahl der Schüler:innen, SD = Standardabweichung, t = Testgröße vom t -Test, p = Signifikanz, p_{korrr} = Signifikanz mit FalseDiscoveryRate Korrektur, d = Effektstärke)

MZP _A	M _A	N _A	SD _A	MZP _B	M _B	N _B	SD _B	t	p	p _{korrr}	d
2	2.50	10	0.76	14	1.86	12	1.15	1.561	0.135	0.162	
2	2.50	10	0.76	17	1.47	10	1.02	2.571	0.020	0.034	1.145
8	2.92	12	0.67	2	2.50	10	0.76	1.354	0.192	0.192	
8	2.92	12	0.67	3	2.10	10	0.80	2.563	0.020	0.034	1.121
8	2.92	12	0.67	14	1.86	12	1.15	2.749	0.013	0.031	1.126
8	2.92	12	0.67	17	1.47	10	1.02	3.855	0.002	0.012	1.715
9	3.03	12	0.77	2	2.50	10	0.76	1.614	0.123	0.162	
9	3.03	12	0.77	3	2.10	10	0.80	2.749	0.013	0.031	1.187
9	3.03	12	0.77	14	1.86	12	1.15	2.919	0.009	0.031	1.196
9	3.03	12	0.77	17	1.47	10	1.02	3.981	0.001	0.012	1.750
15	2.53	12	1.22	14	1.86	10	1.15	1.379	0.182	0.192	
15	2.53	12	1.22	17	1.47	10	1.02	2.223	0.038	0.057	

Auswertung Klassenstufe 8

Die Auswertung der aktuellen Lernmotivation für die 8.Klasse hinsichtlich der intrinsischen Motivation verläuft in der gleichen Art und Weise wie bei der Klassenstufe 7. Ausgangspunkt bildet die farbcodierte Tabelle 13.

Die intrinsische Motivation der Schüler:innen fällt im MZP 7, 12, 13, 14, 15 und 16 höher aus als die durchschnittliche intrinsische Motivation aller Messzeitpunkte ($M = 2.1$). Dabei ist die intrinsische Motivation beim MZP 13 am höchsten ($M = 2.7$). Die intrinsische Motivation fällt beim MZP 1, 5, 6 und 11 am geringsten aus ($1.5 < M < 1.8$).

Tabelle 13 Auswertung der aktuellen Lernmotivation der Klassenstufe 8

Mittelwerte Aktuelle Motivation (Klasse 8)					
Messzeitpunkt	intrinsisch	identifiziert	extrinsisch	amotiviert	introjiert
1	1.6	1.6	2.7	1.6	2.2
2	2	2.1	2.2	1.7	2.6
3	2.3	1.8	2.6	1.6	2.5
4	2.3	2.3	2.4	1.6	2.3
5	1.8	2.1	2.9	1.8	2.5
6	1.5	1.8	2.4	1.8	2.3
7	2.4	1.9	2.2	1.7	2.2
8	2	1.9	2.1	1.8	2.2
9	2.1	2	2	1.4	2.6
10	2.2	2.2	2.3	1.4	2.3
11	1.5	1.5	2.4	1.9	2.1
12	2.4	2	2.1	1.5	2.4
13	2.7	2.1	2.1	1.4	2.4
14	2.5	2.2	2.2	1.8	2.1
15	2.4	2.2	2.9	1.4	2.6
16	2.4	2.3	2.4	1.3	2.6
gesamt	2.1	2	2.4	1.6	2.4

Entsprechend der Mittelwertvergleiche innerhalb der intrinsischen Motivation können zehn Messzeitpunkte in der Klassenstufe 8 ausgewählt werden, die im nächsten Schritt hinsichtlich der Übereinstimmung mit den motivationsförderlichen und -hemmenden Instruktionen (siehe Tabelle 14) überprüft werden.

Tabelle 14 Instruktionen der Lehrperson, Schwerpunkte motivationsfördernder und -hemmender Aspekte für die Unterrichtseinheit (↑ = motivationsfördernd einsetzen, ↓ = motivationshemmend einsetzen)

Messzeitpunkt	Anpassungen
1	Interessantheit ↑, adäquate Freiräume ↓, Fehlerklima ↓
5	adäquate Strukturiertheit ↑, wertschätzender fürsorglicher Umgang ↑
6	Leistungsdifferenzierung ↓, Fehlerklima ↑, Kognitive Aktivierung ↑, adäquate Freiräume ↓
7	wertschätzender fürsorglicher Umgang ↓, positive Schüler:innen-Interaktion ↓, Interessantheit ↑, anstrengungsbezogenes Feedback ↑

Messzeitpunkt	Anpassungen
11	Kognitive Aktivierung ↑, adäquate Strukturiertheit ↓, Relevanz ↓
12	Relevanz ↑, Fehlerklima ↑
13	keine Anpassung
14	adäquate Strukturiertheit ↑, Interessantheit ↑, positive Schüler:innen-Interaktion ↑
15	demotivierend
16	demotivierend

Dies bedeutet für die einzelnen Messzeitpunkte im Hinblick auf die Auswahl der Vignetten Folgendes:

- MZP 2, 3, 4, 8, 9 und 10 werden nicht für die Auswahl motivationsförderlicher bzw. -hemmender Situationen herangezogen, da sie hinsichtlich der intrinsischen Motivation keine relevante Ausprägung aufzeigen.
- Die Messzeitpunkte 1, 5, 6 und 11 zeigen eine geringe intrinsische Motivation auf.
- Die geplante Instruktion wenig Freiräume und einen motivationshemmenden Umgang mit Fehlern zum MZP 1 umzusetzen, ist in Einklang mit der geringen intrinsischen Motivation zu bringen. Die Interessantheit des Themas scheint eine geringere Auswirkung zu haben. Daher wird der MZP 1 für die Auswahl geeigneter Vignetten berücksichtigt.
- Im MZP 6 lässt sich anhand der Dokumentation nachvollziehen, dass das Handeln der Lehrperson in dieser Unterrichtseinheit vor allem direkter Art ist. Der Umgang mit Fehlern wirkt wenig motivationsfördernd. Adäquate Freiräume werden, nicht geschaffen. Ebenso erfolgt keine Leistungsdifferenzierung. Eine Transkription der Vignette scheint sinnvoll, um genauen Aufschluss über das tatsächliche Handeln der Lehrperson zu geben.
- Die theoretische Planung für den MZP 11 passt ebenfalls zur geringen intrinsischen Motivation. Ausgenommen ist die Förderung der kognitiven Aktivierung, die „zu einem vertieften Nachdenken über den Unterrichtsgegenstand“ (Hanisch 2015, S. 187) führen soll. Hier kann die Vermutung nahegelegt werden, dass sich eine Auseinandersetzung mit kognitiv aktivierendem Material negativ auf die intrinsische Motivation auswirken kann, sofern die Schüler:innen nicht daran gewohnt sind.
- Im MZP 5 lässt sich die geringe intrinsische Motivation nicht mit der theoretischen Planung in Zusammenhang bringen. Daher wird MZP 5 zunächst nicht verwendet.
- MZP 7 zeichnet sich durch eine erhöhte intrinsische Motivation aus. Anhand der Kurzdokumentation wird deutlich, dass die Lehrperson besonders im zweiten Teil der Stunde den Fokus auf die Vergabe eines anstrengungsbezogenen Feedbacks gelegt hat. Daher soll der MZP 7 transkribiert und kategorisiert werden.

- Im MZP 9 wird keine Anpassung vorgenommen. Ebenso gibt es keine Fokussierung eines motivationsförderlichen oder -hemmenden Aspekts im MZP 13. Eine ausführliche Analyse könnte gegebenenfalls zusätzliches Material liefern. Aus ökonomischen Gründen wird darauf aktuell verzichtet.
- Im MZP 12 und 14 fällt die intrinsische Motivation höher aus als im Durchschnitt über alle Messzeitpunkte, dies passt zum geplanten motivierenden Handeln der Lehrperson.
- MZP 15 und 16 sollen entsprechend der Instruktion demotivierend wirken. Bei beiden Messzeitpunkten ist die intrinsische Motivation höher als im Durchschnitt über alle MZP. Aus diesen Gründen sollen beide Messzeitpunkte weiter analysiert werden, um das tatsächliche Handeln der Lehrperson in einen Zusammenhang mit der aktuellen Lernmotivation bringen zu können. Anders als erwartet ist die tatsächliche intrinsische Motivation leicht über den Durchschnitt über alle Messzeitpunkte. Eine Transkription dieser Unterrichtseinheiten kann das motivationsförderliche Handeln der Lehrperson aufdecken. Besonders interessant ist im MZP 15 die hohe Ausprägung der extrinsischen Motivation. Für das aktuelle Ziel wird allerdings auf eine weitere Analyse verzichtet.
- Zusammenfassend werden folgende MZP vorausgewählt: 1, 6, 7, 11, 12 und 14.

Um die Mittelwertsunterschiede statistisch absichern zu können, erfolgt zunächst die Testung auf Normalverteilung der Daten. Anhand der Ergebnisse des Shapiro-Wilk-Tests sind alle Daten der ausgewählten Messzeitpunkte normalverteilt (siehe Tabelle 15).

Tabelle 15 Testung auf Normalverteilung Klassenstufe 8 (MZP = Messzeitpunkt, W = Testgröße des Shapiro Wilk-Tests, p = Signifikanz)

MZP	W	p	MZP	W	p	MZP	W	p
1	0.936	0.248	7	0.905	0.083	12	0.956	0.582
6	0.940	0.582	11	0.906	0.087	14	0.903	0.054

Die Überprüfung der Mittelwertsunterschiede wird entsprechend der Normalverteilung mittels t-Test durchgeführt. Während die MZP 1, 7 und 11 eine niedrige intrinsische Motivation aufweisen, sind die Werte zu den MZP 6, 12 und 14 hoch. Daher erfolgt eine Gegenüberstellung der MZP, um zu überprüfen, ob alle Messpaare sich statistisch hinsichtlich der Mittelwerte zur intrinsischen Motivation unterscheiden (siehe Tabelle 16). Die zentralen Ergebnisse zur intrinsischen Motivation in der Klassenstufe 8 sind an dieser Stelle zusammengefasst.

- 1) Es lassen sich signifikante Unterschiede zwischen folgenden Messzeitpunkten nachweisen:
- MZP 1 unterscheidet sich signifikant von MZP 7 ($t = -2.98$, $p_{\text{korr}} = 0.02$, $d = 1.02$), MZP 12 ($t = -2.55$, $p_{\text{korr}} = 0.03$, $d = 0.91$) und MZP 14 ($t = -3.02$, $p_{\text{korr}} = 0.02$, $d = 0.99$)
 - MZP 6 unterscheidet sich signifikant von MZP 7 ($t = -2.28$, $p_{\text{korr}} = 0.04$, $d = 0.94$) und MZP 14 ($t = -2.32$, $p_{\text{korr}} = 0.04$, $d = 0.92$)
 - MZP 11 unterscheidet sich signifikant von MZP 7 ($t = -2.82$, $p_{\text{korr}} = 0.02$, $d = 0.96$), MZP 12 ($t = -2.45$, $p_{\text{korr}} = 0.03$, $d = 0.86$) und MZP 14 ($t = -2.86$, $p_{\text{korr}} = 0.02$, $d = 0.85$)
 - Die Effektstärken der signifikanten Ergebnisse liegen alle im hohen Bereich ($d > 0.8$).
- 2) Zwischen folgenden Messzeitpunkten ist der Unterschied in der intrinsischen Motivation der Lernenden nicht statistisch signifikant:
- lediglich MZP 6 und 12 bleiben unterhalb des Signifikanzniveaus ($p_{\text{korr}} = 0.06$)

Tabelle 16 Mittelwertvergleich zwischen der intrinsischen Motivation zu verschiedenen Messzeitpunkten – Klassenstufe 8 (MZP = Messzeitpunkt, M = Mittelwert, SD = Standardabweichung, N = Anzahl der Schüler:innen, t = Testgröße vom t-Test, p = Signifikanz, p_{korr} = Signifikanz mit FalseDiscoveryRate Korrektur, d = Effektstärke)

MZP _A	M _A	N _A	SD _A	MZP _B	M _B	N _B	SD _B	t	p	p_{korr}	d
1	1.57	18	0.64	7	2.43	17	1.01	-2.984	0.006	0.018	1.024
1	1.57	18	0.64	12	2.38	16	1.10	-2.547	0.018	0.032	0.914
1	1.57	18	0.64	14	2.46	19	1.09	-3.015	0.005	0.018	0.989
6	1.48	9	1.02	7	2.43	17	1.01	-2.277	0.037	0.042	0.937
6	1.48	9	1.02	12	2.38	16	1.10	-2.048	0.055	0.055	
6	1.48	9	1.02	14	2.46	19	1.09	-2.316	0.033	0.042	0.917
11	1.55	17	0.81	7	2.43	17	1.01	-2.822	0.008	0.018	0.961
11	1.55	17	0.81	12	2.38	16	1.10	-2.445	0.021	0.032	0.863
11	1.55	17	0.81	14	2.35	19	1.05	-2.856	0.007	0.018	0.847

Videoanalyse gesamt

Entsprechend der Ergebnisse der Datenauswertung der 7.Klasse und der 8.Klasse können 12 Messzeitpunkte für die Videoanalyse herausgestellt werden. Die Transkription und Kategorisierung dieser Unterrichtseinheiten ergibt 21 Videoausschnitte, die für die Erstellung der Vignetten geeignet sind. Anhand der Abbildung 19 wird beispielhaft die Übersicht für das Schneiden der Vignetten in Camtasia 8 dargestellt.

<i>IBAS - Anpassung</i>	<i>Datum</i>	<i>Zeiten</i>	<i>Inhalt</i>
<i>Strukturiertheit ↓</i>	25.01.18 I	00:25 – 05:10	
<i>Strukturiertheit ↑</i>	19.01.18 I	00:10 – 00:22	Rückblick
		00:50 – 01:39	Rückblick
		02:15 – 03:07	Überblick und Überleitung zur aktuellen UE
		03:08 – 03:46	Fahrplan der Stunde, Überleitung zum Video
		22:59 – 23:03	Überleitung und Rückblick
		23:18 – 23:21	Fahrplan Verweis
		23:40 – 23:45	Ausblick nächste Stunde
<i>Strukturiertheit ↑</i>	01.02.18 II	00:34 – 01:52	Orientierung, Thema einbindung, Stundeninhalt/Vorgehensweise, Plakatregeln
<i>Freiräume ↑</i>	09.02.18 II	01:38 – 01:59	Wahlaufgaben
		02:38 – 03:49	Wahlaufgaben
		03:50 – 03:52	Wahl der Methode Einzel- oder Partnerarbeit
		04:23 – 04:48	Zusatzaufgaben
		04:52 – 05:09	Wahlfreiheit am S-Beispiel erklärt
		08:35 – 08:50	Wahlfreiheit
		10:28 – 11:30	Zusatzaufgaben -> Tafel notiert

Abbildung 19 Videoschnitt-Plan

7.4 Interpretation und Schlussfolgerungen

Auf Basis der erhobenen Daten zur aktuellen Motivation der Schüler:innen können bei einer Übereinstimmung mit der theoretisch fundierten Anpassung der Unterrichte 12 gefilmte Unterrichtsabschnitte in die Vorauswahl für die Erstellung der Videovignetten genommen werden. Die deskriptiven Analysen mit den Mittelwertvergleichen zwischen den verschiedenen Messzeitpunkten zeigen mit BH-Korrektur signifikante Unterschiede zwischen einzelnen Unterrichtseinheiten (z.B. intrinsische Motivation der Klassenstufe 8 zum MZP1: $M = 1.57$, $SD = 0.64$; MZP7: $M = 2.43$, $SD = 1.01$; $t = -2.98$, $p_{\text{korrr}} = 0.02$, $d = 1.02$). Aufgrund der Fokussierung einzelner motivationaler Aspekte während der Unterrichtseinheiten (z.B. MZP1: Reduktion der Adäquaten Freiräume und des Positiv-konstruktiven Fehlerklimas; MZP7: Geben eines anstrengungsbezogenen Feedbacks) können Rückschlüsse auf die unterschiedliche Beeinflussung dieser Aspekte auf die Lernmotivation gezogen werden.

Die Ergebnisse erweitern den bisherigen methodischen Ansatz der inhaltlichen Validierung von Videovignetten, die in Videostudien über Expert:innenratings erfolgt sind (Oser et al. 2010; u.a. Meschede et al. 2015; Seidel et al. 2010; Star und Strickland 2008). Daraus ergeben sich wichtige Implikationen für folgende Forschungen, die sich ebenfalls mit Konstrukten wie der Motivation beschäftigen, die insbesondere einen inneren Prozess beschreiben und damit optimal durch eine Selbsteinschätzung beurteilt werden sollten.

Insgesamt können 21 Unterrichtssequenzen für die Erstellung von geeigneten motivationsförderlichen und -hemmenden Situationen herausgestellt werden. Aufgrund des weiteren Vorgehens der *Entwicklung und Validierung passender Analyseaufträge* zu den Vignetten (siehe Abschnitt 8.1 und 8.2) wird eine weitere Eingrenzung des Materials für die vorliegende Arbeit vorgenommen. Dies ergibt sich aus ökonomischen Aspekten, da die weiteren Schritte sehr zeitaufwendig sind. Insgesamt werden in einem ersten Schritt sieben Vignetten (*Beschreibung der Vignetten* siehe Abschnitt 7.5) erstellt. Die

Schwerpunkte hinsichtlich der motivationsrelevanten Aspekte im Handeln der Lehrperson liegen auf der Leistungsdifferenzierung, der individuellen Bezugsnormorientierung, der Relevanz, den adäquaten Freiräumen, der adäquaten Strukturierung und dem positiv-konstruktiven Fehlerklima. Nichtsdestotrotz können die hier vorgenommenen Auswertungen und Analysen für die Erstellung weiterer Vignetten genutzt werden. Dies ist vor allem im Hinblick auf die Entwicklung eines umfassenden Videotools zur Förderung und Analyse der Professionellen Unterrichtswahrnehmung hinsichtlich motivationsrelevanter Aspekte sinnvoll. Mit den zusätzlichen Vignetten können perspektivisch auch die weiteren Facetten der Motivierung in Anlehnung an das IBAS-Modell (siehe im Abschnitt 2.2) ergänzt werden.

Resultat zur Forschungsfrage 1:

Die Vignetten stellen repräsentative motivationsförderliche oder -hemmende Situationen dar. Anhand der Auswertung der tatsächlichen Motivation der Schüler:innen im Abgleich mit der theoretisch fundierten motivationsförderlichen bzw. -hemmenden Anpassung können entsprechende Videovignetten ausgewählt und inhaltlich validiert werden.

Die Entscheidung für die Verwendung von multiplen t-Tests mit alpha-Fehler-Korrektur zum Aufdecken von Unterschieden in der Motivation der Lernenden zwischen den Unterrichtseinheiten statt eines Linearen Mixed Models kann kritisch betrachtet werden. Ziel war es nicht, individuelle Entwicklungsverläufe darzustellen, sondern in erster Linie signifikante Unterschiede zwischen den einzelnen Messzeitpunkten für die Auswahl geeigneter Videosequenzen zu beschreiben. Daher eignen sich die paarweisen t-Tests, vor allem auch im Hinblick auf die nicht konsistente Schüler:innenschaft. Zur Reduktion der alpha-Fehler Kumulation wurde eine BH-Korrektur vorgenommen. Auch hier bleibt anzumerken, dass diese Methode von einer weniger konservativen Annahme ausgeht als die Bonferroni-Methode. Die BH-Korrektur bietet jedoch den Vorteil, mehr echte Effekte bei ausreichender Teststärke aufzudecken. Das Risiko der Annahme falsch-positiver Signifikanzen bei den mehrfach durchgeführten paarweisen t-Tests bleibt gegenüber der Bonferroni-Methode jedoch höher.

7.5 Beschreibung der Vignetten

Die Vignetten zeigen sieben reale Unterrichtssituationen aus dem Erdkundeunterricht. Diese können mit den entsprechenden Analyseaufträgen (siehe Kapitel 8) hinsichtlich der Professionellen Unterrichtswahrnehmung zur Motivationsförderung analysiert werden. An dieser Stelle werden die einzelnen Vignetten kurz vorgestellt. Dazu werden zum einen das Thema der Stunde und der Schwerpunkt der Motivationsförderung genannt, zum anderen folgt eine zusammenfassende Beschreibung der Sequenz sowie eine

Auflistung der Schwerpunkte der Items des Erklärens und Vorhersagens. Für eine bessere Nachvollziehbarkeit (vor allem für die Raschskalierung siehe Abschnitt 10.2.2) werden die Items mit ihrem Kürzel an entsprechender Stelle zugeordnet. Beim Itemcode wird eine grundlegende Struktur eingehalten, die an folgenden Beispielen erklärt wird:

V1BS1 (V1 = Vignette 1, BS = Beschreiben, 1 = Item 1),

V4E12 (V4 = Vignette 4, E = Erklären, 1 = 1.Fokus, 2 = Item 2),

V6V43 (V6 = Vignette 6, V = Vorhersagen, 4 = 4.Fokus, 3 = Item 3).

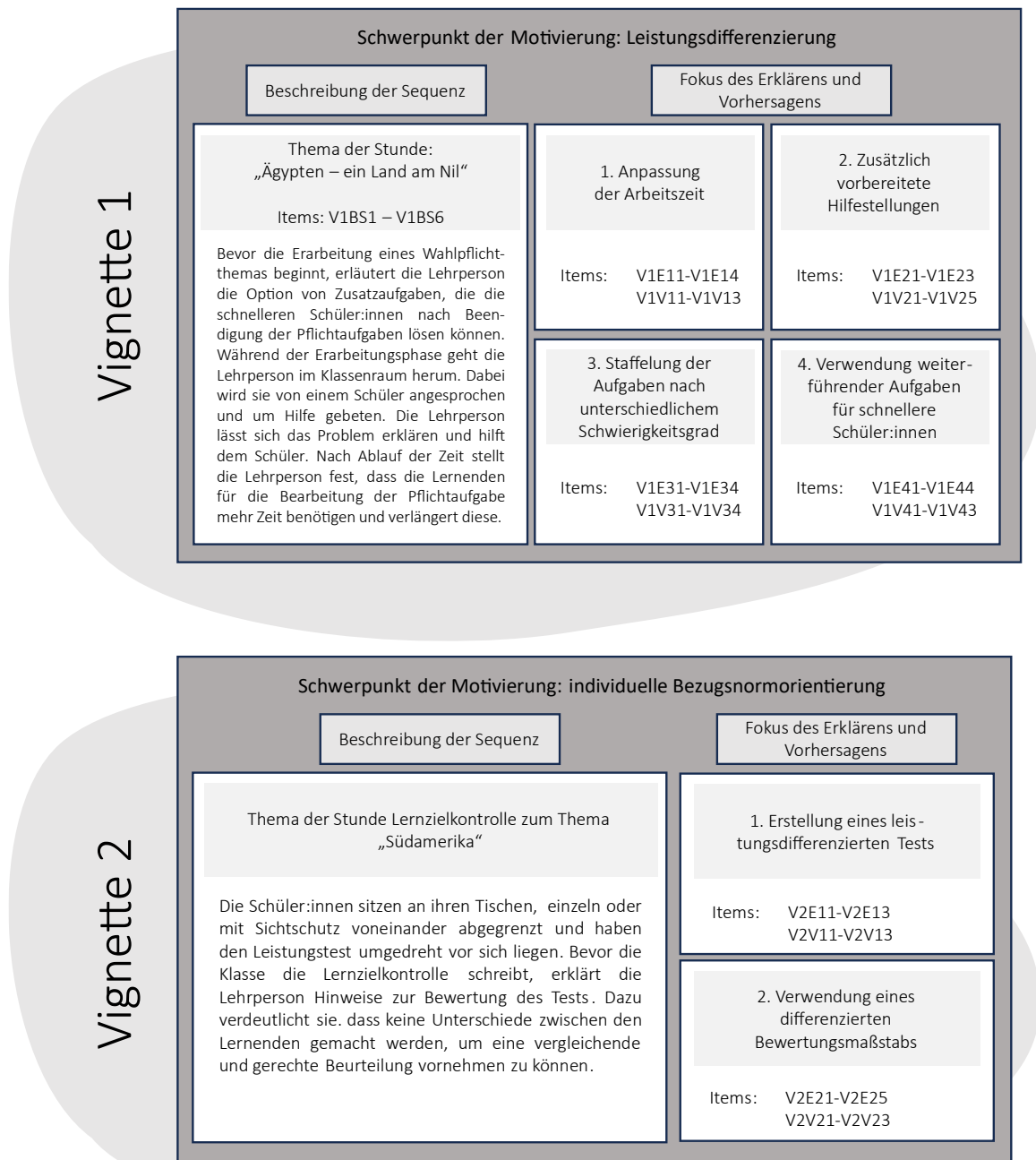


Abbildung 20 Beschreibung der Vignetten 1 und 2

Die Vignette 3 stellt eine Besonderheit dar. Als einzige Vignette zeigt sie einen demotivierenden Aspekt. Die Lehrperson eröffnet die Unterrichtssequenz mit zwei Leitfragen, die sie an die Tafel anbringt und vorliest. Anschließend folgt die Erarbeitung des Themas in Einzel- oder Partnerarbeit. Die dazugehörigen Arbeitsaufträge werden auf einer Folie präsentiert. Auf die Relevanz des Themas geht die Lehrperson nicht ein. Daher handelt es sich um einen motivationshemmenden Aspekt hinsichtlich der Subdimension der Relevanz (siehe IBAS-Modell im Abschnitt 2.2.1). Daher werden zu dieser Vignette keine Analysefragen zu den Aspekten des Erklärens und Vorhersagens formuliert. Die Vignette 4 greift den Schwerpunkt der Relevanz aus motivationsförderlicher Sicht auf und setzt an der Vignette 3 mit dem Vorstellen der Leitfragen an.

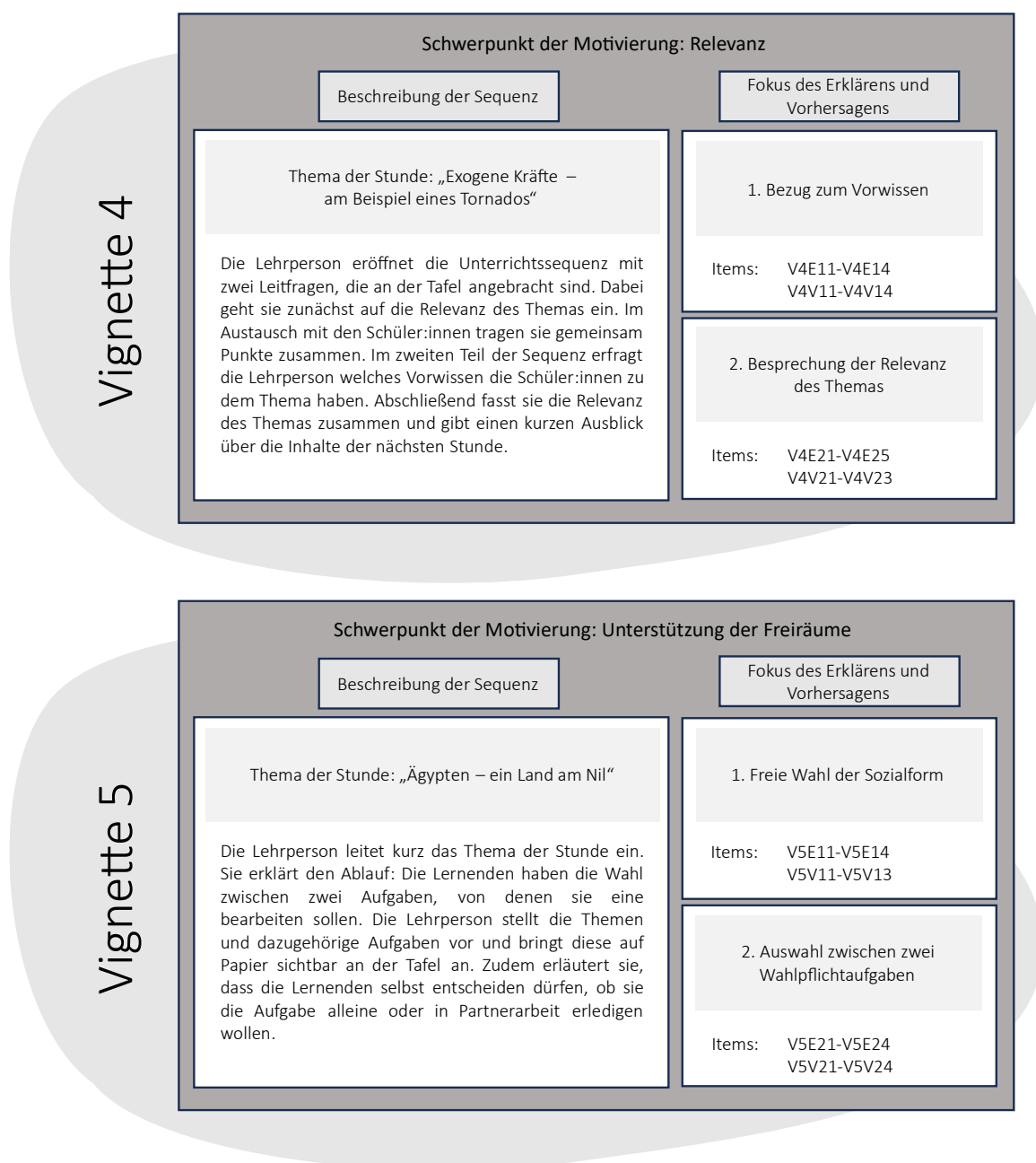


Abbildung 21 Beschreibung der Vignetten 4 und 5

Vignette 6

Schwerpunkt der Motivierung: adäquate Strukturiertheit		
Beschreibung der Sequenz	Fokus des Erklärens und Vorhersagens	
<p>Thema der Stunde: „Exogene Kräfte“</p> <p>Das Thema der Unterrichtseinheit „Exogene Kräfte“ soll in arbeitsteiligen Gruppen erarbeitet werden. Dazu bespricht die Lehrperson die Rahmenbedingungen. Sie geht insbesondere auf die Plakatregeln ein, die die Lernenden bei der Anfertigung des Plakats als Ergebnissicherung berücksichtigen sollen. Die Plakatregeln sind an der Tafel verschriftlicht und dienen als Bewertungskriterien. Ausblick: die Lehrperson bespricht die einzelnen Punkte mit den Lernenden.</p>	<p>1. Benennung des Themas der Stunde</p> <p>Items: V6E11-V6E13 V6V11-V6V13</p>	<p>2. Beschreibung der Lernziele</p> <p>Items: V6E21-V6E24 V6V21-V6V24</p>
	<p>3. Beschreibung des methodischen Vorgehens der Gruppenarbeit</p> <p>Items: V6E31-V6E35 V6V31-V6V35</p>	<p>4. Beschreibung der Leistungserwartung der Gruppenarbeit</p> <p>Items: V6E41-V6E44 V6V41-V6V44</p>

Vignette 7

Schwerpunkt der Motivierung: Unterstützung der Freiräume	
Beschreibung der Sequenz	Fokus des Erklärens und Vorhersagens
<p>Thema der Stunde: „Exogene Kräfte – am Beispiel eines Tornados“</p> <p>Die Lehrperson möchte die Folgen eines Tornados mit der Klasse besprechen. Sie ermutigt die Lernenden sich zu äußern und ihre Ansichten zu erklären, damit man sie nachvollziehen oder korrigieren kann. Anschließend teilen einzelne Lernende ihre Ideen mit. Im zweiten Teil der Sequenz stellt die Lehrperson eine weitere Frage und ermutigt die Lernenden zum Antworten. Dabei betont sie besonders, dass Fehler erlaubt sind und zum Lernprozess dazu gehören. Am Ende lobt sie eine Schülerin für ihre Rückmeldung.</p>	<p>1. Fehler als Teil des Lernprozesses</p> <p>Items: V7E11-V7E15 V7V11-V7V14</p>
	<p>2. Eingehen auf Fehler</p> <p>Items: V7E21-V7E25 V7V21-V7V24</p>

Abbildung 22 Beschreibung der Vignetten 6 und 7

8 Entwicklung und Validierung der Analyseaufträge

Um aus den Videovignetten ein Analysetool generieren zu können, bedarf es in den nächsten Schritten der *Entwicklung und Validierung der Analyseaufträge* (siehe Abschnitt 8.1 und 8.2). Die Zusammenführung beider Elemente in das bestehende Videotool ViviAn ergeben das vorliegende videobasierte Instrument (siehe Abschnitt 8.3 *Implementierung in ViviAn*).

8.1 Entwicklung der Analyseaufträge

Das erstellte und validierte Videomaterial wird in einem nächsten Schritt mit Analyseaufträgen auf Basis der aktuellen Literatur (Helmke 2015) ergänzt. „Zur Nutzung jeder Vignette in der Lehrerbildung gehören **Aufgaben**, die in spezifischer Weise zur Bearbeitung bzw. Analyse der Vignette auffordern“ (von Aufschnaiter et al. 2017, S. 86). Thematischer Schwerpunkt der Vignetten ist der Einsatz motivationsfördernder Aspekte im Unterricht. Daher orientieren sich die Analyseaufträge an den bereits erwähnten Motivationstheorien und -modellen und an der aktuellen Literatur und Forschung zur Motivation und Motivierung im Unterricht (siehe Kapitel 2). Zur Analyse und Förderung der Professionellen Unterrichtswahrnehmung sollen die Analyseaufträge entsprechend der Prozesse des Noticing und Knowledge-based Reasoning (Seidel et al. 2010; Sherin und van Es 2008; van Es 2009) ausgerichtet werden. Im ersten Schritt der Entwicklung des videobasierten Instrumentes wird der Prozess des Noticings nicht explizit berücksichtigt, sondern könnte perspektivisch durch das Setzen von Markern impliziert werden. Somit liegt der derzeitige Fokus auf dem Bereich des Knowledge-based Reasoning mit den drei Stufen des Beschreibens, Erklärens und Vorhersagens (Seidel et al. 2010; Sherin und van Es 2008; van Es 2009). Die Analyseaufträge werden entsprechend dieser Phasen und auf Grundlage des theoretischen Hintergrundes formuliert.

1. Beschreiben: Hier sollen die Komponenten des Unterrichts (durch Ankreuzen) beschrieben werden.
2. Erklären: Einzelne Unterrichtshandlungen aus dem Video sollen auf Basis wissenschaftlicher Theorien und Befunde erklärt werden.
3. Vorhersagen: Die Wirkungen dieser Unterrichtshandlungen auf den weiteren Lehr-Lernprozess sollen vorhergesagt werden.

Alle Fragen (Items) werden über eine vierstufigen Likert-Skala beantwortet. Es besteht die Möglichkeit zwischen 1 – stimmt genau, 2 – stimmt eher, 3 – stimmt eher nicht und 4 – stimmt gar nicht zu wählen.

Zur Veranschaulichung, wie die Analyseaufträge für den jeweiligen Teilprozess inklusive des Itemstamms umgesetzt und innerhalb des Videotools implementiert werden, dienen die folgenden Beispielfragen (siehe Abbildung 23-25).

Ein Beispiel: Die Vignette 7 handelt vom Thema „Exogene Kräfte – am Beispiel eines Tornados“. In dieser Einheit bespricht die Lehrperson die Folgen eines Tornados mit der Klasse. Der Ablauf ist folgendermaßen vorgesehen (siehe Anhang F ViviAn – Anleitung für Studierende): Die Studierenden können sich in der Lernumgebung ViviAn zunächst zusätzliche Informationen zur Klasse, zur Stunde sowie die verwendeten Lehr-Lernmaterialien ansehen. Danach starten sie das Video. Im Anschluss gelangen sie über den Button „Analyseauftrag“ zu den Aufträgen, strukturiert nach den Teilprozessen: 1. Beschreiben, 2. Erklären und 3. Vorhersagen.

Entscheiden Sie, inwieweit die BESCHREIBUNGEN auf die gesehene Vignette zutreffen.				
	stimmt genau	stimmt eher	stimmt eher nicht	stimmt gar nicht
Die Lehrperson nimmt auf gemachte Aussagen Bezug.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fehler werden als Teil des Lernprozesses gesehen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Lehrperson korrigiert falsche Aussagen der Schülerinnen und Schüler sofort.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 23 Analyseaufträge zum Beschreiben: Beispielhaft 3 Items (von 7) für das Beschreiben der Vignette 6

Beurteilen Sie, welche der folgenden ERKLÄRUNGEN bezüglich der Motivationsförderung theoretisch angemessen sind.

Die Lehrperson **verdeutlicht, dass Fehler Teil des Lernprozesses sind,**

	stimmt genau	stimmt eher	stimmt eher nicht	stimmt gar nicht
um Fehler nicht in Verbindung mit negativen Konsequenzen zu bringen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
um die Risikobereitschaft der Schülerinnen und Schüler zu erhöhen, sich auch an schwierige Aufgaben zu trauen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
um den Schülerinnen und Schülern Aufgaben stellen zu können, bei denen es mehrere Lösungen gibt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 24 Analyseaufträge zum Erklären: Items für das Handeln der Lehrperson „Verdeutlichen, dass Fehler Teil des Lernprozesses sind“. Es gibt für diese Vignette weitere Handlungen, die aus theoretischer Perspektive erklärt werden sollen.

Beurteilen Sie, welche AUSWIRKUNGEN **Fehler als Teil des Lernprozesses zu sehen** für den weiteren Lehr-Lernprozess hat.

	Stimmt genau	stimmt eher	stimmt eher nicht	stimmt gar nicht
Deshalb werden Schülerinnen und Schüler, die sich nicht verbessern, gelobt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Deshalb üben die Schülerinnen und Schüler, Fehler aufzudecken und zu korrigieren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Deshalb trauen sich die Schülerinnen und Schüler, ihre Vermutungen zu äußern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 25 Analyseaufträge zum Vorhersagen: 3 Items, passend zu der vorherigen Erklärung.

Anhand der Beispiele wird deutlich, dass die Analyseaufträge entsprechend der Ebenen des Knowledge-based Reasonings einer sprachlichen Struktur folgen. Alle Itemstämme des Erklärens und Vorhersagens beziehen sich auf ein in der Vignette gezeigtes Lehrpersonenhandeln. Eine Begründung für die Benennung des Handelns gebührt der lokalen stochastischen Unabhängigkeit. „Die Wahrscheinlichkeit, ein Item zu lösen, beeinflusst nicht die Wahrscheinlichkeit, ein anderes Item zu lösen“ (Rentzsch et al. 2009, S. 155). Wenn eine Situation nicht richtig beschrieben werden kann, könnte ohne Vorgabe des Lehrpersonenhandelns nicht überprüft werden, ob richtige Erklärungen für dieses Handeln oder Vorhersagen für den weiteren Lehr-Lernprozess abgerufen werden können. Möglich wäre, dass auf Grund der komplexen Situation das Handeln nicht wahrgenommen wurde.

Die Items des Erklärens und Vorhersagens unterscheiden sich in der Verwendung der einleitenden Wörter: „um“ bzw. „deshalb“. Alle Items einer Kategorie beginnen immer gleich, um eine förderliche Struktur im Sinne einer Kontinuität herzustellen. Damit trägt es auch zur Augenscheinvalidität bei, da die Proband:innen gleich wissen, worauf das Item zielt (Bühner 2011, S. 62) (siehe Abschnitt 5.3).

Eine Besonderheit in der Entwicklung der Analyseaufträge stellen motivationshemmende Situationen dar. Im Vergleich zum Videotool OBSERVER werden nicht nur lernwirksame Unterrichtsausschnitte verwendet (Seidel et al. 2010, S. 299). Eine Lehrperson sollte ebenfalls in der Lage sein, motivationshemmende Aspekte wissensbasiert wahrzunehmen und beschreiben zu können. Allerdings ist es wenig zielführend, Erklärungen für Fehlkonzepte abzurufen. Daher wird der Itemstamm durch eine Umpolung der gezeigten Handlung in Form eines Konjunktivs angepasst. Im folgenden Beispiel soll dies verdeutlicht werden (Abbildung 28):

Beurteilen Sie, welche der folgenden ERKLÄRUNGEN
bezüglich der Motivationsförderung theoretisch angemessen sind.

Angenommen, die Lehrperson **würde die Lernziele der Stunde beschreiben**,
dann geschieht dies...

	stimmt genau	stimmt eher	stimmt eher nicht	stimmt gar nicht
um zur Strukturierung der Unterrichtsstunde beizutragen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
um die Schülerinnen und Schüler nicht zu überfordern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
um einen persönlichen Bezug herzustellen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 26 Erklärungen für ein nicht gezeigtes, aber motivationsförderliches Lehrpersonenhandeln. 3 Beispielim des Erklärens aus Vignette 6.

Eine erste inhaltliche Überprüfung der Items, inwiefern sie das zu messende Konstrukt abbilden (siehe Abschnitt 8.2 Validierung der Analyseaufträge), wird durch Konsensgespräche innerhalb einer kleinen Expert:innengruppe getroffen. Dabei wird jedes einzelne Item überprüft und ggf. angepasst.

8.2 Validierung der Analyseaufträge

Im Anschluss an die Entwicklung der Analyseaufträge müssen diese validiert werden. Dazu kommen Expert:innenratings zum Einsatz. Dieses Vorgehen zählt als Standardvariante zur Überprüfung der Validität von Testitems (Brunner et al. 2006, S. 531). Ziel ist es, ein Tool zu entwickeln, das hinsichtlich der Beurteilung der Analyseaufträge eine hohe Auswertungsobjektivität aufzeigen kann. Dieser Abschnitt widmet sich daher der folgenden Forschungsfrage:

Forschungsfrage

- 2 Ist die Übereinstimmung der Expert:innen bei der Auswertung der Rating-Items ausreichend hoch, um von einer hohen Auswertungsobjektivität ausgehen zu können?*

8.2.1 Stichprobe

Die Validierung der Analyseaufträge wird in einem zweiteiligen Prozess mit zwei unterschiedlichen Expert:innengruppen durchgeführt. Beim Beschreiben sind die Analyseaufträge direkt mit den Videos verknüpft. Da Noviz:innen das Beschreiben relevanter Situationen im Vergleich zum Erklären einfacher gelingt (Frommelt et al. 2016, S. 369; Schäfer und Seidel 2015), wird der Fokus der Expertise im Besonderen auf die Arbeit mit Videos gerichtet und weniger auf die Lehrtätigkeit in Schulen. Das Beurteilen von Videos ist eine herausfordernde Aufgabe. Deswegen sind alle im wissenschaftlichen Bereich beschäftigten Expert:innen mit der Analyse von Videoausschnitten durch ihre Tätigkeit sehr vertraut. Zudem haben alle Rater:innen einen universitären pädagogischen Abschluss und können somit auch eine angemessene inhaltliche Aussage treffen.

Die Komponenten des Erklärens und Vorhersagens beziehen sich insbesondere auf empirische Studien und Theorien aus dem Bereich der Motivationsforschung. Daher werden hierbei für das Rating Expert:innen ausgewählt, die mit diesen Themen vertraut sind. Insgesamt werden fünf Dozierende, die auf universitärer Ebene Vorlesungen oder Seminare zum Thema Motivation und Motivierung halten, alle 156 Items des Erklärens und Vorhersagens bearbeiten.

8.2.2 Methodisches Vorgehen

Zur Überprüfung der richtigen Antworten der einzelnen Analyseaufträge (225 Items) werden zwei typische Verfahren angewendet: Zum einen wird auf Ergebnisse aus der pädagogischen und psychologischen Forschung, speziell im Bereich der Motivation und Motivierung, zurückgegriffen, zum anderen werden Expertenratings durchgeführt (Brunner et al. 2006, S. 531; Carlson 1990, S. 161). In der Kategorie des Beschreibens soll beurteilt werden, ob und wie häufig das beschriebene Handeln der Lehrperson im Video zu beobachten ist. Dabei beantworten die Expert:innen die Items entsprechend der vierstufigen Likert-Skala von stimmt gar nicht bis stimmt genau. Die 72 Analyseaufträge der Kategorie des Beschreibens werden vollständig von fünf Rater:innen beurteilt. Fünf weitere Rater:innen bewerten die 156 Items aus den Kategorien des Erklärens und Vorhersagens. Ähnlich wie bei Meschede et al. (2015, S. 325) ist geplant, das Rating zum einen zum Validieren der Items und zum anderen als Vorlage für eine Musterlösung zu nutzen. Zunächst soll die Übereinstimmung der einzelnen Urteile der Expert:innen bestimmt werden. Dazu wird ermittelt, welche Antwortmöglichkeit pro Item am häufigsten durch die Rater:innen genannt wird. Wenn es eine größtmögliche Übereinstimmung gibt, so wird diese Antwort als „richtig“ festgelegt. Bei einer Abweichung wird die Tendenz betrachtet. Die Antworten der Expert:innen werden zudem nach einem zustimmenden und einem ablehnenden Pol dichotomisiert. Dies ist möglich, da es sich bei der Einteilung der vierstufigen Skala eindeutig um zwei Pole handelt, die auch von den Studierenden so wahrgenommen wird (Tepner und Dollny 2014, S. 317). Zu beachten ist, dass sich alle Antworten auf das Handeln von Lehrpersonen im komplexen Unterrichtsgeschehen beziehen. Da es schwierig ist, im Unterricht von der einen richtigen motivierenden Handlung auszugehen, wird bei der Auswertung auf dichotomisierte Antworten zurückgegriffen und bei der Auswertung die Tendenz betrachtet (Carlson 1990, S. 161). Sofern die Abweichung der Urteile der Expert:innen innerhalb der Tendenz liegt, so wird ebenfalls von einer Übereinstimmung ausgegangen (Tepner und Dollny 2014, S. 317). Bei Abweichungen, die diesen Bereich überschreiten, muss das Item überprüft werden.

„Geringe Beobachterübereinstimmung deutet darauf hin, dass das Beobachtungssystem unklar bzw. messungenau ist und hinsichtlich einzelner Kategorien überarbeitet werden muss und/oder dass mindestens ein Beobachter verzerrte Beurteilungen abgibt“ (Döring und Bortz 2016, S. 346). Um die Höhe der Übereinstimmung der Expert:innen zu bestimmen, wird Fleiss-Kappa berechnet. Dabei handelt es sich um ein Verfahren, dass bei mehr als zwei Rater:innen genauere Ergebnisse erzielt als der paarweise ermittelte Cohens Kappa-Koeffizient mit anschließendem Median-Score über alle Ratingpaare hinweg (Döring und Bortz 2016, S. 346; Fleiss 1971, S. 378; Roth 1995, S. 141). Zur Interpretation der Ergebnisse wird die in der Literatur verbreitete Einteilung nach Landis und Koch (1977, S. 164–165) verwendet:

Tabelle 17 Interpretation Fleiss-Kappa (nach Landis und Koch 1977, S. 164–165)

Kappa Statistic	Strength of Agreement
< 0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Bei einer zufriedenstellenden Übereinstimmung (Fleiss-Kappa > 0.61) werden die Antworten als Musterlösung verwendet und dienen als Grundlage zur Bewertung der Anwender:innen (siehe Anhang G).

8.2.3 Ergebnisse und Interpretation

Wie beschrieben erfolgt das Rating in zwei Schritten. Das Expert:innenrating zu den Items des Beschreibens erreicht entsprechend der Einteilung nach Landis und Koch (1977, S. 164–165) ein zufriedenstellendes Fleiss-Kappa Ergebnis (Fleiss' $\kappa = 0.7$). Im zweiten Expert:innenrating zu den Items des Erklärens und Vorhersagens ist die Übereinstimmung etwas höher (Fleiss' $\kappa = 0.78$). Diese Ergebnisse sprechen für eine gute Übereinstimmung der Expert:innen bezüglich der Beantwortung der Items. Damit decken sich die Ergebnisse mit bisherigen Forschungsarbeiten, die bei der Validierung der Analysefragen methodisch in gleicher Weise vorgegangen sind und im Einklang mit einer ähnlich hohen Übereinstimmung die Items ausgewertet haben. Dazu gehören beispielsweise die Studien zum OBSERVER von Seidel et al. (2010, S. 303) und der Arbeitsgruppe von Meschede et al. (2015, S. 325).

Resultat zur Forschungsfrage 2:

Die Expert:innen erreichen bei der Auswertung der Rating-Items eine gute Übereinstimmung. Es kann von einer ausreichend hohen Auswertungsobjektivität ausgegangen werden.

Insgesamt kann, wie in den früheren Studien davon ausgegangen werden, dass die Lösungen der Analyseaufträge zum Beschreiben, Erklären und Vorhersagen den theoretischen Kompetenzen entsprechen und somit als Mustervorlage genutzt werden können. Dabei soll folgendermaßen vorgegangen werden: Entspricht die Beantwortung eines Items genau der Meinung des Expertenratings, so werden zwei Punkte vergeben. Bei einer richtigen Tendenz gibt es einen Punkt, ansonsten null Punkte (Meschede et al. 2015, S. 327). In der Auswertung wird dementsprechend ein Partial-Credit-Modell (siehe Abschnitt 10.2.1 *Item-Response-Theorie*) verwendet.

8.3 Implementierung in ViviAn

Die Videovignetten mit den entsprechenden Analyseaufträgen zum Thema Motivation und Motivierung (siehe Anhang G) werden in das bestehende Video-Tool der RPTU ViviAn „Videovignetten zur Analyse von Unterrichtsprozessen“ implementiert. Das Videotool ist online über die Homepage <https://vivian-training.de> abrufbar. Durch die Verwendung des Tools können die Anwender:innen ihr Wissen, ihre Fähigkeiten und Fertigkeiten hinsichtlich der Professionellen Unterrichtswahrnehmung im Bereich der Motivation und Motivierung analysieren. Die Beschreibung der Lernumgebung von ViviAn mit dem Lehransatz und Lernzielen sind im Abschnitt 4.3 *Das Videotool ViviAn* zusammengefasst.

Im Rahmen der vorliegenden Arbeit konnten sieben Vignetten entwickelt werden (*Beschreibung der Vignetten* siehe Abschnitt 7.5). Entsprechend der Validierung des Videotools (siehe Kapitel 10) musste eine Vignette entfernt werden, somit werden final sechs Videos mit den dazu passenden, validierten Analyseaufträgen in das Videotool ViviAn eingebettet. Dabei ist jede Lernumgebung so aufbereitet, dass zusätzliche Informationen zur gezeigten Sequenz wie zum Beispiel verwendete Unterrichtsmaterialien oder zeitliche und inhaltliche Rahmenbedingungen integriert und über blaue Buttons jederzeit abrufbar sind (Bartel und Roth 2017, S. 48–49). Dadurch kann das Video besser in den Kontext der Klasse eingeordnet werden (von Aufschnaiter et al. 2017, S. 95). Die folgende Abbildung zeigt ein Beispiel für den Aufbau der Lernumgebung:



Abbildung 27 Lernumgebung ViviAn am Beispiel eines Videos zum Themenbereich Motivierung im Unterricht

Aktuell ist das Videotool im Bereich der Vignetten zur Motivation und Motivierung so konzipiert, dass es primär der Analyse der Fähigkeiten hinsichtlich der Professionellen Unterrichtswahrnehmung dienen kann. Die Nutzer:innen des Tools gelangen nach der

einmaligen Betrachtung der Videos über einen Link zu den Analysefragen. Das einmalige Betrachten der Videoszene entspricht eher einer realen Unterrichtssituation und soll die tatsächlichen Fähigkeiten in diesem Moment erfassen (von Aufschnaiter et al. 2017, S. 92). Auf Grund der Komplexität von Videos, kann das einmalige Betrachten des Videos auch zu einer kognitiven Überforderung der Studierenden führen (Syring et al. 2015, S. 680).

Perspektivisch soll aus dem Analysetool ein Trainingstool entwickelt werden, sodass unter anderem die Studierenden im Lehramt bereits die Möglichkeit haben, ihre Fähigkeiten der Professionellen Unterrichtswahrnehmung zu fördern. Dadurch können die Studierenden ihr bisher erworbenes Wissen zu fundierten Theorien mit dem praxisbezogenen Handeln verknüpfen. Das Videotool dient dann neben der Erfassung der Fähigkeiten auch der Förderung der Kompetenzen (von Aufschnaiter et al. 2017, S. 92). Durch ein wiederholtes Abrufen des Videos mit einer gemeinsamen Analyse und Diskussion des Videos kann die Beschreibung der Situation mit der wissenschaftlichen Erklärung und Vorhersage nachvollziehbar werden (von Aufschnaiter et al. 2017, S. 92).

9 Akzeptanz des Tools ViviAn – Pilotierungsstudie

Der erste Einsatz des Videotools innerhalb der ersten Phase der Lehrpersonenbildung dient dazu, die Akzeptanz der Studierenden in der Anwendung konkret zu den Kriterien Relevanz, Interessantheit und Gestaltung der Lernumgebung zu untersuchen. Ziel ist es, Aussagen über das Gütekriterium der Nützlichkeit (siehe Abschnitt 5.4 *Nebenkriterien*) treffen zu können. Dazu soll folgende Forschungsfrage beantwortet werden:

Forschungsfrage

- 3 *Können die Vignetten zum Themenbereich Motivierung bei den Studierenden das Interesse an der Arbeit mit Videos wecken sowie eine hohe Relevanz wahrnehmen lassen?*

Weiterhin sollen durch die Evaluation der Vignetten und Analyseaufträge konkrete Empfehlungen zur Optimierung des Tools abgeleitet werden, um vor der Haupterhebung Anpassungen vornehmen zu können.

9.1 Stichprobe

In der ersten Testphase wird das Videotool ViviAn bei Studierenden im Bachelor Seminar „Lernumgebungen gestalten“ im Sommersemester 2018 eingesetzt. Das Seminar findet in Form einer Blockveranstaltung mit jeweils zwei Freitags- und Samstagsterminen statt. Zwischen den beiden Blöcken liegen 14 Tage. Mit dem Videotool wird an den jeweiligen Samstagen im Rahmen der Veranstaltung gearbeitet. Beim ersten Einsatz bearbeiten die Studierenden insgesamt vier Vignetten zu den Schwerpunkten Relevanz und Leistungsdifferenzierung. Beim zweiten Einsatz kommen drei Vignetten zu den Themen Freiräume, Fehlerklima und Strukturierung zur Anwendung. Insgesamt nehmen 25 Studierende (17 weiblich, 8 männlich) an der Bearbeitung aller Vignetten teil.

9.2 Methodisches Vorgehen

Teil 1

Nach der ersten Bearbeitungsphase der Videos sollen die Studierenden die Arbeit mit dem Videotool evaluieren. Dazu wird zunächst ein reliabler und valider Fragebogen der Arbeitsgruppe Bartel und Roth (2020) eingesetzt, der unter anderem die Interessantheit, Relevanz und die Gestaltung des Tools überprüft (Bartel und Roth 2020, S. 304) (siehe Tabelle 18). Diese Aspekte sollen Auskunft über die Akzeptanz des Tools liefern.

Tabelle 18 Kurzbeschreibung des Fragebogens (Bartel und Roth 2020)

Skala	Items	Beispielitem
Einschätzung des Interesses an der Arbeit mit Videos	5	„Ich habe Interesse an der Bearbeitung der Videos.“
Einschätzung der Relevanz an der Arbeit mit Videos	5	„Ich halte die Bearbeitung der Videos im Hinblick auf meine berufliche Ausbildung als sinnvoll.“
Einschätzung der Gestaltung der Lernumgebung	4	„Ich finde die Benutzeroberfläche der Lernumgebung ViviAn übersichtlich.“

Alle Fragen werden über eine vierstufige Likert-Skala beantwortet (Bartel und Roth 2020, S. 305): Die Bewertung erfolgt auf einer bipolaren Skala, ausgehend von einem positiven Pol „trifft voll zu“ zu einem negativen Pol „trifft gar nicht zu“. Die Auswertung und Interpretation der geschlossenen Items erfolgt über die Bildung von Summenscores. Bei der Skala Relevanz und Interessantheit können Summenscores zwischen 0 (sehr niedrige Ausprägung) und 15 Punkten (sehr hohe Ausprägung) erreicht werden (Bartel und Roth 2020, S. 311). Die Gestaltung der Lernumgebung kann einen maximalen Summenscore von 12 (sehr hohe Ausprägung) ergeben (Bartel und Roth 2020, S. 312). Zur Interpretation der Ergebnisse werden deskriptiv die Lage- und Streuungsmaße in Form von Box-Plots dargestellt (Sedlmeier und Renkewitz 2008, S. 190). Innerhalb der Box-Plots wird ebenfalls das arithmetische Mittel angegeben.

Teil 2

Ein weiteres Instrument zur Beurteilung der Vignetten und Analyseaufträge (Seidel und Prenzel 2007) wird im Anschluss der zweiten Bearbeitungsphase eingesetzt. Am Ende der zweiten Blockveranstaltung sollen alle Vignetten und Items eingeschätzt werden. Der Fragebogen kommt ebenfalls in der Pilotierungsstudie des OBSERVERs als Evaluationsbogen zum Einsatz (Seidel et al. 2010, S. 301) und kann dementsprechend zum Vergleich herangezogen werden. Die Beantwortung der Items erfolgt über eine vierstufige Likert-Skala: Die Studierenden haben die Auswahl zwischen „Trifft nicht zu“, „Trifft eher nicht zu“, „Trifft eher zu“ und „Trifft zu“. Die Auswertung erfolgt deskriptiv durch die Bestimmung des arithmetischen Mittelwerts (Sedlmeier und Renkewitz 2008, S. 193) sowie über Häufigkeitsverteilungen (Sedlmeier und Renkewitz 2008, S. 185). Für eine vereinfachte Interpretation und Veranschaulichung werden die Antwortkategorien zusammengefasst zum ablehnenden Pol „trifft (eher) nicht zu“ und zustimmenden Pol „trifft (eher) zu“.

9.3 Ergebnisse

Teil 1

Die Mittelwerte der Summenscores der Skalen der Interessantheit und Relevanz weisen eine Ausprägung über den theoretischen Mittelwert (7,5 Punkte) auf. Die Interessantheit ($M = 9.69$, $SD = 3.17$) wird leicht geringer beurteilt als die Relevanz der Arbeit mit Videos ($M = 10.52$, $SD = 2.52$). Die Interessantheit weist zudem eine stärkere Streuung als die Relevanz auf (Abbildung 28 und 29). Die Studierenden des Seminars bewerten die Lernumgebung ViviAn im Mittelwert ($M = 9.56$, $SD = 2.20$) im oberen positiven Bereich (Abbildung 30).

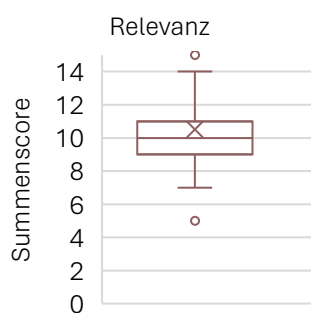


Abbildung 28 Einschätzung der Relevanz an der Arbeit mit Videos von den Studierenden des Seminars

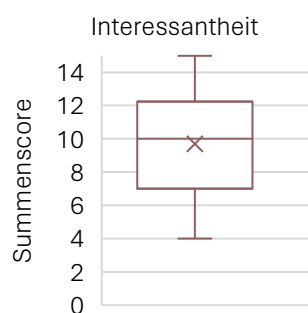


Abbildung 29 Beurteilung der Interessantheit an der Arbeit mit Videos von den Studierenden des Seminars

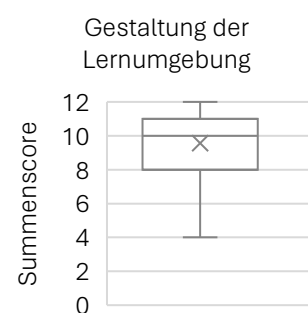


Abbildung 30 Einschätzung der Gestaltung der Lernumgebung von den Studierenden des Seminars

Teil 2

Analysiert werden zwei Abschnitte: die Beurteilung der Vignetten und die der Analyseaufträge. Die Kennwerte sind in der Tabelle 19 festgehalten.

1) Beurteilung der Vignetten

Die Videovignetten werden im arithmetischen Mittel als ergiebig, aussagekräftig, nicht zu kurz, typisch, authentisch und interessant beurteilt. Alle Werte befinden sich über dem theoretischen Mittelwert der Skala ($M = 1.5$) und werden dementsprechend als Zustimmung gesehen. Die prozentuale Häufigkeit der Antworten liegt zwischen 64% und 80% auf dem zustimmenden Pol. Hinsichtlich der prozentualen Verteilung zeigt sich bei der Einschätzung der Aussagekraft der Vignetten mit 40% Ablehnung und 60% Zustimmung die geringste Eindeutigkeit.

2) Beurteilung der Analyseaufträge

Bei den Analyseaufträgen zeigt sich ein weniger eindeutiges Bild. Der prozentuale Unterschied in den Antworthäufigkeiten zwischen Ablehnung und Zustimmung fällt bei der Beurteilung der Interessantheit, Erhellung und Produktivität am geringsten aus. Die Analyseaufträge werden zu 58% als interessant bzw. produktiv beurteilt und zu 46% als erhellend. Angemessenheit und Passung werden mit einer Häufigkeit von

72% als positiv beurteilt. Der Aspekt der Schwierigkeit deutet daraufhin, dass die Items als leicht empfunden (80%) werden. Die Bearbeitung der Vignetten wird im Mittel als nicht aufwendig, nicht anstrengend, technisch nicht schwierig und nicht zu umfangreich eingeschätzt.

Tabelle 19 Beurteilung der Vignetten und Analyseaufträge (M = Mittelwert, SD = Standardabweichung)

	prozentuale Häufigkeiten		M	SD
	Ablehnend: Trifft (eher) nicht zu	Zustimmend: Trifft (eher) zu		
Die Vignetten waren...				
ergiebig	25.0	75.0	1.92	0.65
aussagekräftig	40.0	60.0	1.80	0.87
nicht zu kurz	20.0	80.0	2.12	0.93
typisch	36.0	64.0	1.76	0.66
authentisch	28.0	72.0	1.92	0.70
interessant	28.0	72.0	1.84	0.62
Die Analyseaufträge zu den Vignetten waren...				
produktiv	41.7	58.3	1.58	0.78
nicht schwierig	12.0	78.0	2.16	0.62
interessant	41.7	58.3	1.63	0.71
angemessen	28.0	72.0	1.84	0.85
erhellend	54.2	45.8	1.42	0.58
passend	28.0	72.0	1.84	0.75
Die Beantwortung der Analyseaufträge war...				
zu umfangreich	70.8	29.2	1.13	0.99
technisch nicht schwierig	16.0	84.0	2.40	0.76
inhaltlich schwierig	80.0	20.0	0.96	0.68
anstrengend	68.0	32.0	1.20	0.96
aufwendig	84.0	16.0	0.96	0.84
herausfordernd	76.0	24.0	1.16	0.55
anregend	40.0	60.0	1.56	0.58
abwechslungsreich	76.0	24.0	1.04	0.68

9.4 Interpretation und Schlussfolgerungen

Die Beurteilung des vorliegenden Videotools ViviAn zum Themenbereich der Motivierung zeigt anhand des Fragebogens der Arbeitsgruppe Bartel und Roth (2020), dass die Studierenden das Videotool als interessant, relevant und gut gestaltet wahrnehmen. „Damit sich (angehende) Lehrpersonen auf die Analyse der Vignette einlassen können, muss diese als angemessen realitätsnah und relevant erlebt werden“ (von Aufschnaiter et al. 2017, S. 98). Mit den Ergebnissen der Beurteilung der wahrgenommenen Relevanz kann gezeigt werden, dass das vorliegende Instrument „zur Lösung praktischer Probleme (praktische Relevanz) bei[trägt]“ (Döring und Bortz 2016, S. 90). Damit wird dem Gütekriterium der Nützlichkeit über die Evaluation der Studierenden zugestimmt. Insgesamt

wird das Tool positiv bewertet. Die Resultate sind konsistent mit den bisherigen Befunden zum Videotool ViviAn von Bartel und Roth (2020).

Resultat zur Forschungsfrage 3:

Die Forschungsfrage hinsichtlich der Nützlichkeit kann positiv beantwortet werden: Die Studierenden zeigen eine gute Akzeptanz des Videotools. Es wird als interessant und relevant beurteilt.

Die differenzierten Ergebnisse der Beurteilung der Videovignetten und Analyseaufträge anhand des Fragebogens von Seidel und Prenzel (2007) können als Grundlage für die Anpassung der Items genutzt werden. Im Vergleich erzielt der Einsatz des Evaluationsbogens in der vorliegenden Erhebung ähnliche Ergebnisse wie die Pilotierungsstudie von Seidel et al. (2010, S. 303–304). Die Vignetten werden im arithmetischen Mittel beim OBSERVER leicht besser beurteilt. Der aufsummierte Mittelwert der Items zur Beurteilung der Vignetten beträgt beim vorliegenden Tool $M_{Sum} = 11.36$, beim OBSERVER $M_{Sum} = 13.75$. Bei beiden Instrumenten zeigt sich, dass die Analyseaufträge als nicht schwierig eingestuft werden und die Bearbeitung nicht aufwendig bzw. nicht anstrengend ist (Vgl. Seidel et al. 2010, S. 303–304). Inwiefern diese Wahrnehmung die Ergebnisse hinsichtlich der erreichten Punkte widerspiegeln, muss überprüft werden.

Im Anschluss an die erste Pilotierungsstudie werden die Analyseaufträge hinsichtlich ihrer Formulierungen unter anderem auf Präzision und Wertfreiheit überprüft und angepasst (Beispiele Tabelle 20). Zudem wird der Itemstamm der Items des Erklärens und Vorhersagens für eine klare Struktur und bessere Verständlichkeit voneinander getrennt, da sie zuvor zusammengefasst waren. Die adaptierte Version des Tools wird für den nächsten Schritt der Validierung verwendet.

Tabelle 20 beispielhafte Anpassung der Items

Itemformulierung alt	Itemformulierung neu
Die Lehrperson bringt auch trockene Themen interessant rüber.	Die Lehrperson bringt das Thema interessant rüber.
...um den Schülerinnen und Schülern für die Bearbeitung der Aufgabe und somit für die Erreichung des Lernziels die erforderliche Zeit einzuräumen	1) ...um den Schülerinnen und Schülern für die Bearbeitung der Aufgabe die erforderliche Zeit einzuräumen 2) ...um den Schülerinnen und Schülern die Erreichung des Lernziels zu ermöglichen.
Schätzen Sie nun ein, warum die Lehrperson so gehandelt hat und welche Auswirkungen ihr Verhalten auf den weiteren Lehr-Lernprozess haben kann.	1) Beurteilen <u>Sie</u> , welche der folgenden ERKLÄRUNGEN bezüglich der Motivationsförderung theoretisch angemessen sind. Die Lehrperson [...] 2) Beurteilen <u>Sie</u> , welche AUSWIRKUNGEN [...] für den weiteren Lehr-Lernprozess hat.

10 Validierung des Videotools Teil I: Raschskalierung

Im folgenden Abschnitt wird ausführlich auf die Haupterhebung zur Validierung des Instruments eingegangen. Ziel ist es, die Gültigkeit des angenommenen Modells zu überprüfen. Lässt sich mittels eines Rasch-Modells die dreidimensionale Struktur der professionellen Unterrichtswahrnehmung als Beschreiben, Erklären und Vorhersagen abbilden? Andernfalls soll ermittelt werden, welche Dimensionalität angenommen werden kann. Für die Überprüfung der Modellpassung und -güte wird das Verfahren der Stichprobenhalbierung angewendet (Rost 2004, S. 73). Nach erfolgter Raschskalierung mit einer Hälfte der Stichprobe, wird überprüft, ob das ermittelte Modell auch für die zweite Hälfte der Stichprobe angenommen werden kann. Folgende Forschungsfragen sollen in diesem Abschnitt beantwortet werden:

Forschungsfragen

- 4 *Bilden die Items die Skalen des Knowledge-based Reasonings reliabel ab?*
- 5 *Zeigen die Items modellkonforme Itemfitwerte?*
- 6 *Differenziert das vorliegende Instrument zwischen den drei Phasen des Knowledge-based Reasoning: Beschreiben, Erklären und Vorhersagen?*
- 7 *Lassen sich die Ergebnisse an einer zweiten Stichprobe replizieren?*

10.1 Stichprobe

Die Stichprobe der zweiten Validierung setzt sich aus Studierenden der Universität Koblenz-Landau vom Campus Landau mit dem Schwerpunkt Lehramt ($N = 239$) und Lehramtsanwärter:innen (LAA) aus dem staatlichen Studienseminar für das Lehramt an Realschulen plus der Teildienststelle Landau-Land ($N = 41$) zusammen.

Tabelle 21 Verteilung nach Semestern – Gesamtstichprobe (TN = Anzahl der teilnehmenden Personen pro Semester, LAA = Lehramtsanwärter:innen)

Semester	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	LAA
TN	17	92	20	30	11	25	6	18	6	6	3	1	1	1	1	1	41

Die Lehramtsanwärter:innen sind zum Zeitpunkt der Erhebung seit sieben Monaten im Vorbereitungsdienst. Aufgrund der Corona-Pandemie fand der Unterricht im Studienseminar hauptsächlich digital und in reduzierter Form statt. Zudem konnte der praktische Einsatz in den Schulen nicht wie geplant erfolgen, sodass die Lehramtsanwärter:innen kaum bzw. keine zusätzlichen praktischen Erfahrungen sammeln konnten.

Die Stichprobe wird für die Raschskalierung und Überprüfung der Modellgüte in zwei gleichgroße Stichproben geteilt ($N_1 = 140$, $N_2 = 140$). Das Kriterium für die Aufteilung ist die gleichmäßige Verteilung der Semester, in denen sich die Studierenden befinden, bzw. die gleichmäßige Verteilung der Lehramtsanwärter:innen (siehe Tabelle 22). Damit soll vermieden werden, dass die Beantwortung des Tests nicht durch Personeneigenschaften beeinflusst wird (Rost 2004, S. 73). Denkbar wäre, dass mit zunehmender Anzahl an Semestern gewonnene Erfahrungen und angeeignetes Wissen andere Beurteilungsprozesse bedingen. Die Zuordnung zu den Gruppen erfolgt zufällig. Die Stichprobe N_1 wird für die Rasch-Skalierung verwendet. Zur Überprüfung der Modellpassung dient die Stichprobe N_2 .

Tabelle 22 Verteilung nach Semestern - Teilstichproben N1 und N2 (TN = Anzahl der teilnehmenden Personen pro Semester, LAA = Lehramtsanwärter:innen)

Semester	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	LAA
TN - N1	9	46	10	15	5	13	3	9	3	3	1	1	0	1	0	1	20
TN - N2	8	46	10	15	6	12	3	9	3	3	2	0	1	0	1	0	21

10.2 Methodisches Vorgehen

In diesem Kapitel wird für die Auswertung die zu Grunde gelegte *Item-Response-Theorie* (Abschnitt 10.2.1) kurz dargestellt. Im Anschluss erfolgt die differenzierte Beschreibung der Raschskalierung mit den verwendeten Itemkennwerten zur Überprüfung der Skaleneigenschaften und Validierung des Videotools.

10.2.1 Item-Response-Theorie

Neben der Klassischen Testtheorie (KTT) gilt die Item-Response-Theorie (IRT) als eine der meist angewendeten Testtheorien zur Entwicklung und Auswertung von Testverfahren (Moosbrugger et al. 2020, S. 252).

Die Klassische Testtheorie ist eine Messfehlertheorie, die davon ausgeht, dass der beobachtete Testwert einer Person einem Messfehler unterliegt (Steyer und Eid 2001, S. 102). Ziel ist es, mit der KTT den wahren Messwert (true score) zu berechnen (Steyer und Eid 2001, S. 102; Döring und Bortz 2016, S. 462). Dabei wird die Beantwortung eines Items nicht mit dem Merkmal assoziiert (Bühner 2011, S. 53). Dies erfolgt im Gegensatz dazu bei der Item-Response-Theorie (Steyer und Eid 2001, S. 215). Mit der Beantwortung der Items können Folgerungen auf die Personenmerkmale (latente Variablen) getroffen werden (Hartig und Goldhammer 2010, S. 3). Zentral ist bei der Item-Response-Theorie die Bestimmung der Wahrscheinlichkeit der Beantwortung bzw. Lösung eines Items (*item responses*) in Abhängigkeit von den Item- und Personenmerkmalen (Döring und Bortz 2016, S. 462; Kelava und Moosbrugger 2020b, S. 371; Bühner 2011, S. 494; Steyer und Eid 2001, S. 215). Die Itemschwierigkeit stellt das maßgebliche Itemmerkmal dar (Hartig und

Goldhammer 2010, S. 3). Im Gegensatz zur KTT kann die Itemschwierigkeit unabhängig vom Fähigkeitsniveau der Testpersonen (viele oder wenige richtige Antworten) ermittelt werden (Hartig und Goldhammer 2010, S. 5). Die IRT bietet im Gegensatz zu anderen Methoden den Vorteil, dass die Item- und Personenparameter auf der gleichen Skala abgebildet werden und dadurch direkt in Verbindung gesetzt werden können (Bühner 2011, S. 495). Beide Parameter werden in Logiteinheiten, ermittelt durch eine sogenannte Logit-Transformation, angegeben (Bühner 2011, S. 495). Es können Werte zwischen $-\infty$ und $+\infty$ erreicht werden (Rost 2004, S. 118). Positive Werte kennzeichnen Personen mit einer hohen Fähigkeit bzw. Items mit einer hohen Schwierigkeit, während negative Werte gegenteilige Ergebnisse aussagen (Bühner 2011, S. 496). Das Besondere beim Rasch-Modell ist der Aspekt der suffizienten Statistik der Personenparameter. Dies bedeutet, dass die Beurteilung der Personenfähigkeit unabhängig von der Beantwortung bestimmter Items ist, sondern allein von der Anzahl korrekt gelöster Items (Strobl 2012, S. 15). Je mehr Antworten eine Person richtig gelöst hat, desto höher ist die Fähigkeit ausgeprägt. Die Lösungswahrscheinlichkeit eines Items wird mit $p(x_{vi} = 1)$ bezeichnet (Rost 2004, S. 116). Damit wird die Wahrscheinlichkeit ermittelt, inwiefern eine Person v ein Item i richtig lösen bzw. beantworten kann.

Eine Voraussetzung für die Anwendung von IRT-Modellen ist die Ordnung der Kategorien (Hartig und Goldhammer 2010, S. 4). Diese kann durch dichotome oder polytome (z.B. als Partial-Credit-Modell) Antwortkategorien hergestellt sein (Hartig und Goldhammer 2010, S. 4). Tests mit dichotomen und ordinalen Antwortkategorien werden häufig mittels der IRT analysiert (Moosbrugger et al. 2020, S. 252). Es kommen verschiedene IRT-Modelle zur Schätzung der Wahrscheinlichkeit in Betracht (Rost 2004, S. 133). Das 1PL-Modell (Ein-Parameter-Logistisches-Modell) bezieht die latenten Personenmerkmale bzw. -fähigkeiten θ (Theta) sowie die Itemschwierigkeit σ (Sigma) bei der Berechnung der Wahrscheinlichkeit ein (Geiser und Eid 2010, S. 313). Ein 2PL-Modell wird durch den Itemparameter der Trennschärfe und ein 3PL-Modell zusätzlich zur Trennschärfe durch die Ratewahrscheinlichkeit erweitert (Bühner 2011, S. 514). Nachteil bei der Berücksichtigung zusätzlicher Parameter ist die Notwendigkeit einer größeren Stichprobe ($N = 500-1000$) (Eid und Schmidt 2014, S. 214). Diese liegt für die vorliegende Validierung der Analyseaufträge nicht vor. Zusätzlich ist beim 3-PL-Modell nachteilig, dass aufgrund der drei miteinander verbundenen Itemparameter die Schätzungen instabil werden können (Eid und Schmidt 2014, S. 210). Zu den 1PL-Modellen gehören das dichotome Rasch-Modell sowie dessen Erweiterungen für ordinale Daten wie das Ratingskalen-Modell und das Partial-Credit-Modell (Moosbrugger et al. 2020, S. 266). Da im vorliegenden Videotool Partial-Credit-Items verwendet werden, wird der Fokus in der Beschreibung auf dieses Modell gelegt. Auf das 2PL- und 3PL-Modell sowie das dichotome Raschmodell als Hauptvertreter des 1PL-Modells wird in dieser Arbeit nicht weiter eingegangen.

Das Partial-Credit-Modell

Das Partial-Credit-Modell ist zurückzuführen auf Masters (1982). Im Vergleich zum dichotomen Modell soll mit dem Partial-Credit-Modell eine präzisere Beurteilung der Personenfähigkeit möglich sein (Masters 1982, S. 150). Es wird nicht nur zwischen 0 – falsche Antwort und 1 – richtige Antwort differenziert, sondern es ermöglicht eine stufenweise Bewertung mit weiteren Punktkategorien. Die Kategorien im vorliegenden Modell sind dreistufig angelegt und wie folgt kodiert: 0 – falsche Antwort, 1 – teilrichtige Antwort, 2 – richtige Antwort. Diese drei Stufen sind geordnet. Das bedeutet, „dass mit der Wahl einer höheren Antwortkategorie auch eine höhere Eigenschaftsausprägung einhergeht.“ (Bühner 2011, S. 515) Der Übergang von einer Kategorie zur nächsten wird als Schwelle bezeichnet (Rost 2004, S. 205). Die Wahrscheinlichkeit einer Person p_s mit einer bestimmten Fähigkeit θ_v ein Item mit einer bestimmten Schwierigkeit σ_{ic} zu lösen bzw. eine Schwelle zu überschreiten kann mit folgender Formel berechnet werden:

$$p_s = \frac{e^{\theta_v - \sigma_{ic}}}{1 + e^{\theta_v - \sigma_{ic}}}$$

θ_v = Fähigkeit einer Person v

σ_{ic} = Schwierigkeit eines Items i der Kategorie c

(Eid und Schmidt 2014, S. 232; Rost 2004, S. 206)

Neben der Schwellenwahrscheinlichkeit p_s kann auch die Kategorienwahrscheinlichkeit p ermittelt werden. Anders als beim Rasch-Modell müssen beim Partial-Credit-Modell aufgrund der zusätzlichen Antwortkategorie(n) zur Bestimmung der Kategorienwahrscheinlichkeit Summen über jede Antwortkategorie eines Items berechnet werden (Strobl 2012, S. 55). Eine genaue Herleitung der Kategorienwahrscheinlichkeit und des Zusammenhangs mit der Schwellenwahrscheinlichkeit lässt sich bei Rost (2004, S. 207–209) finden.

Die Lösungswahrscheinlichkeit für ein Item kann mithilfe der CCC (Category characteristic curve) pro Kategorie in Abhängigkeit von der Personenfähigkeit grafisch dargestellt werden (Strobl 2012, S. 55–56) (Abbildung 31). Anhand dieser Abbildung können verschiedene Eigenschaften der CCC abgelesen werden:

1. Die Wahrscheinlichkeit in der Kategorie 0 zu antworten, nimmt mit zunehmender Personenfähigkeit ab (Vergleich $P(x_1 = 0)$), während die Wahrscheinlichkeit, die Kategorie 1 oder 2 zu erreichen, steigt. Wobei die Kategorie 1 als mittlere Kategorie (Vergleich $P(x_1 = 1)$) nur bis zu einem bestimmten Fähigkeitswert zutrifft und danach die Wahrscheinlichkeit wieder abnimmt bzw. die Wahl der Antwortkategorie 2 zunimmt (Eid und Schmidt 2014, S. 237). In der Abbildung 31 ist beispielsweise für eine Personenfähigkeit $\theta = 0$ die Wahrscheinlichkeit, in Kategorie 1 zu antworten, am höchsten.

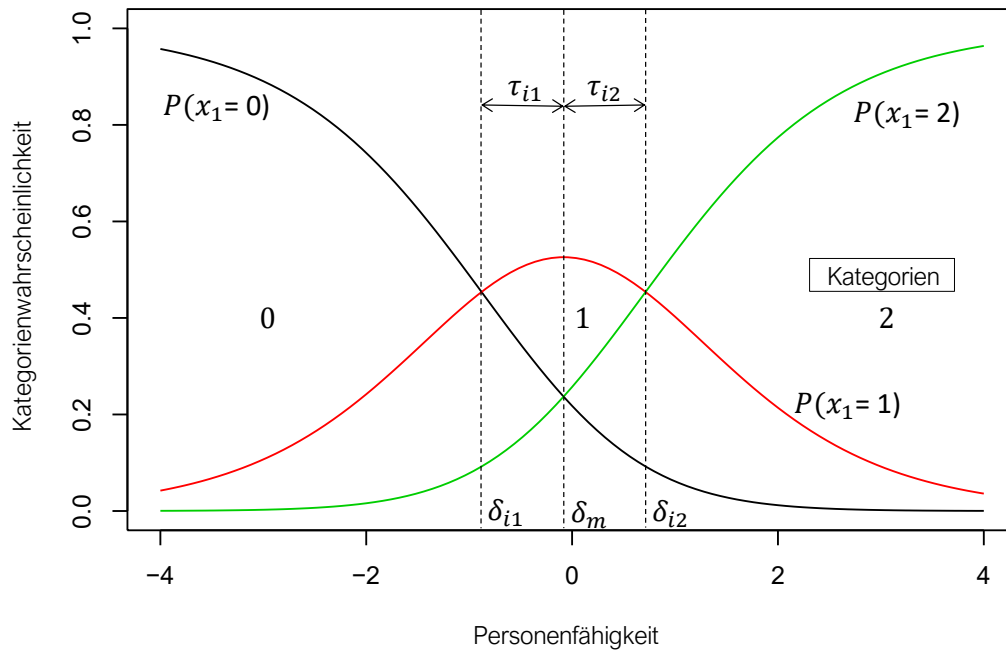


Abbildung 31 Category Characteristic Curve (δ = Schwellenparameter, τ = Abstand zwischen den Schwellenparametern)

2. Die Punkte, in denen sich die Kategorienkurven schneiden, stellen die Schwellen dar. An jeder Schwelle ist die Wahrscheinlichkeit für das Erreichen in einer der beiden Antwortkategorie gleich groß (Bühner 2011, S. 516). Der Schwellenparameter δ_{i2} (Delta) markiert den Schnittpunkt zwischen Kategorie 1 und 2 (Eid und Schmidt 2014, S. 237). Er entspricht der Schwierigkeit, die Schwelle zu überwinden (Eid und Schmidt 2014, S. 233). Dazu wird ein Lot vom Schnittpunkt der beteiligten Kurven auf die x-Achse gefällt. An der Schnittstelle des Lotes mit der Rechtswertachse lässt sich die zugehörigen Personenfähigkeit ablesen (siehe Abbildung 31). Die mittlere Itemschwierigkeit des Items kann mit dem Schwellenparameter δ_m , als Schwelle zwischen Kategorie 0 und 2, angesehen werden (Wu et al. 2016, S. 168).
- Es wird erwartet, dass die Schwellenparameter geordnet sind (Rost 2004, S. 205): $\delta_{i1} < \delta_{i2}$. Andernfalls passen die aufsteigenden Personenfähigkeiten nicht zu den aufeinander folgenden Kategorien (Rost 2004, S. 205). In der Literatur finden sich gegenläufige Meinungen, die zumindest beim Partial-Credit-Modell eine ungeordnete Struktur der Schwellen erlauben (Bühner 2021, S. 326). Eine mangelnde Ordnung der Schwellenparameter kann mit dem Antwortverhalten der Personen in Form von Meidung einer Kategorie erklärt werden und bedeutet nicht zwingend eine Verletzung des Partial-Credit-Modells (Wu et al. 2016, S. 164–165). In der Abbildung 32 ist beispielhaft eine Unordnung der Schwellenparameter $\delta_{i1} > \delta_{i2}$ dargestellt.

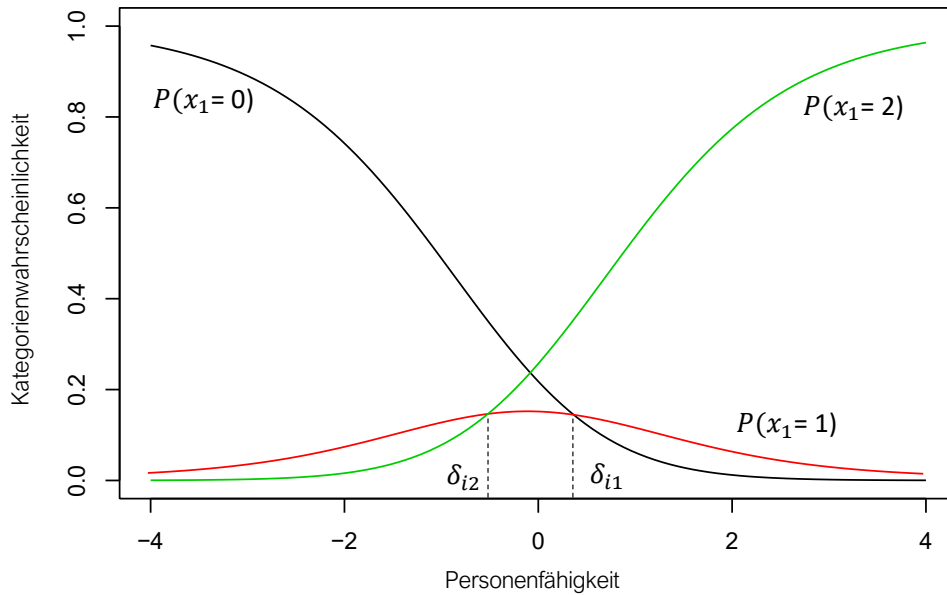


Abbildung 32 ungeordnete Schwellenparameter (eigene Darstellung, in Anlehnung an (Wu et al. 2016, S. 164)

Die Kategorie 1 wird weniger häufig als Antwortkategorie genutzt, dadurch verschieben sich die Schwellenparameter in eine Unordnung. Dennoch bleibt die Annahme bestehen, dass an der Schwelle δ_{i1} die Personen mit zunehmender Personenfähigkeit eher die Kategorie 1 wählen als die Kategorie 0 bzw. an der Schwelle δ_{i2} die Personen mit zunehmender Personenfähigkeit eher die Kategorie 2 wählen als die Kategorie 1 (Wu et al. 2016, S. 164). Dies führt laut Wu et al. (2016, S. 164) zu keiner Verletzung des Partial-Credit-Modells. Anhand der Häufigkeiten der Antworten kann festgestellt werden, ob eine Antwortkategorie unterbesetzt ist. Das Partial-Credit-Modell ist in der Lage, mit derartigen Verteilungsmustern arbeiten zu können (Bühner 2021, S. 326). Items mit ungeordneten Schwellenparametern sollten dennoch inhaltlich bzw. sprachlich überprüft werden (Bühner 2021, S. 327).

- Die Abstände τ (Tau) zwischen den Schwellenparametern müssen beim Partial-Credit-Modell nicht gleich sein (Wu et al. 2016, S. 168). Das bedeutet, der Abstand τ_{i1} muss nicht dem Abstand τ_{i2} entsprechen und kann damit ungeordnet sein. Begründet werden kann dies damit, dass die Überwindung einer Schwelle höhere oder niedrigere Personenfähigkeiten erwartet und dementsprechend die Abstände unterschiedlich sein können (Hartig und Goldhammer 2010, S. 16). Im Vergleich soll beim Ratingskalenmodell eine Ordnung zwischen den τ -Werten vorhanden sein (Strobl 2012, S. 63).

Die Schätzung der Parameter sowie die verwendeten Richtwerte der Analyse sind im *Abschnitt 10.2.2 Raschskalierung* detailliert aufgeführt. Aufgrund der theoretischen Annahme, dass die Professionelle Unterrichtswahrnehmung die Dimensionen Beschreiben, Erklären und Vorhersagen abdeckt (Seidel et al. 2010; Sherin und van Es 2008; van Es 2009), wird von einem mehrdimensionalen Rasch-Modell ausgegangen. Das Partial-Credit-Modell lässt eine multidimensionale Modellierung zu (Eid und Schmidt 2014, S. 353), daher ist dies kein Ausschlusskriterium für die Verwendung des Partial-Credit-Modells. Die Zuordnung eines Items erfolgt gezielt einer Dimension (Rost 2004, S. 259). Es handelt sich um einen strukturprüfenden Ansatz mit dem hypothesengeleitet die Mehrdimensionalität überprüft wird (Moosbrugger und Kelava 2020a, S. 34–35). Bei der Analyse der Item- und Personenfähigkeitsparameter wird die dimensionale Zuweisung an entsprechenden Stellen vorgenommen (z.B. zur Berechnung der Personenfähigkeit – WLE).

10.2.2 Raschskalierung

Zur Überprüfung, ob das theoretische Modell der Professionellen Unterrichtswahrnehmung auf die Daten übertragbar ist, wird ein 1PL-Rasch-Modell verwendet.

Für die Berechnungen wird das Programm R (R Core Team 2019) unter Verwendung der Benutzeroberfläche RStudio (RStudio Team 2019) eingesetzt. R bietet die Möglichkeit, diverse statistische Auswertungen vorzunehmen. Um die entsprechenden Analysen durchführen zu können, werden einzelne Pakete „packages“ heruntergeladen und damit das Programm erweitert (R Core Team 2019).

In einem ersten Schritt werden die aufbereiteten Daten deskriptiv mit der Funktion *describe* aus dem Paket *psych* (Revelle 2020) ausgewertet. Dazu werden die 166 Items hinsichtlich der Mittelwerte, der Varianz und der Range betrachtet. Um mögliche Boden- oder Deckeneffekte auszuschließen, werden alle Items mit Mittelwerten $M > 1.9$ (95%) und $M < 0.1$ (5%) selektiert. Diese Items deuten darauf hin, dass sie von den Personen „zu einfach“ bzw. „zu schwer“ gelöst werden können (Rost 2004, S. 92).

Um Items mit einer geringen oder keiner Streuung aufzudecken ($Var < 0.1$), wird die Itemvarianz ermittelt. „Unter Itemvarianz versteht man ein Maß für die Differenzierungsfähigkeit eines Items i in der untersuchten Stichprobe.“ (Kelava und Moosbrugger 2020a, S. 151) Die Itemvarianz $Var(y_i)$ kann mit folgender Formel berechnet werden:

$$Var(y_i) = \frac{\sum_{v=1}^n (y_{vi} - \bar{y}_i)^2}{n}$$

i - Item

v - Person

(Kelava und Moosbrugger 2020a, S. 152)

Die Range (Spannweite) gibt Aufschluss darüber, wie viele Antworten besetzt sind. Sie wird definiert als „die Differenz aus dem höchsten beobachteten Testwert Y_{max} und dem niedrigsten beobachteten Testwert Y_{min} “ (Kelava und Moosbrugger 2020c, S. 161). Eine

Range von 1 deutet bei einem Partial-Credit-Modell mit drei Kategorien darauf hin, dass eine Kategorie nicht belegt ist. Eine Unterbesetzung einer Antwortmöglichkeit, d.h. eine Kategorie wird von keiner Person aus der Stichprobe zur Beantwortung des Items genutzt, wird als Null-Kategorie bezeichnet (Wilson und Masters 1993, S. 87). Im Fall, dass die Kategorie 1 nicht belegt ist, kann dennoch eine Range von 2 (bei $Y_{min} = 0$ und $Y_{max} = 2$) auftreten. Um zu überprüfen, ob alle drei Kategorien besetzt sind, wird zusätzlich die Anzahl der Kategorien (N_{Kat}) ermittelt. Dies ist notwendig, da eine Skalierung in R automatisch die Verwendung eines Items mit einer Null-Kategorie ausschließt. Es gäbe die Möglichkeit, sofern dies begründet werden kann, diesen Ausschluss zu umgehen (Wilson und Masters 1993, S. 87).

Im nächsten Schritt soll festgestellt werden, welche dimensionale Struktur sich am besten für die Modellskalierung eignet. Dazu werden Modelle mit verschiedenen Dimensionsannahmen aufgestellt und verglichen. Zunächst wird das theoretisch angenommene dreidimensionale Modell mit der Bezeichnung „mod_B+E+V“ aufgestellt. Dieses basiert auf dem theoretischen Modell der Professionellen Unterrichtswahrnehmung und gliedert sich in die Dimensionen des Beschreibens, Erklärens und Vorhersagens (Seidel et al. 2010; Sherin und van Es 2008; van Es 2009). Alle Items werden den drei Dimensionen zugewiesen: Entsprechend der Konzeption werden alle Items mit „BS“ im Itemcode (z.B. V1**BS**1) der Dimension F1=Beschreiben, alle Items mit „E“ im Itemcode (z.B. V2**E**11) der Dimension F2=Erklären und alle Items mit einem zweiten „V“ im Itemcode (z.B. V6**V**41) der Dimension F3=Vorhersagen zugeordnet.

Mittels der Funktion *tamaan* aus dem Paket *TAM* (Robitzsch et al. 2020b) wird das Modell analysiert. Die Funktion bietet die Möglichkeit, verschiedene IRT-Modelle zu berechnen (Robitzsch et al. 2020a, S. 199). Mit dem Argument *tam.method* kann das Modell spezifiziert werden (Robitzsch et al. 2020a, S. 200). Dazu wird für das Rasch-Messmodell bzw. Partial-Credit-Modell der Befehl *tam.method*="tam.mml" verwendet (Robitzsch et al. 2020a, S. 113). Insgesamt werden auf diese Weise fünf Modelle mit unterschiedlichen Annahmen der Dimensionalität aufgestellt. Die Tabelle zeigt die Zuordnung der Dimensionen der einzelnen Modelle.

Tabelle 23 aufgestellte Lavaan-Modelle für Modellvergleiche (F = Faktor, B = Beschreiben, E = Erklären, V = Vorhersagen)

Modell	Dimensionen
mod_B+E+V	F1=Beschreiben, F2=Erklären, F3=Vorhersagen.
mod_B+E/V	F1=Beschreiben, F2=Erklären und Vorhersagen
mod_B/E/V	F1=Beschreiben, Erklären und Vorhersagen
mod_B/E+V	F1=Beschreiben und Erklären, F2=Vorhersagen
mod_B/V+E	F1=Beschreiben und Vorhersagen, F2=Erklären

Der Vergleich zwischen den Modellen erfolgt mittels der Funktion *anova* aus dem Paket *TAM* (Robitzsch et al. 2020b). Dabei wird schrittweise vorgegangen und das besser passende Modell mit dem nächsten Modell verglichen. Zur Interpretation der Ergebnisse werden die Informationskriterien AIC und BIC betrachtet (Bühner 2006, S. 352). Beide Kriterien können nur Aufschluss darüber geben, welches Modell besser zu den Daten passt und sagen nicht, welches Modell als perfekt gilt (Rost 2004, S. 89) bzw. welche Qualität das Modell besitzt (Bühner 2006, S. 352). Das Akaike Information Criterion (AIC) ist „ein Maß für die Anpassungsgüte des geschätzten Modells an die vorliegenden empirischen Daten (Stichprobe) unter Berücksichtigung der Komplexität des Modells.“ (Moosbrugger und Kelava 2020b, S. 744) Es wird anhand der logarithmierten Likelihood-Funktion und der Anzahl der Parameter ohne Gewichtung berechnet (Rost 2004, S. 340).

$$AIC = -2\ln \cdot L + 2n_p$$

$$BIC = -2\ln \cdot L + (\ln N) \cdot n_p$$

L – Likelihood

n_p – Anzahl der Parameter des entsprechenden Modells

N – Stichprobengröße

(Bühner 2011, S. 542)

Das Bayesian Information Criterion (BIC) dient ebenfalls als Vergleichskriterium zur Beurteilung der Passung des Modells an die Daten, bevorzugt jedoch Modelle mit weniger Komplexität bei gleicher Eignung gegenüber komplexen Modellen (Moosbrugger und Kelava 2020b, S. 744). Zusätzlich werden beim BIC die Parameterzahl mit der logarithmierten Stichprobengröße (N) gewichtet (Rost 2004, S. 342). Im Vergleich zwischen zwei Modellen sprechen niedrigere Werte beider Informationskriterien für eine bessere Modellpassung (Bühner 2006, S. 352). Das Modell mit der besten Passung wird mit der Bezeichnung „model1“ für die anschließende Modellspezifizierung verwendet. Beim Konvergieren des Modells werden bereits die im ersten Schritt extrahierten Items (siehe Abschnitt 8.4.3) nicht berücksichtigt. Zur Analyse der Modellpassung werden verschiedene Kriterien herangezogen und im Folgenden näher erläutert.

Erste Ergebnisse können über die Funktion *summary* des Pakets *base* (R Core Team 2019) abgerufen werden. Unter anderem wird hierbei die Korrelationsmatrix zwischen den Dimensionen ausgegeben. Unter Korrelation versteht „man den (linearen) Zusammenhang zwischen zwei quantitativen Variablen“ (Rost 2004, S.34). Die Korrelation kann dabei positive und negative Zusammenhänge zwischen -1 und +1 aufzeigen (Rost 2004, S. 34). Zur Beurteilung des Korrelationskoeffizienten wird häufig die Einteilung nach Cohen (1977) verwendet: Eine kleine Effektstärke liegt ab $r = 0.1$ vor, eine mittlere Effektstärke ab $r = 0.3$ und eine große Effektstärke ab $r = 0.5$ (Cohen 1977, S. 79–80).

Die dargestellten EAP-Reliabilitäten (expected a posteriori, R_{EAP}) geben „aufgrund der latenten Verteilung“ Auskunft über die „optimalen Schätzwerte für die individuellen Messwerte“ (Rost 2004, S. 382). Zur Beurteilung der EAP-Reliabilität kann ein Vergleich mit den Kennwerten von Cronbachs α herangezogen werden (Rost 2004, S. 382). Dabei muss berücksichtigt werden, dass es bei der EAP-Reliabilität zu einer Unterschätzung des Reliabilitätswerts kommen kann (Rost 2004, S. 382). In der Literatur findet man hinsichtlich der Interpretation des Cronbachs α häufig Angaben, dass Werte ab $\alpha = 0.8$ als gut und $\alpha = 0.7$ als akzeptabel gelten (Vgl. Pospeschill 2010, S. 21). Für Fähigkeitstest sprechen Werte ab $\alpha = 0.7$ für eine gute Reliabilität (Field et al. 2012, S. 806).

Anschließend werden die Itemparameter analysiert. Diese geben Auskunft über die Schwellenschwierigkeiten einer jeden Itemkategorie, die eine aufsteigende Ordnung haben sollten (Lamprianou 2020, S. 164; Rost 2004, S. 210). Als Schwelle werden die Schnittpunkte zwischen den Antwortkategorien bezeichnet (Rost 2004, S. 205) (siehe Abschnitt 10.2.1). Als alternatives Kriterium für die Beurteilung der Schwierigkeit können die Thurstonian Thresholds betrachtet werden (Wu et al. 2016, S. 171–172). Diese bilden die Schnittpunkte zwischen den Antwortkategorien ab, bei denen die Wahrscheinlichkeit von 50% vorliegt, diese oder die höhere Antwortkategorie zu wählen (Wu et al. 2016, S. 170; Trendtel et al. 2016, S. 195). Beim Partial-Credit-Modell sind die Thurstonian Thresholds immer ansteigend (Wu et al. 2016, S. 172). Für die Interpretation werden die Mittelwerte der Thurstonian Thresholds zwischen benachbarten Antwortkategorien herangezogen (Kennedy und Wilson 2007, S. 19). Dabei gilt zu beachten, dass es für die unterste Antwortkategorie keinen mittleren Thurstonian Threshold gibt (Kennedy und Wilson 2007, S. 19). Aus diesem Grund werden im vorliegenden Partial Credit Modell zur Beurteilung der Schwierigkeit die Itemparameter anstelle der Thurstonian Thresholds verwendet. Dies ist möglich, da beide Varianten als Auswertungskriterium für die Schwierigkeit betrachtet werden können.

Um weitere Modellfit-Werte erhalten zu können, wird die Funktion *tam.modelfit* des Pakets *TAM* verwendet (Robitzsch et al. 2020a). Mit der *summary* Funktion (R Core Team 2019) werden unter anderem der SRMR und die zusammengefasste Q3-Statistik ausgegeben.

Als ein Kriterium für eine gute Modellpassung gilt ein SRMR-Wert (Standardized-Root-Mean-Residual) $SRMR < 0.08$ (Hu und Bentler 1999, S. 1). „Der *SRMR* kennzeichnet die

standardisierte durchschnittliche Abweichung (Residuum) zwischen der beobachteten und der implizierten Korrelationsmatrix“ (Bühner 2011, S. 427). Alle Werte des SRMR > 0.11 sind inakzeptabel (Bühner 2011, S. 427).

Eine der Grundannahmen von IRT-Modellen ist die lokale stochastische Unabhängigkeit der Items (Christensen et al. 2017, S. 178; Lamprianou 2020, S. 152; Wu et al. 2016, S. 124). „Liegt lokale stochastische Unabhängigkeit vor, so hängt die Lösungswahrscheinlichkeit einer Aufgabe (oder die Wahrscheinlichkeit, eine Aufgabe im Sinne des Konstrukts zu beantworten) ausschließlich von der Ausprägung des latenten Merkmals ab und wird nicht durch das Lösen oder Nichtlösen (bzw. die Beantwortung) einer vorangegangenen Aufgabe beeinflusst“ (Moosbrugger et al. 2020, S. 261). Mit der Q3 Statistik kann die lokale stochastische Unabhängigkeit überprüft werden (Christensen et al. 2017, S. 178). Dabei sollten alle Q3-Werte (auch der adjustierte Q3-Wert) nahe Null liegen, um von einem guten Modellfit sprechen zu können (Lamprianou 2020, S. 152–153). Die MA-DaQ3-Statistik beschreibt den Mittelwert der aQ3-Werte (Lamprianou 2020, S. 153). Die aQ3-Statistik ist eine adjustierte Version, die den Mittelwert aller Q3-Statistiken von Q3 subtrahiert (Lamprianou 2020, S. 152). Das Problem bei der Verwendung von Q3-Richtwerten ist, dass diese, obwohl sie ein häufig angegebenes Qualitätskriterium für den Modellfit darstellen, scheinbar willkürlich ausgewählt werden (Christensen et al. 2017, S. 178). Mitunter schwanken die Cut-off-Werte in Studien stark: $0.1 < Q3 < 0.7$ (Christensen et al. 2017, S. 178). Christensen et al. (2017, S. 192) kommen anhand ihrer Analyse zu dem Schluss, dass bei Werten des $Q3 > 0.3$ (positiv und negativ) von einer lokalen stochastischen Abhängigkeit auszugehen ist.

Zusätzlich können der SGDDM (Standardized Generalized Dimensionality Discrepancy Measure) und WSGDDM (gewichtete Variante des SGDDM) betrachtet werden. Mit diesen Werten können Aussagen über die mittlere absolute, bedingte Korrelation für Item-Paare getroffen werden (Lamprianou 2020, S. 154). Auch hier sollten beide Werte möglichst nahe 0 liegen (Lamprianou 2020, S. 154).

Um zu überprüfen, inwieweit die Daten mit den Modellannahmen übereinstimmen, wird die Chi-Quadrat-Fit-Statistik in Form der Analyse der Infit- und Outfit-Werte verwendet (Bond und Fox 2012, S. 238). Der Outfit, auch bezeichnet als Unweighted Mean Square, beurteilt die Passung eines Items, indem die standardisierten Residuen des Items quadriert und über die Stichprobe N ermittelt werden (Wright und Masters 1982, S. 99):

$$Outfit_i = \frac{\sum_{v=1}^N z_{vi}^2}{N}$$

N = Stichprobengröße z = Residuum

i = Item

v = Person

(Wright und Masters 1982, S. 99)

Dabei ist zu beachten, dass die Outfit-Werte aufgrund fehlender Gewichtung stärker auf Ausreißer reagieren (Bond und Fox 2012, S. 57; Wright und Masters 1982, S. 99). Eine alternative Berechnung bietet die Infit-Statistik. Dabei werden die quadrierten Residuen so gewichtet, dass die Antworten der Personen mit einer Personenfähigkeit nahe der Itemschwierigkeit mehr Einfluss auf den Infit-Wert haben als Personen, für die das Item ungeeignet scheint (Bond und Fox 2012, S. 57; Wright und Masters 1982, S. 99). Der Infit als Weighted Mean Square lässt sich mittels folgender Formel berechnen:

$$\text{Infit}_i = \frac{\sum_{v=1}^N y_{vi}^2}{\sum_{v=1}^N \text{Var}_{vi}}$$

Var = Varianz

$$y_{vi}^2 = z_{vi}^2 \text{Var}_{vi} = \text{Residuen} \quad (\text{Wright und Masters 1982, S. 99})$$

Diese Unterscheidung macht deutlich, warum in der Literatur den Infit-Werten bei der Analyse der Modellpassung mehr Bedeutung als den Outfit-Werten gegeben wird, wobei die Outfit-Werte nicht unberücksichtigt bleiben sollen (Bond und Fox 2012, S. 57). Die Interpretation der Infit- und Outfit-Werte sollte in Abhängigkeit von der Stichprobengröße erfolgen (Wu et al. 2016, S. 148; Smith et al. 1998, S. 71–75). Wenn die Stichprobe eher klein ist, kann von einer größeren Abweichung der Infit- und Outfit-Mean-Squares ausgegangen werden, z.B. sollten bei einer Stichprobe von $N = 200$ die $\text{Infit}_i / \text{Outfit}_i$ -Werte zwischen 0.8 und 1.2 liegen (Wu et al. 2016, S. 148). Bond und Fox (2012, S. 310) geben als allgemeine Richtlinie für den $\text{Infit}_i / \text{Outfit}_i$ eine Spannweite von 0.7 bis 1.3 an. Smith et al. (1998, S. 78) unterscheiden dagegen spezifisch nach Infit- und Outfit-Werten und in Abhängigkeit der Stichprobengröße. Bei einer Stichprobengröße von $N = 150$ sollten folgende Werte nicht überschritten werden: $\text{Infit}_{i\max} = 1.16$ bzw. $\text{Outfit}_{i\max} = 1.48$ (Smith et al. 1998, S. 78). Bei der Beurteilung der Items des vorliegenden Modells wird sich dem angeschlossen und zwischen Infit- und Outfit-Werten differenziert: Für den Outfit werden Itemkennwerte von $0.7 < \text{Outfit}_i < 1.3$, für die Infit-Werte, die kritischer betrachtet werden sollen, von $0.8 < \text{Infit}_i < 1.2$ akzeptiert. Zudem empfiehlt es sich, bei der Selektion der Items eher Items mit Werten leicht unter 1 zu verwenden, da diese zwischen Personenparametern besser trennen (Wu et al. 2016, S. 152). Die Mittelwerte über alle Infit- bzw. Outfit-Werte sollten nahe 1 liegen und sind im Gegensatz zu den einzelnen Itemwerten auch bei verschiedenen Stichprobengrößen und Testlängen stabil (Smith et al. 1998, S. 71).

Bisher liegt der Fokus der methodischen Auswertung auf die Analyse der Itemparameter. Neben diesen Parametern gilt es auch, die Personenparameter zu überprüfen (Rost 2004, S. 309). Dafür können verschiedene Schätzmethode zum Einsatz kommen (Rost 2004, S. 309). Zur Berechnung der Personenfähigkeit in der vorliegenden Raschskalierung wird der WLE-Schätzer verwendet. Dieser Schätzer wird unter anderem von der Arbeitsgruppe Wu et al. (2016, S. 123) und von Rost (2004, S. 315) präferiert. „Die nach der WLE-

Methode geschätzten Personenparameter sind die besten *Punktschätzer* der individuellen Messwerte“ (Rost 2004, S. 315). Da im vorliegenden Modell von einer 3-dimensionalen Struktur ausgegangen wird, erfolgt die Berechnung der Personenparameter pro Dimension. Beurteilt wird hierbei die punktbiseriale Korrelation (r_{pbis}), die angibt wie hoch die Korrelation des Itemscores mit dem Gesamtscore ist (Wu et al. 2016, S. 84). Die punktbiseriale Korrelation wird auch als „Item-Diskriminierung“ (Wu et al. 2016, S. 84) bzw. Trennschärfe bezeichnet. Damit kann festgestellt werden wie gut ein Item zwischen Personen mit verschiedenen Fähigkeiten unterscheiden kann (Rost 2004, S. 98). Bei einem Partial-Credit-Modell sollen die r_{pbis} -Werte entsprechend der Kategorien ansteigend sein (Wu et al. 2016, S. 85). Bei der Kategorie 0 wird aufgrund der Vorgehensweise der Berechnung ein negativer Wert erwartet (Wu et al. 2016, S. 86).

Modelloptimierung

Nach der Auswertung der Daten zur Raschskalierung schließt sich eine Modelloptimierung an. Dabei wird schrittweise vorgegangen und Items entsprechend der beschriebenen Kennwert- sowie Inhaltsanalyse selektiert. Inhaltsanalyse bedeutet in diesem Fall, dass die Items hinsichtlich ihrer Sinn- und Zweckhaftigkeit überprüft werden (beispielsweise wird der Frage nachgegangen, ob das Item für das Abbilden des Motivationsaspektes relevant ist). Während des Prozesses der Itemselektion werden die statistischen Kennwerte der Items nicht losgelöst von der inhaltlichen Interpretation beurteilt. Eine Testkonstruktion steht oft im Spannungsfeld zwischen guten Messeigenschaften und inhaltlicher Validität (Wu et al. 2016, S. 87). Der Prozess der Modellanpassung durch Itemselektion gestaltet sich als konvergierend und iterativ. Bei jeder vorgenommenen Anpassung (bereits, wenn ein Item entfernt wird), verändern sich einzelne Kennwerte. Dementsprechend werden nach jeder Selektion von Items die Daten neu ausgewertet. Dabei wird der bereits beschriebene Prozess der Raschskalierung beibehalten. Aufgrund der Komplexität der Datenanalyse werden die einzelnen Schritte nicht explizit aufgelistet, sondern ausschließlich des ersten Modells mit der Bezeichnung „model 1“ und des optimierten Modells „finalmod“ mit den wesentlichen Kenndaten in der vorliegenden Arbeit dokumentiert.

Modellvergleich mit zweiter Stichprobe

Um zu überprüfen, ob die Konstruktvalidität auch bei einer zweiten Stichprobe gewährleistet ist, wird die Raschskalierung an einer weiteren Stichprobe durchgeführt. Dafür wird die Stichprobe N_2 verwendet, die zu Beginn der Raschskalierung als Zufallsstichprobe von der Gesamtstichprobe gezogen wird (siehe Abschnitt 10.1 *Stichprobe*). Damit stehen zwei Stichproben zur Aufstellung und Überprüfung des Modells zur Verfügung. Zur Beurteilung der Passung des skalierten und optimierten Modells (an der Stichprobe N_1) für die zweite Stichprobe werden zum einen die Modell-Fit-Werte, zum anderen der

Modellvergleich mit der Stichprobe N_1 herangezogen. Diese Vorgehensweise wird von Rost (2004, S. 74) als „sparsamere Form der Geltungsprüfung“ empfohlen. Andernfalls werden sehr hohe Stichproben benötigt, um möglichst jedes Antwortmuster (bei 10 dichotomen Items gibt es bereits 1024 unterschiedliche Antwortmuster) sicher (das 5fache der Antwortmuster) abdecken zu können (Rost 2004, S. 74).

10.3 Ergebnisse

Raschskalierung

Die Auswertung der Raschskalierung erfolgt entsprechend der methodischen Beschreibung (10.2.2 *Raschskalierung*). Anhand der ersten deskriptiven Analysen zeigen sich vier Items hinsichtlich der Mittelwerte, Varianz und Besetzung der Kategorien auffallend (siehe Tabelle 24) und werden für die weiteren Analysen ausgeschlossen.

Tabelle 24 Deskriptive Statistik (N = Stichprobengröße, M = Mittelwerte, SD = Standardabweichung, Var = Varianz, N_{Kat} = Anzahl vorkommender Antwortkategorien)

Item	N	M	SD	Var	N_{Kat}	Range	Skew	Kurtosis
V2E11	139	0.32	0.47	0.22	2	1	0.78	-1.40
V3BS1	140	0.08	0.30	0.09	3	2	15.86	0.02
V7BS1	140	0.11	0.47	0.22	2	2	12.34	0.04
V7BS5	140	1.92	0.34	0.12	3	2	-4.54	20.59

Das Item V2E11 wird der Kategorie des Erklärens zugeordnet: „Angenommen die Lehrperson würde einen Test mit differenzierten Schwierigkeitsgraden erstellen, dann geschieht dies, um die Zielerreichung aller Schülerinnen und Schüler in der Unterrichtsreihe vergleichend überprüfen zu können.“ Keiner der Studierenden erreicht die volle Punktzahl (2 Punkte). Diese erhalten die Studierenden, wenn sie „stimmt gar nicht“ ankreuzen. Anhand des Mittelwerts wird deutlich, dass dieses Item als eher schwer eingestuft werden kann ($M = 0.32$). Die Besetzung der Kategorie 2 wird zumindest von Studierenden mit einer hohen Fähigkeit erwartet. Da es sich hierbei um ein strukturelles Problem handelt, wird dieses Item aus den weiteren Analysen ausgeschlossen. Es liegt kein dringender Grund vor, dieses Item trotz geringer Range zuzulassen.

Beim Item V7BS1 „Das Vorwissen der Schülerinnen und Schüler wird aktiviert“ liegt entsprechend der Anzahl vorkommender Antwortkategorien ($N_{Kat} = 2$) eine Null-Kategorie vor. Es stellt sich heraus, dass keine Person die Kategorie 1 besetzt. In diesem Fall wird die Raschskalierung in R beeinflusst. Da ähnlich wie beim Item V2E11 kein dringender Anlass besteht, dass Item aus inhaltlichen Gründen zwingend weiterzuverwenden, wird es nicht für die Skalierung verwendet.

Das Item V3BS1 weist eine geringe Varianz ($Var = 0.09$) und einen Bodeneffekt ($M = 0.08$) auf. Es entstammt der Vignette 3, die eine Besonderheit in der Konzeption aufweist (siehe Abschnitt 7.5 *Beschreibung der Vignetten*). Sie enthält ausschließlich Items der Kategorie des Beschreibens. In weiteren Analysen zeigt sich, dass die meisten Items der Vignette 3 nicht raschkonform sind. Daher kann an dieser Stelle bereits berichtet werden, dass die Vignette 3 vollständig aus den weiteren Analysen ausgeschlossen wird.

Das Item V7BS5 „Fehler werden als Teil des Lernprozesses gesehen“ gehört ebenfalls zur Kategorie des Beschreibens. Es ist ein Item, das sehr leicht von den Studierenden beantwortet werden kann ($M = 1.92$). Die Lehrperson betont mehrfach im Video, dass Fehler dazu gehören und man daraus nur lernen kann. Die Studierenden können diesen Aspekt so gut beschreiben, dass fast alle Studierenden (95%) die Höchstpunktzahl für die Beantwortung dieses Items erreichen. Daher wird dieses Item aus den weiteren Analysen ausgeschlossen.

Zu Beginn der Raschskalierung wird das theoretisch dreidimensionale Modell aufgestellt und mit weiteren Modellen unterschiedlicher Dimensionsannahmen verglichen. Die Analyse hinsichtlich der Überprüfung der am besten passenden dimensional Modellstruktur ergibt Folgendes: Der Vergleich der fünf aufgestellten Modelle weist anhand der Informationskriterien daraufhin, dass das theoretisch angenommene Dreifaktoren-Modell der Professionellen Unterrichtswahrnehmung mit den Faktoren Beschreiben, Erklären und Vorhersagen für die weiteren Analysen am geeignetsten ist. Es weist die geringsten AIC- und BIC-Werte auf. Alle Differenzen (ΔAIC , ΔBIC) sprechen für einen Vorteil der Modellannahme der Dreidimensionalität. Die Ergebnisse der einzelnen Berechnungen sind in der folgenden Tabelle dargestellt.

Tabelle 25 Modellvergleiche mittels Anova (B = Beschreiben, E = Erklären, V = Vorhersagen, / = zusammengefasste Dimensionen, Dim = Dimensionalität, AIC = Akaike Information Criterion, Δ = Differenz, BIC = Bayesian Information Criterion)

Model	Dim	AIC	ΔAIC	BIC	ΔBIC
mod_B+E/V	2-dim	33290.94		34258.74	
mod_B+E+V	3-dim	30476.58	- 2.814.36	31453.21	- 2.805.53
mod_B/E/V	1-dim	38369.81		39331.73	
mod_B+E+V	3-dim	30476.58	- 7.893.23	31453.21	- 7.878.52
mod_B/E+V	2-dim	33293.90		34261.70	
mod_B+E+V	3-dim	30476.58	- 2.817.32	31453.21	- 2.808.49
mod_B/V+E	2-dim	33294.05		34261.85	
mod_B+E+V	3-dim	30476.58	- 2.817.47	31453.21	- 2.808.64

Für die weitere Modellanpassung wird demgemäß das Modell „mod_B+E+V“ zugrunde gelegt. Wie bereits erwähnt, gilt zu berücksichtigen, dass der Modellvergleich lediglich Auskunft über die beste Passung im Vergleich zu anderen Modellen gibt, aber nicht über die Qualität des Modells (Bühner 2006, S. 352). Das Modell „mod_B+E+V“ wird entsprechend der deskriptiven statistischen Auswertung angepasst und als „model1“ weiter spezifiziert.

Auswertung der Raschskalierung für das „model1“

Dem vorliegenden Modell liegen drei Antwortkategorien vor. Diese entsprechen den möglichen Punkten für die Beantwortung eines Items. Die teilnehmenden Personen können 0 Punkte, 1 Punkt oder 2 Punkte erreichen (siehe Abschnitt 8.2 *Validierung der Analyseaufträge*). Damit liegen im vorliegenden Partial-Credit-Modell zwei Schwellen vor. Ein fehlender Anstieg der Schwellenparameter kann durch eine ungleichmäßige Besetzung der Kategorien bedingt sein (Wu et al. 2016, S. 164; Adams et al. 2012, S. 561). Die Parameterschätzungen sind völlig unabhängig von den Fähigkeiten der Studierenden und hängen nur von der Anzahl der Beobachtungen in jeder Antwortkategorie ab (Adams et al. 2012, S. 561). Daher werden alle ungeordneten Items entsprechend der Häufigkeit der einzelnen Antwortkategorien näher betrachtet. Alle Items, die keinen Anstieg der Schwellenparameter ($\delta_1 > \delta_2$) aufzeigen, aber eine Unterbesetzung einer Kategorie ($f < 10\%$) aufweisen, werden nicht ausgeschlossen. In diesem Fall kann bei einer Unordnung der Schwellenparameter nicht von einer Verletzung der Annahmen des Partial-Credit-Modells ausgegangen werden (Wu et al. 2016, S. 165). Diese Items können dennoch gute Fit-Kennwerte aufweisen (Adams et al. 2012, S. 561). Die Überprüfung erfolgt im weiteren Verlauf und kann gegebenenfalls doch zum Ausschluss des Items führen. Die Analyse für das vorliegende Modell bringt folgende Items hervor, die als kritisch zu betrachten sind:

Tabelle 26 Schwellenparameter (S = Schwelle, δ = Schwellenparameter, SE = Standardfehler)

Item	S	δ	SE	Item	S	δ	SE	Item	S	δ	SE
V2BS3	1	0.10	0.18	V1E12	1	1.33	0.18	V1V12	1	0.52	0.18
	2	-0.12	0.18		2	-1.60	0.17		2	-0.54	0.18
V2BS5	1	0.02	0.18	V1E23	1	1.43	0.17	V1V13	1	1.25	0.17
	2	-0.19	0.18		2	-1.20	0.18		2	-1.12	0.17
V2BS6	1	0.66	0.17	V1E34	1	-0.53	0.23	V1V22	1	0.54	0.18
	2	0.00	0.20		2	-0.63	0.17		2	-1.00	0.17
V2BS7	1	0.28	0.19	V2E22	1	0.12	0.18	V1V33	1	0.54	0.20
	2	-0.77	0.17		2	-0.30	0.17		2	-1.58	0.18

V3BS4	1	-0.18	0.22	V2E23	1	1.22	0.17	V1V41	1	0.92	0.18
	2	-1.04	0.17		2	-0.78	0.18		2	-1.19	0.17
V6BS2	1	-0.41	0.24	V2E24	1	-0.20	0.20	V5V22	1	0.93	0.18
	2	-1.09	0.18		2	-0.54	0.17		2	-1.54	0.17
V6BS3	1	0.61	0.17	V6E32	1	-0.36	0.21	V6V12	1	-0.60	0.23
	2	-0.38	0.18		2	-0.44	0.17		2	-0.67	0.17
V6BS5	1	-0.26	0.20	V6E34	1	1.00	0.17	V6V31	1	-0.32	0.22
	2	-0.29	0.17		2	-0.75	0.18		2	-0.92	0.17
V6BS7	1	0.78	0.17	V6E42	1	0.50	0.19	V6V33	1	-0.56	0.27
	2	0.75	0.25		2	-1.23	0.17		2	-1.23	0.18
V6BS9	1	0.00	0.20								
	2	-0.68	0.17								

Der SRMR als Kriterium für die Passgenauigkeit des Modells liegt im „model1“ mit SRMR = 0.088 unter dem kritischen Wert von 0.11. Dies spricht für eine akzeptable Modellpassung.

Die Beurteilung der lokalen stochastischen Unabhängigkeit in Form der MADaQ3-Statistik beschreibt den Mittelwert der aQ3-Werte (Lamprianou 2020, S. 153) und ist mit $M_{aQ3} = 0.088$ leicht über dem angestrebten 0-Wert. Anhand der Zusammenfassung der Q3-Statistik zum vorliegenden Modell (Tabelle 27) wird deutlich, dass die minimalen und maximalen Werte ($Q3_{\min} = -0.44$ und $Q3_{\max} = 0.54$) zu stark von 0 abweichen und einer genaueren Analyse bedürfen. Der SGDDM = 0.086 und WSGDDM = 0.089 (Abkürzung siehe Tabelle 27) sind erwartungsgemäß gering.

Tabelle 27 Zusammenfassung der Q3-Statistik (M = Mittelwert, SD = Standardabweichung, min/ max = minimaler und maximaler Wert, SGDDM = Standardized Generalized Dimensionality Discrepancy Measure, WSGDDM = gewichtete Variante des SGDDM)

type	M	SD	min	max	SGDDM	WSGDDM
Q3	-0.0178	0.1122	-0.4434	0.5415	0.0885	0.0886
aQ3	0.0000	0.1122	-0.4612	0.5237	0.0876	0.0877

Dabei wird festgestellt, dass einige Items den zuvor definierten Richtwert von 0.3 (positiv und negativ) nicht erfüllen. In einem ersten Schritt werden die besonders kritischen Items mit einem Wert $-0.40 > Q3 > 0.40$ ausgewählt, inhaltlich überprüft und entsprechend aus weiteren Analysen ausgeschlossen. Anhand der Tabelle 28 zeigt sich, dass die Items V4E11, V4E13, V5E23, V6E41 und V4V13 mehrfach mit anderen Items eine stochastische Abhängigkeit eingehen.

Tabelle 28 Q3 Statistik „model1“

Item 1	Q3	Item 2	Item	Q3	Item 2	Item	Q3	Item 2
V1BS2	0.44	V1BS3	V4E13	0.44	V4E21	V5E23	0.45	V4V14
V5BS6	0.42	V5BS7	V4E13	0.46	V5E23	V5E23	0.54	V4V22
V6BS2	0.42	V6BS5	V4E13	0.44	V6E41	V5E23	0.51	V7V13
V7BS2	0.42	V7BS3	V4E13	0.46	V7E11	V6E24	0.43	V6V24
V1E11	0.40	V1E42	V4E13	-0.44	V7E12	V6E41	0.40	V7E11
V1E31	0.49	V1E33	V4E13	0.46	V4V13	V6E41	-0.42	V7E12
V1E44	0.40	V5E13	V4E13	0.48	V4V14	V6E41	0.40	V6V24
V1E44	0.40	V5E14	V4E13	0.49	V4V22	V6E41	0.47	V7V13
V4E11	0.44	V4E21	V4E13	0.49	V7V13	V1V23	0.49	V1V24
V4E11	0.46	V5E23	V4E21	-0.40	V5E11	V2V12	0.53	V2V13
V4E11	0.41	V6E41	V4E21	0.41	V4V13	V2V14	0.48	V2V23
V4E11	0.46	V7E11	V4E21	0.41	V4V14	V4V13	0.48	V4V14
V4E11	-0.44	V7E12	V4E21	0.45	V4V22	V4V13	0.42	V4V22
V4E11	0.46	V4V13	V5E23	0.42	V6E24	V4V13	0.42	V7V13
V4E11	0.48	V4V14	V5E23	0.47	V6E41	V4V14	0.49	V4V22
V4E11	0.49	V4V22	V5E23	0.43	V7E11	V4V14	0.48	V7V13
V4E11	0.49	V7V13						

Die Analyse der Infit-Statistik zeigt, dass alle Infit-Werte innerhalb der Cut-Off-Grenzwerte, demnach zwischen 0.8 und 1.2 liegen. Da die Infit-Werte als kritischer betrachtet werden (Bond und Fox 2012, S. 57), ist dies ein positives Ergebnis für die Modellpassung. Die Outfit-Werte einiger Items (siehe Tabelle 29) sollten in Verbindung mit den weiteren kritischen Kennwert-Ergebnissen bei der Modellanpassung durch Itemselektion berücksichtigt werden. Dabei ist zu bedenken, dass die Überschreitung der annehmbaren Outfit-Werte ($0.7 < \text{Outfit} < 1.3$) durch erwartungswidrige Antworten von Personen aufgrund ungeeigneter Items auftreten können (Wright und Masters 1982, S. 99). Die Mittelwerte des Infits ($M = 1.00$) und Outfits ($M = 1.00$) sind gut.

Tabelle 29 kritische Outfitwerte der Items im „model 1“ (S=Schwelle)

Item	S	Outfit	Item	S	Outfit	Item	S	Outfit
V2BS1	1	1.51	V7BS7	1	1.39	V6E41	1	0.41
V5BS1	1	1.39	V7BS7	2	1.40	V4V11	2	1.36
V4BS4	1	1.52	V4E13	1	0.41			

Bei der Interpretation der Kennwerte der WLE-Reliabilitäten (R_{WLE}) gilt zu beachten, dass sie im Vergleich zu den MLE-Schätzern geringer ausfallen (Rost 2004, S. 314). Die WLE-Reliabilitäten für die einzelnen Dimensionen befinden sich dennoch im niedrigen Bereich: Beschreiben $R_{WLE} = 0.551$, Erklären $R_{WLE} = 0.541$, Vorhersagen $R_{WLE} = 0.627$.

Die Trennschärfe der Items in Bezug auf die Personenparameter stellt sich bei einigen Items als ungeordnet dar (Tabelle 30-32). Bei diesen Items sollte überprüft werden, ob die Schwierigkeit (zu leicht oder zu schwer) Ursache für die niedrige Trennschärfe ist (Wu et al. 2016, S. 87).

Tabelle 30 Trennschärfe (r_{pbis}) der Items der Dimension des Beschreibens mit Angabe der Itemschwierigkeit (σ , zu leicht $\sigma < -1$ oder zu schwer $\sigma > 1$ ist fett markiert) und der Häufigkeit (F) der Antworten pro Kategorie (K)

Item σ	K	F	r_{pbis}	Item σ	K	F	r_{pbis}	Item σ	K	F	r_{pbis}
V1BS1	0	0.06	-0.05	V3BS2	0	0.18	0.07	V5BS8	0	0.93	-0.03
-0.91	1	0.58	-0.15	-0.15	1	0.58	-0.08	1.38	1	0.01	0.07
	2	0.36	0.18		2	0.24	0.04		2	0.06	0.00
V1BS4	0	0.07	-0.07	V3BS4	0	0.18	-0.20	V6BS1	0	0.24	0.14
-1.21	1	0.18	-0.15	-0.61	1	0.21	-0.23	-0.07	1	0.47	-0.07
	2	0.75	0.17		2	0.61	0.35		2	0.28	-0.05
V1BS6	0	0.01	0.01	V4BS1	0	0.04	-0.08	V6BS4	0	0.09	-0.14
-2.04	1	0.12	-0.14	-0.90	1	0.70	0.07	-1.04	1	0.16	-0.24
	2	0.87	0.14		2	0.26	-0.01		2	0.75	0.29
V2BS1	0	0.06	-0.07	V4BS2	0	0.04	-0.01	V6BS5	0	0.25	-0.10
-1.25	1	0.15	-0.32	-1.06	1	0.66	0.04	-0.28	1	0.32	-0.33
	2	0.78	0.32		2	0.30	-0.04		2	0.43	0.40
V2BS6	0	0.49	-0.24	V4BS5	0	0.21	-0.08	V6BS8	0	0.19	-0.16
0.33	1	0.25	0.24	-0.39	1	0.34	-0.19	-0.31	1	0.47	0.09
	2	0.25	0.04		2	0.45	0.24		2	0.35	0.04
V2BS7	0	0.29	-0.04	V5BS1	0	0.09	-0.04	V6BS9	0	0.25	-0.03
-0.25	1	0.22	0.07	-1.12	1	0.09	-0.27	-0.39	1	0.25	-0.18
	2	0.48	-0.02		2	0.82	0.23		2	0.50	0.18
				V5BS4	0	0.18	-0.32	V7BS2	0	0.09	-0.04
				-0.15	1	0.58	0.23	-0.98	1	0.29	0.07
					2	0.24	0.03		2	0.62	-0.02

In der Kategorie des Beschreibens zeigen sieben Items eine hohe oder niedrige Schwierigkeit (σ , in der Tabelle 30 fett markiert) auf. Die übrigen 13 Items weisen sich als kritisch hinsichtlich der Trennschärfe (r_{pbis}) aus und müssen weiter überprüft und gegebenenfalls ausgeschlossen werden.

Die Trennschärfe (r_{pbis}) in der Kategorie des Erklärens ist bei 31 Items nicht kontinuierlich über alle drei Kategorien ansteigend (σ , in der Tabelle 31 fett markiert). Bei neun

Items wird der fehlende Anstieg mit der Itemschwierigkeit begründet. Es erfolgt ein sukzessiver Ausschluss der 23 kritischen Items unter Berücksichtigung weiterer Itemparameter und inhaltlicher Abklärung.

Tabelle 31 Trennschärfe (r_{pbis}) der Items der Dimension des Erklärens mit Angabe der Itemschwierigkeit (σ , zu leicht $\sigma < -1$ oder zu schwer $\sigma > 1$ ist fett markiert) und der Häufigkeit (F) der Antworten pro Kategorie (K)

Item σ	K	F	r_{pbis}	Item σ	K	F	r_{pbis}	Item σ	K	F	r_{pbis}
V1E11 -1.08	0 1 2	0.02 0.17 0.81	-0.04 -0.21 0.24	V2E24 -0.37	0 1 2	0.23 0.28 0.49	-0.20 0.11 0.07	V6E11 -0.97	0 1 2	0.08 0.38 0.54	-0.07 -0.11 0.15
V1E12 -0.13	0 1 2	0.39 0.10 0.51	-0.20 0.23 0.06	V4E11 -0.83	0 1 2	0.06 0.64 0.30	0.14 -0.01 -0.06	V6E13 -0.56	0 1 2	0.16 0.36 0.48	0.07 -0.14 0.09
V1E14 0.56	0 1 2	0.51 0.32 0.17	0.02 -0.08 0.06	V4E12 2.07	0 1 2	0.90 0.09 0.01	-0.16 0.12 0.11	V6E21 -1.00	0 1 2	0.05 0.58 0.37	-0.02 0.02 -0.01
V1E22 -0.34	0 1 2	0.20 0.41 0.39	-0.13 -0.16 0.27	V4E14 0.71	0 1 2	0.77 0.04 0.19	-0.10 0.06 0.08	V6E31 -1.15	0 1 2	0.01 0.84 0.14	0.00 0.15 -0.15
V1E23 0.12	0 1 2	0.49 0.12 0.39	-0.34 0.36 0.11	V4E23 0.77	0 1 2	0.51 0.38 0.11	-0.38 0.28 0.17	V6E32 -0.40	0 1 2	0.21 0.31 0.48	0.08 0.01 -0.08
V1E31 -0.58	0 1 2	0.04 0.16 0.79	-0.15 -0.18 0.24	V5E11 -0.96	0 1 2	0.06 0.49 0.44	0.04 0.08 -0.10	V6E42 -0.37	0 1 2	0.27 0.16 0.56	-0.33 0.17 0.17
V1E32 0.86	0 1 2	0.45 0.47 0.08	-0.46 0.32 0.26	V5E12 -0.56	0 1 2	0.17 0.33 0.50	-0.01 -0.13 0.13	V6E43 -1.30	0 1 2	0.05 0.27 0.68	-0.20 -0.20 0.29
V1E41 -0.83	0 1 2	0.09 0.45 0.46	0.03 -0.10 0.08	V5E13 -0.39	0 1 2	0.11 0.66 0.24	-0.17 0.15 -0.04	V6E44 1.57	0 1 2	0.83 0.14 0.04	-0.26 0.20 0.17
V2E12 -1.08	0 1 2	0.07 0.31 0.62	-0.15 -0.18 0.25	V5E21 -1.08	0 1 2	0.09 0.17 0.74	-0.06 -0.09 0.12	V7E12 -1.52	0 1 2	0.01 0.84 0.15	0.10 0.05 -0.07
V2E22 -0.09	0 1 2	0.32 0.29 0.39	-0.24 0.22 0.02	V5E22 0.12	0 1 2	0.55 0.02 0.43	-0.31 0.11 0.28	V7E23 0.43	0 1 2	0.27 0.69 0.02	-0.17 0.13 0.13
								V7E24 -0.84	0 1 2	0.29 0.09 0.41	-0.05 0.03 0.00

Tabelle 32 Trennschärfe (r_{pbis}) der Items der Dimension des Vorhersagens mit Angabe der Itemschwierigkeit (σ , zu leicht $\sigma < -1$ oder zu schwer $\sigma > 1$ ist fett markiert) und der Häufigkeit (F) der Antworten pro Kategorie (K)

Item σ	K	F	r_{pbis}	Item σ	K	F	r_{pbis}	Item σ	K	F	r_{pbis}
V1V12	0	0.38	-0.34	V4V11	0	0.87	0.02	V6V13	0	0.09	-0.12
-0.01	1	0.23	0.31	1.85	1	0.11	0.02	-0.74	1	0.50	-0.13
	2	0.39	0.08		2	0.02	-0.09		2	0.41	0.20
V1V21	0	0.11	-0.33	V4V12	0	0.13	-0.07	V6V22	0	0.06	-0.25
-0.82	1	0.32	0.13	-0.69	1	0.35	-0.14	-0.96	1	0.50	0.13
	2	0.57	0.08		2	0.52	0.19		2	0.44	0.00
V1V22	0	0.32	-0.30	V4V21	0	0.14	-0.05	V6V23	0	0.38	-0.33
-0.23	1	0.18	0.18	-0.65	1	0.36	-0.07	0.54	1	0.49	0.20
	2	0.50	0.14		2	0.50	0.10		2	0.13	0.18
V1V24	0	0.66	-0.37	V5V11	0	0.57	0.00	V6V32	0	0.05	-0.14
1.45	1	0.30	0.30	0.21	1	0.05	0.01	-0.82	1	0.69	0.15
	2	0.04	0.20		2	0.38	-0.01		2	0.26	-0.09
V1V41	0	0.37	-0.28	V5V12	0	0.77	-0.16	V6V41	0	0.05	-0.24
-0.14	1	0.15	0.16	1.79	1	0.21	0.16	-1.07	1	0.52	0.16
	2	0.49	0.16		2	0.02	0.02		2	0.43	-0.05
V1V42	0	0.21	-0.18	V5V13	0	0.76	-0.04	V6V42	0	0.56	0.11
-0.27	1	0.42	0.09	0.60	1	0.01	0.05	1.29	1	0.39	-0.10
	2	0.37	0.06		2	0.23	0.03		2	0.04	-0.02
V2V11	0	0.29	-0.02	V5V23	0	0.46	-0.32	V7V12	0	0.10	0.00
-0.08	1	0.36	-0.07	0.61	1	0.40	0.20	-0.74	1	0.46	0.03
	2	0.35	0.09		2	0.14	0.18		2	0.44	-0.03
V2V13	0	0.64	-0.30	V6V11	0	0.12	-0.13	V7V13	0	0.02	-0.14
0.38	1	0.06	0.13	-0.70	1	0.39	0.08	-1.85	1	0.11	-0.26
	2	0.30	0.25		2	0.49	0.01		2	0.86	0.30
				V6V12	0	0.16	-0.46	V7V23	0	0.26	-0.13
				-0.63	1	0.29	0.26	0.11	1	0.42	-0.16
					2	0.56	0.10		2	0.32	0.30

Nach Überprüfung der Itemschwierigkeit (siehe Tabelle 32) bleiben 20 Items in der Kategorie des Vorhersagens hinsichtlich der Trennschärfe kritisch zu betrachten. Diese Items werden weiter analysiert.

Modelloptimierung

Nachdem alle Kennwerte ermittelt und ausgewertet sind, erfolgt die Modelloptimierung durch Itemselektion. Hier wird schrittweise vorgegangen. Zunächst werden alle Items mit mehreren kritischen Kennwerten bestimmt. Durch die Extraktion dieser Items wird ein

verbesserter Modellfit erwartet. Nach erneuter Rasch-Skalierung werden die gewonnenen Daten wie zuvor beschrieben ausgewertet und entsprechend kritische Items entfernt. Dieser Prozess verläuft konvergierend und iterativ. Mit Abschluss dieses mehrstufigen Analyseprozesses werden folgende Items aus dem Modell „model1“ nicht mehr im finalen Modell berücksichtigt:

Tabelle 33 ausgeschlossene Items entsprechend den zuvor festgelegten Kennwerten

Dimension	ausgeschlossene Items mit kritischen Kennwerten
Beschreiben	V1BS1, V1BS4, V1BS6, V2BS1, V2BS5, V2BS6, V2BS7, V3BS1 - V3BS5, V4BS1, V4BS2, V4BS5, V5BS1, V5BS2, V5BS4, V5BS8, V6BS1, V6BS4, V6BS5, V6BS8, V6BS9, V7BS1, V7BS3, V7BS5, V7BS7
Erklären	V1E12, V1E14, V1E31, V1E41, V1E44, V2E11, V2E22, V2E24, V4E11, V4E13, V4E22, V4E24, V5E11 - V5E13, V6E13, V6E21, V6E31, V6E32, V6E41, V6E44, V7E12, V7E23
Vorhersagen	V1V22, V1V33, V2V11, V4V13, V4V14, V4V22, V5V11, V5V22, V6V11, V6V22, V6V32, V6V41, V7V12

An dieser Stelle wird vor der Präsentation der finalen Modellparameter darauf hingewiesen, dass einige Items trotz kritischer Werte beibehalten werden. Im Abschnitt 10.4 *Interpretation und Schlussfolgerung* wird auf diese Items nochmal eingegangen und eine Begründung gegeben, warum sie beibehalten werden.

Die Raschskalierung des finalen Modells ergibt folgende Kennwerte. Die EAP-Reliabilitäten sind in allen drei Dimensionen akzeptabel und erfüllen bei einem Leistungstest vollkommen die Anforderungen: Beschreiben $R_{EAP} = 0.756$, Erklären $R_{EAP} = 0.735$, Vorhersagen $R_{EAP} = 0.718$

Die Korrelationen zwischen den Dimensionen lassen sich folgendermaßen beschreiben: Zwischen Beschreiben und Erklären liegt eine hohe Korrelation vor ($r = 0.61$), zwischen Erklären und Vorhersagen eine mittlere Korrelation ($r = 0.34$) und zwischen Beschreiben und Vorhersagen eine niedrige Korrelation ($r = 0.18$). Der SRMR = 0.086 ist deutlich geringer als der Grenzwert (SRMR > 0.11) und stellt damit einen akzeptablen Modellfit dar. Die Mittelwerte der Infit- ($M = 1.001$) und Outfitwerte ($M = 1.008$) liegen nahe 1 und zeigen eine gute Passung. Es kann angenommen werden, dass die theoretischen und empirischen Daten übereinstimmen. Die minimalen (Outfit_{min} = 0.62) und maximalen (Outfit_{max} = 1.52) Outfit-Werte befinden sich bei acht Items außerhalb des zuvor definierten Grenzbereiches ($0.7 < \text{Outfit} < 1.3$) (siehe Tabelle 34).

Tabelle 34 kritische Outfitwerte der Items im finalen Modell (S=Schwelle)

Item	S	Outfit	Item	S	Outfit	Item	S	Outfit
V1BS5	1	1.47	V7BS4	2	1.47	V6E11	1	1.45
V2BS4	1	0.66	V1E11	1	1.33	V6E24	1	0.62
V4BS4	1	1.52	V5E23	1	0.65	V2V12	2	0.62
V7BS2	1	1.37	V5E24	2	1.38	V6V42	2	1.44

Bei der vorliegenden Stichprobengröße können laut Smith et al. (1998, S. 78) Outfit-Werte zwischen $0.52 < \text{Outfit} < 1.48$ angenommen werden. Demnach bleibt lediglich ein Item V4BS4 (Outfit = 1.52) kritisch zu betrachten, alle weiteren Items liegen innerhalb des Toleranzbereiches. Die Infitwerte sind bei allen Items zwischen $0.83 < \text{Infit} < 1.19$ und damit außerhalb des kritischen Bereiches. Die Q3-Werte ($M = -0.002$) sind ebenfalls nahe Null, aQ3 ($M = 0.000$) ist optimal. Der maximale Q3-Wert liegt bei $Q3_{\max} = 0.490$. Bei vier Itempaaren befindet er sich über dem definierten Akzeptanzbereich:

Tabelle 35 kritische Q3-Werte im finalen Modell

Item 1	Q3	Item 2	Item	Q3	Item 2
V1V23	0.43	V1V24	V2V12	0.46	V2V13
V2V14	0.42	V2V23	V7V13	0.49	V5E23

Im vorliegenden Modell sind die Schwellenparameter von der ersten zur zweiten Schwelle bei allen Items ansteigend. Die mittlere Itemschwierigkeit liegt bei $M_{\sigma} = -0.21$ Logit. Dieser Wert deutet darauf hin, dass der Test weder zu schwer ($\sigma > 1.00$), noch zu leicht ($\sigma < -1.00$) ist. Die WLE Reliabilitäten der Dimensionen befinden sich wie erwartet leicht unterhalb der EAP-Reliabilitäten und gelten als akzeptabel (Beschreiben $R_{WLE} = 0.720$, Erklären $R_{WLE} = 0.656$, Vorhersagen $R_{WLE} = 0.681$). Hinsichtlich der Personenfähigkeit zeigt sich bei einigen Items, dass diese nicht ansteigend ist (siehe Tabelle 36).

Tabelle 36 Trennschärfe (r_{pbis}) pro Kategorie (K) im finalen Modell

Item	r_{pbis} (Kat 1, 2, 3)	Item	r_{pbis} (Kat 1, 2, 3)
V7BS2	-0.14, -0.16, 0.23	V7E13	-0.37, 0.26, 0.23
V1E22	-0.10, -0.15, 0.23	V1V11	-0.30, 0.09, 0.06
V1E23	-0.32, 0.33, 0.11	V1V12	-0.36, 0.34, 0.06
V2E24	-0.24, 0.14, 0.08	V4V12	-0.09, -0.17, 0.22
V4E23	-0.37, 0.26, 0.19	V4V21	0.02, -0.13, 0.11
V6E11	-0.05, -0.16, 0.18	V6V12	-0.45, 0.30, 0.06
V6E34	-0.33, 0.21, 0.17	V6V23	-0.40, 0.24, 0.22
V6E42	-0.38, 0.21, 0.18	V7V23	-0.14, -0.18, 0.32

Mit dem abschließenden Modellvergleich zwischen dem ersten aufgestellten „model1“ und dem finalen Modell soll gezeigt werden, ob nach der Itemanalyse und dem Entfernen kritischer Items das finale Modell eine signifikant bessere Passung ergibt. Die Ergebnisse mit $p = 1.000$ und $\Delta AIC = -5238.28$, $\Delta BIC = -5579.51$ lassen dies annehmen (siehe Tabelle 37).

Tabelle 37 Anova zwischen dem "model1" und dem finalen Modell „finalmod“ (B = Beschreiben, E = Erklären, V = Vorhersagen, Dim = Dimensionalität, AIC = Akaike Information Criterion, Δ = Differenz, BIC = Bayesian Information Criterion)

Model	Dim	AIC	ΔAIC	BIC	ΔBIC
finalmod	3-dim: B+E+V	25170.01		25799.52	
model1	3-dim: B+E+V	30408.29	- 5238.28	31379.03	- 5579.51

Damit kann abschließend eine Übersicht über die Anzahl der Items pro Vignette sowie im Gesamten pro Dimension für das finale Modell erstellt werden (siehe Abbildung 33). Alle relevanten Itemkennwerte der 104 Items sind im Anhang H abgebildet. Auffallend ist, dass die Itemanzahl der Vignetten 1 und 6 im Erklären und Vorhersagen im Vergleich zu den anderen Vignetten deutlich höher ausfällt. Dies ist damit zu erklären, dass es bei beiden Vignetten insgesamt vier Schwerpunkte der Analyseaufträge gibt (siehe Abschnitt 7.5 *Beschreibung der Vignetten*), bei den anderen Vignetten maximal zwei Schwerpunkte.

Vignette	Itemanzahl		
	Beschreiben	Erklären	Vorhersagen
1	3	11	13
2	3	6	6
4	3	5	4
5	4	5	5
6	4	10	9
7	3	5	5
gesamt	20	42	42

Abbildung 33 Übersicht der Items des finalen Modells

Modellvergleich mit zweiter Stichprobe

Die Raschskalierung mit der Stichprobe N_2 weist minimal niedrigere Modellfitwerte auf. Die EAP-Reliabilität sinkt für die Dimension 1 um 0.035, Dimension 2 um 0.041 und Dimension 3 um 0.004. Damit kann die EAP-Reliabilität in den Dimensionen des Beschreibens ($R_{EAP} = 0.721$) und Vorhersagens ($R_{EAP} = 0.714$) mit $R_{EAP} > 0.700$ und in der Dimension des Erklärens mit minimal unter $R_{EAP} = 0.694$ bei einem Leistungstest als akzeptabel angesehen werden. Der Q3-Mittelwert liegt mit $M_{Q3} = -0.004$ um 0.002 unterhalb des finalen Modells mit der Stichprobe N_1 . Der aQ3-Wert ($M_{aQ3} = 0.000$) ist ebenfalls optimal. Der SRMR ist um 0.001 gestiegen und liegt beim Modell mit der Stichprobe N_2 bei $SRMR = 0.087$. Dies bedeutet weiterhin eine akzeptable Passung. Die Infit- ($M = 1.007$) und Outfitmittelwerte ($M = 1.009$) sind gut, wenn auch leicht verschlechtert. Die WLE-Reliabilitäten entsprechen im Allgemeinen den Werten des finalen Modells der Stichprobe N_1 .

N_1 : Beschreiben $R_{WLE} = 0.720$, Erklären $R_{WLE} = 0.656$, Vorhersagen $R_{WLE} = 0.681$

N_2 : Beschreiben $R_{WLE} = 0.704$, Erklären $R_{WLE} = 0.617$, Vorhersagen $R_{WLE} = 0.662$

Der Modellvergleich zwischen beiden Stichproben ergibt minimal niedrigere AIC und BIC-Werte ($\Delta AIC = -79.55$, $\Delta BIC = -79.55$) für das finale Modell der Stichprobe N_1 (Tabelle 38). Der Unterschied zwischen beiden Modellen ist signifikant ($p = 0.000$). Das Modell mit der Datenstruktur der ersten Stichprobe lässt eine bessere Passung zum Modell vermuten.

Tabelle 38 Modellvergleich zwischen Stichprobe N1 und N2 mittels Anova (B = Beschreiben, E = Erklären, V = Vorhersagen, Dim = Dimensionalität, AIC = Akaike Information Criterion, Δ = Differenz, BIC = Bayesian Information Criterion)

Model	Dim	AIC	Δ AIC	BIC	Δ BIC
finalmod N ₁	3-dim: B, E, V	25170.01		25799.52	
model_2 N ₂	3-dim: B, E, V	25249.56	+ 79.55	25879.07	+ 79.55

10.4 Interpretation und Schlussfolgerungen

Das Kapitel 10 stellt die erste Validierung des Videotools dar. Mit der Annahme, die Professionelle Unterrichtswahrnehmung hinsichtlich der drei-dimensionalen Struktur des Knowledge-based Reasonings mit dem Modell abbilden zu können, wurde die Analyse begonnen. Offensichtlich können die vorliegenden empirischen Daten diese Struktur abbilden. Dieser Befund wurde bereits in mehreren OBSERVER-Studien (Seidel und Stürmer 2014, S. 759; Jahn 2014, S. 176) beschrieben und findet hier weitere Bestätigung. Die Ergebnisse der Arbeitsgruppe von Meschede et al. (2015, S. 326), die eine eindimensionale Struktur des Knowledge-based Reasonings validiert haben, können damit nicht unterstützt werden.

Ein Vergleich mittels Anova zeigt, dass das dreidimensionale Modell im Vergleich zum eindimensionalen oder den verschiedenen zweidimensionalen Modellen eine signifikant bessere Passung ergibt. Die AIC- und BIC-Werte sind beim dreidimensionalen Modell deutlich geringer als bei den anderen Modellen. Trotz der höheren Korrelation zwischen der Dimension des Beschreibens und Erklärens ($r = 0.61$) führt eine Zusammenführung beider Dimensionen nicht zu einem besseren Modellfit.

Resultat zur Forschungsfrage 6:

Das vorliegende Instrument weist auf die erwartete Konstruktvalidität eines dreidimensionalen Modells hin. Es kann mit dem vorliegenden Instrument unter den vorhandenen Bedingungen zwischen den drei Faktoren Beschreiben, Erklären und Vorhersagen des Knowledge-based Reasonings differenziert werden. Es muss darauf hingewiesen werden, dass eine absolute Aussage hinsichtlich der Modellstruktur nicht möglich ist, sondern lediglich ein vergleichendes Urteil darstellt.

Zur weiteren Bestätigung der Modellstruktur wurde das aufgestellte dreidimensionale Modell auf seine interne Konsistenz überprüft. Als Kriterium dient hierbei die EAP-Reliabilität, diese können analog zum Cronbachs α interpretiert werden. Die Werte sollten für eine gute Reliabilität mindestens bei $\alpha = 0.8$ liegen (Pospeschill 2010, S. 21) bzw. für Fähigkeitstest auch bei $\alpha = 0.7$ (Field et al. 2012, S. 806). Im Vergleich zum OBSERVER ($\alpha = 0.91-0.97$) (Stürmer und Seidel 2015, S. 60) erzielt ViviAn eine geringere EAP-Reliabilität, die sich jedoch innerhalb der Ergebnisse der erweiterten Version OBSERVER

Extended Research Tool von einer akzeptablen bis guten Reliabilität ($\alpha = 0.65-0.87$) bewegt (Stürmer und Seidel 2015, S. 60).

Resultat zur Forschungsfrage 4:

Die Reliabilität als Zeichen der internen Konsistenz weist eine akzeptable Messung auf. Die R_{EAP} -Werte liegen zwischen 0.718 und 0.756 und sind demnach für Leistungstests ausreichend. Für eine hohe interne Konsistenz müsste die Reliabilität höher ausfallen.

Bei der Itemanalyse des dreidimensionalen Modells wurden verschiedene Kriterien berücksichtigt, u.a. die Schwellenparameter, die punktbiseriale Korrelation, die stochastische Abhängigkeit und die Itemfitwerte. Bereits nach der ersten deskriptiven Analyse mussten vier Items wegen mangelnder Varianz und Range ausgeschlossen werden. Alle weiteren 162 Items wurden im Verlauf des iterativen Prozesses mehrfach analysiert. Im abschließenden Modell weisen alle Items hinsichtlich der Schwellenparameter eine geordnete Struktur auf. Somit liegt diesbezüglich keine Verletzung der Modellannahme vor. Die punktbiseriale Korrelation wird als ein Kriterium der Trennschärfe betrachtet. Sie kann Auskunft darüber geben, inwiefern die einzelnen Antworten der Items mit der Gesamtleistung im Zusammenhang stehen (Wu et al. 2016, S. 84). Im Partial-Credit-Modell wird erwartet, dass die punktbiseriale Korrelation der einzelnen Kategorien aufsteigend sind (Wu et al. 2016, S. 85). Mit dieser Annahme ist eine zunehmende Schwierigkeit verbunden. Das finale Modell weist bei 16 Items keinen Anstieg über alle drei Kategorien vor. Bei diesen Items liegt eine mangelnde Differenzierung zwischen Personen mit unterschiedlichen Fähigkeiten vor. Es wurde überprüft, inwiefern sich die Reliabilität durch den Ausschluss dieser Items (als reduziertes Modell bezeichnet) im Gegensatz zum finalen Modell aus der Raschskalierung verändert. Dabei zeigte sich, dass die Items zur internen Konsistenz beitragen:

Finales Modell:

Beschreiben $R_{EAP} = 0.756$, Erklären $R_{EAP} = 0.735$, Vorhersagen $R_{EAP} = 0.718$

Reduziertes Modell:

Beschreiben $R_{EAP} = 0.759$, Erklären $R_{EAP} = 0.694$, Vorhersagen $R_{EAP} = 0.682$

Eine mögliche Begründung kann im Auftreten von Kompensationseffekten liegen. Dies bedeutet, dass andere Items mit höherer Trennschärfe die mangelnde punktbiseriale Korrelation der Items kompensieren. Zusätzlich werden diese Items als kontextuell relevant wahrgenommen. Ein Ausschluss dieser Items führt dazu, dass die einzelnen Fragen zu den verschiedenen Vignetten mit zu wenig Antwortitems repräsentiert werden.

Zur Erläuterung ein Beispiel einer Frage aus der *Kategorie des Erklärens zum Thema zusätzliche Hilfestellungen zu der Vignette 1* (Übersicht der Vignetten siehe 7.5):

Beurteilen Sie, welche der folgenden Erklärungen bezüglich der Motivationsförderung theoretisch angemessen sind. Angenommen eine Lehrperson bietet zusätzliche Hilfestellungen (für schwierige Aufgaben) vor, dann geschieht dies...

...um die Erreichung der Lernziele zu ermöglichen (V1E21).

...um das Lernen zu erleichtern (V1E22).

...um einen Anreiz zum selbstorganisierten Lernen zu geben (V1E23).

Die Items V1E22 und V1E23 zeigen hinsichtlich der Punktbiserialen Korrelation keine ansteigenden Werte. Auffallend bei beiden Items ist die mittlere Kategorie. Hier ist die Trennschärfe zur oberen oder unteren Kategorie nicht eindeutig. Alle weiteren Skalenergebnisse wie Infit, Outfit und Q3 sind insgesamt gut, sodass die Items robust genug scheinen. Zudem wird das Gesamtmodell nicht substantiell beeinträchtigt. Beim Entfernen beider Items bliebe das Item V1E21 als einziges Item zu diesem Fragestamm erhalten. Dementsprechend tragen beide Items zur Konstruktdeckung bei.

Empfohlen wird dringend eine erneute Überprüfung der Raschskalierung mit einer weiteren Stichprobe. Falls die lokal kritischen Items weiterhin auffallen, sollten sie entfernt oder angepasst werden. Die stochastische Abhängigkeit ist zwischen vier Itempaaren besonders hoch. Die höchste Abhängigkeit wird dabei zwischen den Items V7V13 und V5E23 erreicht. Diese Items bilden verschiedene Schwerpunkte ab.

A) *Wahl zwischen zwei Aufgaben: Beurteilen Sie, welche der folgenden Erklärungen bezüglich der Motivationsförderung theoretisch angemessen sind. Angenommen eine Lehrperson stellt den Schülerinnen und Schülern zwei unterschiedliche Wahlpflichtaufgaben vor, dann geschieht dies...*

...um den Schülerinnen und Schülern die Möglichkeit zu geben, interessenbezogene Aufgaben zu wählen (V5E23).

B) *Fehler als Teil des Lernprozesses: Beurteilen Sie, welche Auswirkungen Fehler als Teil des Lernprozesses zu sehen, für den weiteren Lehr-Lernprozess hat.*

Deshalb üben die Schülerinnen und Schüler Fehler aufzudecken und zu korrigieren (V7V13).

An dieser Stelle wird aus inhaltlichen Gründen an beiden Items festgehalten. Zudem wird ein Item dem Erklären und eins dem Vorhersagen zugeordnet. Mittels einer zusätzlichen Faktorenanalyse aus der Klassischen Testtheorie könnte überprüft werden, ob beide Items einer Dimension zugeordnet werden können. Eine solche Analyse aller Items ist bei der IRT nicht vorgesehen, könnte aber zusätzliche Aufklärung bieten. Bei den anderen Itempaaren wurde wieder die Reliabilität zwischen dem finalen Modell und einem

reduzierten Modell mit Entfernen eines Items aus der jeweiligen Paarung (V1V23 und V1V24, V2V12 und V2V13, V2V14 und V2V23) verglichen. Es werden zwei reduzierte Modelle überprüft. Aufgrund der komplexen Rechenleistung werden die weiteren Varianten nicht analysiert, es wird von ähnlichen Ergebnissen ausgegangen.

Finales Modell:

Beschreiben $R_{EAP} = 0.756$, Erklären $R_{EAP} = 0.735$, Vorhersagen $R_{EAP} = 0.718$

Reduziertes Modell 1 mit Ausschluss der Items V1V23, V2V12, V2V23:

Beschreiben $R_{EAP} = 0.757$, Erklären $R_{EAP} = 0.735$, Vorhersagen $R_{EAP} = 0.672$

Reduziertes Modell 2 mit Ausschluss der Items V1V24, V2V13, V2V14:

Beschreiben $R_{EAP} = 0.756$, Erklären $R_{EAP} = 0.736$, Vorhersagen $R_{EAP} = 0.690$

Bei beiden Varianten verschlechtert sich die Reliabilität, sodass im Vorteil für die höhere interne Konsistenz die Items beibehalten werden. In ihren vergleichenden Analysen decken Christensen et al. (2017, S. 178) auf, dass Cut-off-Werte in Studien bis $Q3 = 0.7$ zugelassen werden, auch wenn damit von einer lokalen stochastischen Abhängigkeit ausgegangen werden muss. Diese Limitation bleibt im hier vorliegenden finalen Modell bestehen.

Als ein wesentliches Kriterium der Skalierung werden die Itemfitwerte gesehen. Es weisen nicht alle Items optimale Outfitwerte $0.7 < \text{Outfit}_i < 1.3$ (u.a. Bond und Fox 2012, S. 310) auf.

Entsprechend der Stichprobengröße kann ein Toleranzbereich von $0.52 < \text{Outfit}_i < 1.48$ (Smith et al. 1998, S. 78) bei 11 Items geltend gemacht werden. Das Item „Die Lehrperson nimmt sich zurück und hat nur einen geringen Sprechanteil“ (V4BS4) liegt außerhalb dieses Bereiches ($\text{Outfit}_i = 1.52$). Es weist hinsichtlich der Itemanalyse keine weiteren kritischen Werte auf, sodass dieses Item im Zuge der Abdeckung der Dimension des Beschreibens der Vignette 4 mit mindestens drei Items beibehalten wird.

Resultat zur Forschungsfrage 5:

Hinsichtlich des Gütekriteriums der Skalierung müssen Limitationen in den Ergebnissen der Itemfitwerte einiger Items berücksichtigt werden. Elf Items können hinsichtlich der Outfitwerte weiterhin als akzeptabel betrachtet werden. Ein Item liegt außerhalb dieses Bereiches und erfüllt das Kriterium nicht. Die Infitwerte aller Items weisen gute bis sehr gute Kennwerte auf. Die mittleren Itemfitwerte deuten zudem daraufhin, dass eine Passung zwischen dem theoretischen Modell und der Datenbasis vorliegt.

Der Modellvergleich an einer zweiten Stichprobe soll als Kriterium der Konstruktvalidität herangezogen werden. An dieser Stelle soll gleich zu Beginn deutlich gemacht werden, dass es sich bei dem Vorgehen in dieser Arbeit um keine „echte“ Validierung handelt. Es kann lediglich als erstes Zeichen einer Konstruktvalidität gesehen werden. Denn sowohl die Raschskalierung als auch die Überprüfung an einer zweiten Stichprobe werden in einer einzigen Erhebung durchgeführt. Im Idealfall erfolgt zunächst eine Raschskalierung mit anschließender Überprüfung der Ergebnisse bei einer neuen Erhebung an einer zweiten Stichprobe. Aufgrund der hohen Stichprobenzahl wurde die Stichprobe vor der Auswertung der Daten geteilt. Mit der einen Stichprobe wurde die Raschskalierung durchgeführt und entsprechend den Ergebnissen wurde das finale Modell an der zweiten Stichprobenhälfte hinsichtlich der Modellgüte überprüft. Es scheint, als würde das Modell mit der zweiten Stichprobe vergleichbare Ergebnisse erzielen (siehe 10.3 Ergebnisse). Im Modellvergleich lassen sich etwas schlechtere AIC- und BIC-Werte mit den Daten der zweiten Stichprobe erzielen ($\Delta\text{AIC} = + 79.55$, $\Delta\text{BIC} = + 79.55$). Die Passung mit der ersten Stichprobe ist dabei signifikant besser ($p = 0.000$).

Resultat zur Forschungsfrage 7:

Es lässt sich schließen, dass in dieser konkreten Erhebung unter gewissen Limitationen die Raschskalierung in einer zweiten Stichprobe zu vergleichbaren Ergebnissen führt. Dies kann als positives Zeichen für die Konstruktvalidität angesehen werden. Dabei handelt es sich nur um ein vergleichendes Urteil und es kann keine absolute Passung garantiert werden.

Zusammenfassend kann gesagt werden, dass das finale Modell akzeptable bis gute Ergebnisse vorweisen kann, sodass weitere Analysen damit erfolgen können.

11 Validierung des Videotools Teil II: Erhebung der Professionellen Wahrnehmung

In diesem Abschnitt werden zwei weitere Aspekte zum Beitrag zur Überprüfung der Konstruktvalidität des videobasierten Instruments untersucht. Es sollen zum einen Kenntnisse darüber gewonnen werden, wie gut das Tool Fähigkeiten hinsichtlich des Knowledge-based Reasoning misst. Ausgehend von Studien von Schäfer und Seidel (2015, S. 50–51) und Stürmer et al. (2015, S. 352) wird erwartet, dass die Teilnehmenden im Beschreiben die meisten Punkte erreichen, gefolgt vom Erklären. Die Analysefragen zum Vorhersagen können in der Regel am schwierigsten beantwortet werden (ebd.). Diese Differenzierung innerhalb des Reasonings soll am vorliegenden Instrument überprüft werden (theoriegeleitete Konstruktvalidität). Zum anderen wird der Frage nachgegangen, ob das Tool in der Lage ist, signifikante Unterschiede zwischen verschiedenen Gruppen von Teilnehmenden aufzeigen zu können (gruppendifferenzierte Konstruktvalidität). Es wird davon ausgegangen, dass die Analysefähigkeit im Verlauf des Studiums zunimmt (Meschede et al. 2015, S. 330).

Forschungsfragen

- 8 *Sind die angenommenen Schwierigkeitsgrade innerhalb des Reasonings mit dem Instrument überprüfbar?*
- 9 *Kann mit ViviAn eine Entwicklung in der Ausprägung der Professionellen Unterrichtswahrnehmung zu Beginn des Studiums im Vergleich zum Ende des Studiums erfasst werden?*

11.1 Stichprobe

Die Stichprobe der zweiten Validierung setzt sich aus Studierenden der Universität Koblenz-Landau am Standort Landau mit dem Schwerpunkt Lehramt ($N = 239$) und Lehramtsanwärter:innen (LAA) aus dem staatlichen Studienseminar für das Lehramt an Realschulen plus der Teildienststelle Landau-Land ($N = 41$) zusammen (siehe Tabelle 39).

Tabelle 39 Verteilung nach Semestern – Gesamtstichprobe (TN = Anzahl der teilnehmenden Personen pro Semester, LAA = Lehramtsanwärter:innen)

Semester	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	LAA
TN	17	92	20	30	11	25	6	18	6	6	3	1	1	1	1	1	41

Die Lehramtsanwärter:innen sind zum Zeitpunkt der Erhebung seit sieben Monaten im Vorbereitungsdienst. Aufgrund der Corona-Pandemie fand der Unterricht im Studienseminar hauptsächlich digital und in reduzierter Form statt. Zudem konnte der praktische

Einsatz in den Schulen nicht wie geplant erfolgen, sodass die Lehramtsanwärter:innen kaum bzw. keine zusätzlichen praktischen Erfahrungen sammeln konnten.

Da laut Forschungsfrage insbesondere interessant ist, ob Unterschiede in der Professionellen Unterrichtswahrnehmung zu Beginn im Vergleich zum Ende des Studiums aufgezeigt werden können, werden einzelne Semester zusammengefasst. Dies bietet zudem den Vorteil große Stichprobenunterschiede auszugleichen. Daher werden das 1. bis 3. Semester für den Anfang des Studiums, das 4. bis 6. Semester für die Mitte des Studiums und das 7. bis 10. Semester für das Ende des Studiums zusammengefasst. Zusätzlich werden die Referendare als eine weitere Gruppe betrachtet. Alle weiteren Semester werden ausgeschlossen, da es sich um Einzelfälle handelt, die sich über der Regelstudienzeit befinden und keine weiteren Informationen zu den Gründen der Verzögerung der individuellen Studienverläufe vorliegen.

11.2 Methodisches Vorgehen

Erhebungsinstrument

Zum Einsatz kamen die Analyseaufträge und Vignetten wie im Abschnitt 7.5 Beschreibung der Vignetten nachzulesen ist. Eingebettet wurden die Fragen und Videos mit Zusatzinformationen in die Lernumgebung ViviAn, die online über die Website abgerufen werden konnte. Alle Teilnehmer:innen haben eine schriftliche Einführung in das Videotool bekommen (siehe Anhang F). Für die Auswertung, ob das Videotool sowohl die Schwierigkeitsgrade als auch Unterschiede zwischen den Studierenden aufzeigen kann, wurde das finale, raschskalierte Modell mit einem reduzierten Itempool verwendet (*Limitationen* in der Vorgehensweise siehe Abschnitt 14.2). Insgesamt wurden für die Auswertung 104 Items berücksichtigt. Die Modellfitwerte sind im Abschnitt 10.4 (*Interpretation und Schlussfolgerung* der Validierung des Videotools Teil 1) zusammengefasst.

Auswertungsmethode

Als Vorbereitung für die Durchführung eines Mittelwertvergleichs wird die prozentual erreichte Punktezahl für die jeweiligen Dimensionen ermittelt. Dies ist erforderlich, da die Itemanzahl zwischen dem Beschreiben mit 24 Items und dem Erklären und Vorhersagen mit jeweils 42 Items unterschiedlich ist. Um die Forschungsfrage beantworten zu können, ob mit dem Instrument Unterschiede innerhalb des Reasonings aufgezeigt werden können, müssen sie zunächst vergleichbar dargestellt werden.

Einfaktorielle Varianzanalyse:

Zur Überprüfung, ob eine semesterabhängige Professionelle Unterrichtswahrnehmung vorliegt, wird eine einfaktorielle Varianzanalyse durchgeführt. Die Verwendung mehrerer t-Tests muss kritisch betrachtet werden. Auf Grund des Fehlerniveaus von 5% muss bei

einer mehrfachen Durchführung von t-Tests von einer α -Fehler Kumulierung ausgegangen werden (Rasch et al. 2014a, S. 3). Somit können signifikante Unterschiede mit der Zunahme der Irrtumswahrscheinlichkeit zusammenhängen (Zöfel 2003, S. 125). Aus diesem Grund kann mit einer einfaktorielle Varianzanalyse und damit dem gleichzeitigen Vergleich mehrerer Gruppen dem Problem begegnet werden (Rasch et al. 2014a, S. 4). In R wird für die Varianzanalyse die Funktion *aov* aus dem Paket *stats* (R Core Team 2023) verwendet. Bei signifikanten Ergebnissen wird mittels Post-hoc-Testung ein paarweiser Vergleich mit α -Fehler Korrektur durchgeführt (Rasch et al. 2014a, S. 28–29). Dazu wird in R das Argument *p.adjust* (R Core Team 2023) hinzugefügt. In der vorliegenden Analyse wird auf die Bonferroni-Adjustierung zurückgegriffen, da sie zu den konservativsten Methoden zählt (Eid et al. 2017, S. 417–418).

Varianzanalyse mit Messwiederholung:

Für die Überprüfung, ob es Unterschiede zwischen dem Beschreiben, Erklären und Vorhersagen gibt, wird eine Anova mit Messwiederholung durchgeführt. Hierbei muss es sich bei Messwiederholungen nicht um Zeitpunkte handeln, sondern um unterschiedliche Bedingungen (Eid et al. 2017, S. 462). Zudem liegt bei der vorliegenden Messung dieselbe Stichprobe vor. In R wird zur Berechnung das Paket *rstatix* (Alboukadel 2023) mit der Funktion *anova_test* angewendet.

11.3 Ergebnisse

Im ersten Teil der Ergebnisse wird der Frage nachgegangen, ob es signifikante Unterschiede zwischen den Kategorien des Beschreibens, Erklärens und Vorhersagens hinsichtlich der erreichten Punktzahl gibt. An dieser Stelle wird die gesamte *Stichprobe* wie in Abschnitt 11.1 beschrieben berücksichtigt. Deskriptiv lassen sich folgende Ergebnisse festhalten.

Tabelle 40 Mittelwerte (M) der normierten erreichten Punktzahl mittels Prozenttransformation über alle Proband:innen mit Standardabweichung (SD)

Kategorie	M	SD
Beschreiben	50.8	12.9
Erklären	53.1	9.28
Vorhersagen	51.1	9.45

Die Anova weist auf signifikante Unterschiede $F(2,492) = 5.02; p = 0.009, f = 0.018$ zwischen den drei Kategorien hin. Beim paarweisen Vergleichen lassen sich adjustierte signifikante Unterschiede zwischen dem Beschreiben und Erklären ($p = 0.008$) sowie zwischen dem Erklären und Vorhersagen ($p = 0.010$) bestimmen. Zwischen dem Beschreiben

und Vorhersagen können keine signifikanten Unterschiede berichtet werden ($p = 1$), hier bleibt die Nullhypothese bestehen.

Im zweiten Teil der Ergebnisse wird geprüft, ob es Unterschiede im Beschreiben, Erklären und Vorhersagen innerhalb der Studierendenschaft bzw. zu den Lehramtsanwärter:innen in Bezug auf die Erfahrung gibt. Gemessen wird die Erfahrung in diesem Fall an der Semesteranzahl und der Zugehörigkeit zum Studienseminar. In der Kategorie des Beschreibens können mittels der Anova keine Unterschiede zwischen dem Anfang, der Mitte und dem Ende des Studiums bzw. des Referendariats festgestellt werden ($F(1,270) = 1.988, p = 0.160$). Ebenso bleibt die Null-Hypothese beim Erklären bestehen, dass keine Unterschiede zwischen den Gruppen existieren ($F(1,270) = 0.517, p = 0.473$). Lediglich im Vorhersagen deuten die Ergebnisse auf signifikante Unterschiede zwischen den Gruppen hin: $F(1,270) = 5.045, p = 0.026$. Der paarweise t-Test mit bonferroni-Adjustierung gibt Aufschluss über die genauen Unterschiede. Hierbei lassen sich ausschließlich zwischen dem Anfang und dem Ende des Studiums Gruppenunterschiede aufdecken ($p = 0.050$).

11.4 Interpretation und Schlussfolgerungen

Die Ergebnisse zur Prüfung von Unterschieden zwischen den drei Kategorien innerhalb des Knowledge-based Reasonings zeigen, dass es Unterschiede zwischen dem Erklären und Vorhersagen gibt. Dabei erreichen die Personen im Mittel im Test in der Kategorie des Erklärens bessere Ergebnisse als im Vorhersagen. Dieses Resultat deckt sich mit anderen Forschungen (Stürmer et al. 2015, S. 352; Schäfer und Seidel 2015, S. 50–51; Meschede et al. 2015, S. 328; Vgl. Frommelt et al. 2016, S. 369). Allerdings gehen diese auch davon aus, dass die Studierenden und Lehramtsanwärter:innen im Beschreiben die höchste Punktzahl erreichen (ebd.). In der vorliegenden Erhebung konnte dieses Ergebnis nicht repliziert werden. Die teilnehmenden Personen haben im Mittel im Beschreiben die geringste Punktzahl erreicht. Gründe dafür können darin liegen, dass es vor allem Noviz:innen schwer fällt, Situationen genau zu beschreiben (Seidel et al. 2010, S. 297–298). Weil der Unterrichtsalltag sehr komplex und dynamisch ist (Oser et al. 2010, S. 6), kann es für die Studierenden herausfordernder sein, präzise die Situation wahrzunehmen und zu beschreiben. Zum einen kann eine Überforderung durch die Fülle an Informationen in der Videosequenz eine Ursache dafür sein (von Aufschnaiter et al. 2017, S. 95; Syring et al. 2015, S. 670). Zum anderen zeigt die Metaanalyse von Keskin et al. (2024), dass Studierende eher den Fokus auf das Wesentliche verlieren. Ein weiterer Aspekt liegt möglicherweise im häufig beklagten fehlenden Theorie-Praxis-Bezug im Studium (Ertl-Schmuck 2019, S. 49), der es den Studierenden erschwert, Theorien an praktischen Beispielen anzuwenden (Ertl-Schmuck 2019, S. 48). Im Studienalltag werden vermehrt der Ansatz einer Evidenzbasierung und theoriegeleitete Lehre verfolgt (Bresges et al. 2019, S. 4). Daher ist es möglich, dass Studierende besser Situationen wissensbasiert erklären

als beschreiben können. Eine Etablierung eines videobasierten Tools zur Verknüpfung von Theorie und Praxis könnte hilfreich sein, diese Fähigkeiten zu fördern. Andernfalls verbessert sich das wissensbasierte Beschreiben erst mit zunehmender beruflicher Erfahrung (Seidel und Prenzel 2007, S. 211–213), unter anderem, weil es Expert:innen leichter gelingt, Situationen zielgerichtet wahrzunehmen (Huang et al. 2023, S. 131). Zur Überprüfung dieser Annahmen wäre eine Expertiseforschung mit dem vorhandenen Instrument notwendig. Da im Rahmen der vorliegenden Erhebung keine Lehrkräfte das Videotool angewendet haben, ist kein Vergleich mit den Studierenden möglich. Ein weiterer Erklärungsansatz dafür, dass die Proband:innen im Beschreiben weniger Punkte bekommen, kann mit dem Aufbau der Fragenstruktur zusammenhängen (siehe Abschnitt 8.1 Entwicklung der Analyseaufträge). Beim Erklären und Vorhersagen wird das zu beobachtende Handeln vorgegeben. Damit ist eine von der Wahrnehmung und Beschreibung unabhängige wissensbasierte Begründung für das Handeln und das Vorhersagen für den weiteren Lehr-Lernprozess möglich. Folglich kann die Beantwortung dieser Items einfacher sein.

Resultat zur Forschungsfrage 8:

Die ursprüngliche Annahme, dass die Schwierigkeit der Beantwortung der Analyseaufträge vom Beschreiben, zum Erklären und Vorhersagen zunimmt, kann mit dem vorliegenden Tool nur bedingt bestätigt werden. Zwischen dem Erklären und Vorhersagen nimmt die Fähigkeit der Studierenden erwartungsgemäß ab. Beim Beschreiben erreichen die Studierenden jedoch im Mittel die geringste Punktzahl. Es lassen sich signifikante Unterschiede zwischen dem Beschreiben und Erklären sowie dem Erklären und Vorhersagen aufzeigen. Zwischen dem Beschreiben und Vorhersagen muss die Nullhypothese, keine Unterschiede zwischen den Gruppen feststellen zu können, bestehen bleiben. Weitere Forschungen sind notwendig, um zu überprüfen, ob diese Ergebnisse mit erfahrenen Lehrpersonen ähnlich ausfallen oder sich bei ihnen die angenommenen Schwierigkeitsgrade innerhalb des Knowledge-based Reasonings mit dem Instrument überprüfen lassen.

Weiterhin wurde erwartet, dass mittels des vorliegenden Instruments Unterschiede innerhalb der Studierendenschaft in der Ausprägung der Professionellen Unterrichtswahrnehmung im Knowledge-based Reasoning aufgezeigt werden können. Die Ergebnisse dieser Analyse lassen lediglich leichte signifikante Unterschiede im Vorhersagen zwischen dem Anfang und dem Ende des Studiums zu. Im Beschreiben und Erklären liegen keine signifikanten Unterschiede vor. Wenn man davon ausgeht, dass Lehrpersonen erst mit zunehmender praktischer Erfahrung Unterrichtsgeschehnisse systematisch beurteilen können (Seidel und Prenzel 2007, S. 214), dann können kaum Unterschiede innerhalb des Studiums erwartet werden. Bereits Stürmer et al. (2015, S. 355) wiesen vergleichbare

Ergebnisse zwischen Studierenden und Referendar:innen nach, was sich beim Beschreiben und Erklären mit dieser Arbeit deckt.

Beim Beschreiben ist im Allgemeinen bekannt, dass Noviz:innen Situationen weniger gut beschreiben können als Expert:innen (Seidel et al. 2010, S. 297–298). Hierbei bleibt die Frage bestehen, ab wann Studierende über ausreichendes Wissen und Erfahrungen verfügen, um genügend Expertise für die zuverlässige Beschreibung von Unterrichtssituationen zu besitzen. Laut Berliner (2004, S. 3) braucht es für den Expertisestatus ausreichend praktische Erfahrung im Klassenzimmer. Unterrichtspraktika oder Erfahrungen als PES-Kraft können Einfluss auf dieses Erfahrungswissen während des Studiums nehmen. Solche Auswertungen wurden in der vorliegenden Erhebung nicht vorgenommen, hätten aber ein differenzierteres Bild über den Erfahrungsstatus liefern können. Möglicherweise können diese ersten Erfahrungen im Klassenzimmer Unterschiede im Vorhersagen zwischen den Anfängen im Studium und dem Ende hervorbringen. Alternativ wäre es denkbar, dass Studierende höheren Semesters tatsächlich besser situativ antizipieren können. Eine vertiefende Folgestudie könnte diesen Aspekt genauer analysieren. Möglicherweise spielen auch Primacy-Effekte eine Rolle. Die vorhergehenden Items zum Erklären können die Denkprozesse für das Vorhersagen beeinflusst haben. Dies müsste auf Grund der gleichen Gegebenheiten für Studierende aller Semester genauer betrachtet werden, um der Frage nachzugehen, warum die Studienanfänger:innen davon weniger profitieren würden.

Resultat zur Forschungsfrage 9:

Unterschiede in der Ausprägung der Professionellen Unterrichtswahrnehmung zu Beginn des Studiums im Vergleich zum Ende des Studiums können lediglich im Vorhersagen erfasst werden. Im Beschreiben und Erklären lassen sich keine signifikanten Differenzen feststellen. Damit ist das Instrument nur eingeschränkt in der Lage, zwischen den Studierenden unterschiedlichen Semesters zu differenzieren. Ursachen dafür müssen weiter untersucht werden.

12 Validierung des Tools Teil III: Förderung der Professionellen Wahrnehmung

Aufbauend auf den Ergebnissen aus Kapitel 11 *Erfassung der professionellen Unterrichtswahrnehmung*, stellt sich die Frage, ob die Professionelle Unterrichtswahrnehmung der Studierenden bereits vor dem täglichen Schuldienst als Lehrkraft gefördert werden kann. In diesem Kapitel wird daher die Sensitivität des Videoinstruments überprüft und folgender Forschungsfrage nachgegangen:

Forschungsfrage

10 Erweist sich ViviAn als geeignetes Instrument, um eine Entwicklung der Professionellen Wahrnehmung hinsichtlich der Motivierung durch eine Intervention aufzuzeigen?

Es soll demnach erforscht werden, ob die Intervention in Form eines Seminars Einfluss auf die Professionelle Unterrichtswahrnehmung hinsichtlich der Motivierung hat. Deswegen wird eine Pre-Post-Messung mit dem vorliegenden Instrument im Abstand von vier Wochen durchgeführt. Es wird erwartet, dass sich die Teilnahme am Seminar positiv in den Ergebnissen der Posttestung widerspiegeln lässt (Vgl. Gold und Hasselhorn 2013, S. 149).

12.1 Stichprobe

Die Stichprobe setzt sich aus Studierenden zusammen, die sich im Rahmen des Freien Workloads im Fachbereich Erziehungswissenschaften für das Seminar „Lern-umgebungen gestalten“ eingeschrieben haben. Es hatten alle Studierenden im Lehramt die Möglichkeit, sich für dieses Seminar aus einem Wahlpflichtpool zu entscheiden. Im Rahmen des Seminars mussten die Studierenden sowohl in einer Pre- als auch Posttestung das Videotool mit den Analyseaufträgen zu motivationsförderlichen Aspekten bearbeiten. Für die folgende Auswertung werden nur Studierende berücksichtigt, die an beiden Testungen teilgenommen und alle Vignetten bearbeitet haben (N=203).

Tabelle 41 Verteilung nach Semestern – Stichprobe Pre- und Posttest (TN = Anzahl der teilnehmenden Personen pro Semester)

Semester	1	2	3	4	5	6	7	8	9	10	16	gesamt
TN	17	88	18	27	11	23	4	10	2	2	1	203

12.2 Beschreibung des Seminars

Besondere Berücksichtigung muss finden, dass das Seminar aufgrund der Coronapandemie nur im digitalen Format stattfinden konnte. Hierzu wurde ein selbstgesteuertes Lernkonzept innerhalb eines Olat-Kurses angelegt. Insgesamt haben sich über 200 Studierende für das Seminar angemeldet. Um den Studierenden ein gutes Lernformat anbieten zu können, wurde das Seminar auf insgesamt 10 Workshops aufgeteilt, sodass nicht mehr als 25 Studierende einem Workshop zugeteilt waren. Alle Workshops waren gleich strukturiert und zeitlich datiert. Jeder Workshop gliederte sich in drei Phasen. In der ersten Phase sollten die Studierenden mit dem Videotool ViviAn reale Unterrichtssituationen analysieren. Dazu haben die Studierenden eine schriftliche Anleitung zum Videotool erhalten (siehe Anhang F). In der zweiten Phase standen Online-Materialien zur Erarbeitung motivationsrelevanter Aspekte im Unterricht zur Verfügung, die über dynamische PDFs und interaktive Elemente (kollaborative Dokumente und Foren) bearbeitet werden sollten. Zusätzlich wurden einzelne Videosequenzen (nicht aus dem Videotool) eingesetzt, die mit Analyse-, Bewertungs- oder Reflexionsfragen versehen waren. Der Fokus lag auf einer hohen Theorie-Praxis-Verzahnung und Reflexion. Eine Sicherstellung der Erarbeitung der Themen und der Ergebnisse konnte über das Hochladen der bearbeiteten Dokumente und der Beiträge der Studierenden innerhalb des Olat-Kurses gewährleistet werden. Zudem konnten die Studierenden ihre Ergebnisse mit Musterlösungen abgleichen. Im Anschluss sollten die Studierenden in der dritten Workshopphase erneut die Videovignetten bearbeiten. Nach Abschluss des Workshops haben alle Studierenden ein persönliches Kompetenzprofil zu ihren Analysefähigkeiten erhalten.

12.3 Methodisches Vorgehen

Als Design wurde ein Pre-/Posttest mit Intervention gewählt (siehe Abbildung 36). Die Pretestung fand vor dem inhaltlichen Input des Seminars statt und die Posttestung 4-5 Wochen später. Nach dem Pretest haben die Studierenden keine Auswertung oder Musterlösung erhalten. Im Posttest wurden exakt die gleichen Vignetten und Fragen eingesetzt, um einen direkten Vergleich ziehen zu können. Die Studierenden wurden im Vorfeld darüber informiert sowie über die Ziele dieses Vorgehens. Am Ende haben die Studierenden ein individuelles Fähigkeitsprofil erhalten.

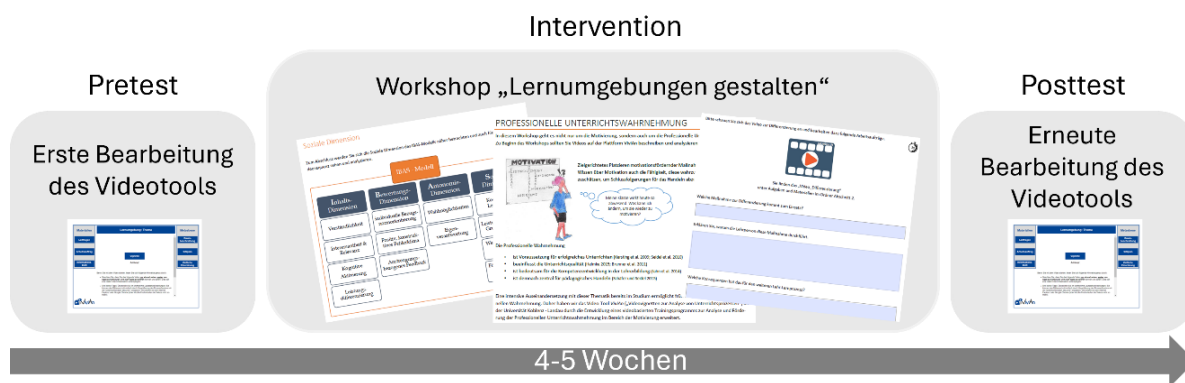


Abbildung 34 Studiendesign

Aufgrund von organisatorischen Gründen, insbesondere wegen der coronabedingten kurzfristigen und zeitintensiven Umstellung des Seminarkonzepts auf ein online-Format, musste auf eine Kontrollgruppe verzichtet werden.

Erhebungsinstrument

Wie bereits im Abschnitt 11.2 beschrieben, wurde auch für diese Untersuchung das vorliegende raschskalierte Modell mit einem reduzierten Itempool angewendet (Vorgehensweise, Ergebnisse und Interpretationen siehe Kapitel 10 *Validierung des Videotools Teil I: Raschskalierung*).

Auswertungsmethode

t-Test für verbundene Stichproben:

Um überprüfen zu können, ob es zufällige oder bedeutsame Unterschiede hinsichtlich der Ausprägung der Professionellen Unterrichtswahrnehmung bei den Studierenden zwischen den zwei Messzeitpunkten (vor und nach dem Seminar) gibt, wird ein t-Test durchgeführt (Vgl. Rasch et al. 2014b, S. 34). Da die Stichprobe die gleiche ist und sich nur der Messzeitpunkt unterscheidet, wird das Testverfahren eines t-Tests für abhängige Stichproben angewendet (Rasch et al. 2014b, S. 62; Luhmann 2015, S. 187). Die Voraussetzungen für den t-Test sind folgende:

- Die Daten müssen metrisch skaliert sein (Luhmann 2015, S. 187).
- Die Messwerte müssen eindeutig zugeordnet werden können (Eid et al. 2017, S. 368). In der vorliegenden Erhebung haben alle Studierenden genau zwei Messwerte, die zusammengehören: ein Wert aus dem Pretest und ein Wert aus dem Posttest.
- Zudem dürfen sich die Personen innerhalb der Stichprobe nicht gegenseitig beeinflussen (Eid et al. 2017, S. 368).
- Die Normalverteilung muss zwischen der Differenz der Messwertpaare vorliegen (Zöfel 2003, S. 128; Eid et al. 2017, S. 71). Bei ausreichend großer Stichprobe ist der t-Test robust hinsichtlich nicht normalverteilter Differenzvariable. Alternativ kann bei

einer kleinen Stichprobe der Wilcoxon-Test herangezogen werden (Eid et al. 2017, S. 371). Bei der vorliegenden Stichprobe mit $N > 200$ ist selbst bei einer sehr schiefen Verteilungsannahme die Verletzung der Normalverteilung nicht ausschlaggebend.

Die Berechnung des t-Tests erfolgt mittels der Funktion *t.test* aus dem Paket *stats* (R Core Team 2023) in R. Zur Formulierung der Abhängigkeit wird die Bedingung *paired = TRUE* gesetzt (Luhmann 2015, S. 190). Die Voraussetzung ist, dass immer zwei Wertepaare vorliegen müssen, ansonsten kann die Berechnung nicht durchgeführt werden. Es wird von der Hypothese ausgegangen, dass die Personen auf Grund der Intervention im Vortest weniger Punkte erreichen als im Nachtest. Daher wird diese Annahme mittels eines einseitigen t-Tests und des zusätzlichen Arguments *alternative = "less"* geprüft (Luhmann 2015, S. 179).

Effektstärke

Die Effektgröße der Mittelwertunterschiede wird mittels der Effektgröße Cohens *d* bestimmt (Döring und Bortz 2016, S. 816). Bei abhängigen Stichproben (Cohen 1988, S. 46) gelten andere Richtwerte als bei unabhängigen Stichproben (Cohen 1977, S. 25–26):

Tabelle 42 Effektstärke nach Cohen (1977, 1988)

Effektstärke	Unabhängige Stichprobe	Abhängige Stichprobe
Kleiner Effekt	$d = 0.2$	$d = 0.14$
Mittlerer Effekt	$d = 0.5$	$d = 0.35$
Großer Effekt	$d = 0.8$	$d = 0.57$

12.4 Ergebnisse

In den Ergebnissen der Raschskalierung zeigte sich, dass das dreidimensionale Modell besser gefittet hat als ein zwei- oder eindimensionales Modell. Daher wird der t-Test für alle drei Dimensionen separat durchgeführt. Es wird angenommen, dass die Unterschiede zwischen dem Pre- und Posttest signifikant sind und die Personen im Nachtest bessere Ergebnisse erzielen. In der Dimension des Beschreibens muss diese gerichtete Hypothese verworfen werden ($t(202) = 6.057, p = 1$). Im Pretest ist der Mittelwert ($M_{Pre} = 24.18, SD_{Pre} = 6.18$) höher als im Posttest ($M_{Post} = 21.98, SD_{Post} = 5.81$). Im Folgenden wird ein t-Test für abhängige Stichproben ohne Richtung durchgeführt. Dabei kann die Signifikanz bestätigt werden ($t(202) = 6.057, p < 0.001$). Die Effektstärke liegt bei einem mittleren Effekt ($d = -0.367$). Die Professionelle Unterrichtswahrnehmung in der Kategorie des Erklärens zeigt im Posttest signifikant höhere Werte ($t(202) = -2.648, p = 0.004, M_{Pre} = 44.9, SD_{Pre} = 7.3, M_{Post} = 45.97, SD_{Post} = 6.78$). Damit kann die Alternativhypothese angenommen werden. Die Effektstärke ist mit $d = 0.152$ klein. Im Vorhersagen kann

ebenfalls die Alternativhypothese bestätigt werden. Der Mittelwert im Posttest ($M_{Post} = 46.59$, $SD_{Post} = 8.07$) ist signifikant höher ($t(202) = -8.310$, $p < 0.001$) als im Pretest ($M_{Pre} = 42.94$, $SD_{Pre} = 7.47$). Auch hier lässt sich ein mittlerer Effekt ($d = 0.469$) berechnen.

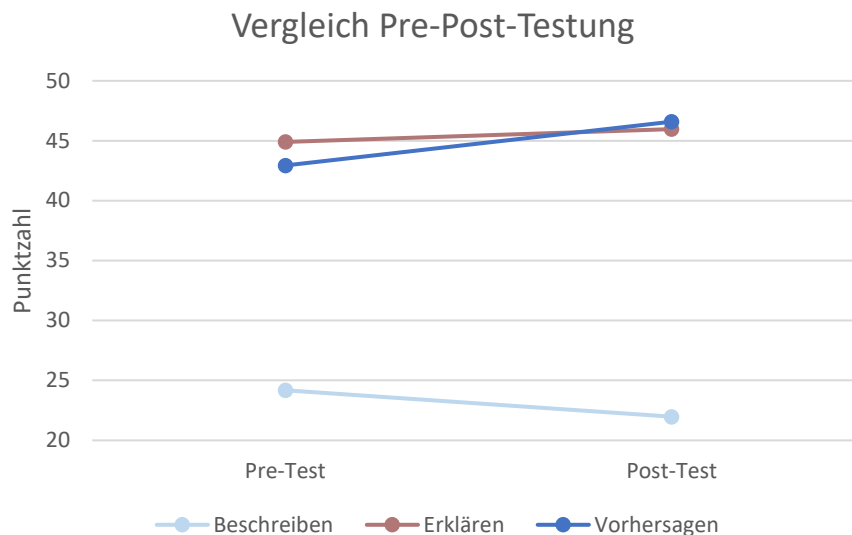


Abbildung 35 mittlerer Summenscore pro Dimension im Vergleich zwischen Pre- und Posttestung (Anmerkung: im Beschreiben können maximal 40 Punkte und im Erklären und Vorhersagen maximal 84 Punkte erreicht werden)

12.5 Interpretation und Schlussfolgerungen

Ein Test-Instrument, das in der Lage ist, auf Veränderungen zu reagieren, kann eine Änderungssensitivität aufzeigen. In der vorliegenden Untersuchung sollte überprüft werden, ob sich das Videotool in einem Seminar zum Thema Motivierung im Unterricht als sensibles Analyseinstrument eignet.

Als Limitation soll zu Beginn klar aufgezeigt werden, dass die Auswertung der Intervention auf Grund der vorhandenen Ressourcen nicht an einer neuen, unabhängigen Stichprobe durchgeführt werden konnte. Dies bedeutet, dass sowohl die Raschskalierung als auch die Konstruktvalidität und Änderungssensitivität an der gleichen Stichprobe überprüft wurde. In diesem Fall handelt es sich um eine erweiterte Auswertung, deren Ergebnisse lediglich als Tendenz gesehen werden können. Zudem kann nicht ausgeschlossen werden, dass Erinnerungseffekte einfließen. Mit einer Kontrollgruppe hätte man solche Effekte berechnen können. Diese war in der aktuellen Erhebung jedoch nicht berücksichtigt worden.

Die Annahme bestand darin, dass die angewandte Intervention in Form einer thematischen Auseinandersetzung mit praktischen Text- und Videobeispielen zur Förderung der Professionelle Unterrichtswahrnehmung beiträgt. Damit könnten die Befunde früherer Studien bestätigt werden, die zeigen, dass sich die Professionelle Unterrichtswahrneh-

mung durch eine Videointervention verbessert (Frommelt et al. 2016, S. 368). Im Gegensatz zur pauschalen Annahme der Förderung weisen Sunder et al. (2016, S. 9) und Gold und Hasselhorn (2013, S. 149) daraufhin, dass die Verbesserung themenspezifisch ist.

Anhand der Ergebnisse wird deutlich, dass das vorliegende Instrument sensibel auf Veränderung reagiert. In allen drei Dimensionen konnten signifikante Unterschiede zwischen den zwei Messzeitpunkten ermittelt werden. Allerdings muss dieser Effekt in der Kategorie des Beschreibens genauer differenziert werden. Hier zeigt sich ein unerwartetes Ergebnis, das von den bisherigen Forschungen abweicht, bei denen in allen drei Kategorien Verbesserungen eintraten (siehe oben genannte). Im Mittelwert werden die Studierenden bei der zweiten Messung im Anschluss des Seminars schlechter. Eine mögliche Begründung kann mit dem Inhalt des Seminars und der konkreten Aufgabenstellung in Zusammenhang gebracht werden (siehe Abbildung 36). Es fällt auf, dass der Fokus der selbstgesteuerten Lernelemente vor allem auf die Analyse von Situationen mit dem Erklären und Vorhersagen gelegt wurde. Hier wurden ausführlichere Antworten erwartet und in Form einer Musterlösung präsentiert. Dies würde auch belegen, dass die Studierenden sich im Mittel in der Dimension des Erklärens und Vorhersagens zum zweiten Messzeitpunkt verbesserten.

Bitte schauen Sie sich das Video zur Differenzierung an und bearbeiten dazu folgende Arbeitsaufträge.



Sie finden das „Video_Differenzierung“ unter Aufgaben und Materialien im Ordner Abschnitt 2.

Welche Maßnahme zur Differenzierung kommt zum Einsatz?

Erklären Sie, warum die Lehrperson diese Maßnahme durchführt.

Welche Konsequenzen hat das für den weiteren Lehr-Lernprozess?

Welche Verbesserungsvorschläge bzw. Hinweise würden Sie der Lehrperson geben?

Abbildung 36 Beispiel für eine Aufgabenstellung im Rahmen des Seminars

Ein weiterer Grund kann das fehlende direkte Feedback zum Beschreiben einer Situation sein. Häufig fällt es Studierenden schwer, wie bereits in Abschnitt 11.4 dargestellt, den Fokus auf das Wesentliche zu richten und differenziert genug zu beschreiben (Seidel et al. 2010, S. 297–298). Aufgrund des coronabedingten Settings fehlte eine direkte Rückmeldung zu den einzelnen Aufgabenbereichen, um diesen Effekten entgegenwirken zu können. Zudem konnte Enenkiel (2022, S. 303) in ihrer Forschungsarbeit aufzeigen, dass der Zeitpunkt des Feedbacks Einfluss auf die aktive Verweildauer auf der Rückmeldung hat. Es war bei der vorliegenden Erhebung im Online-Setting nicht nachvollziehbar, ob und wann die Studierenden auf die Musterlösung zugegriffen haben.

Resultat zur Forschungsfrage 10:

Es lassen sich Rückschlüsse auf die Änderungssensitivität ziehen. Unter den gesetzten Bedingungen kann mit dem Instrument eine Veränderung in allen drei Dimensionen durch den Einsatz einer Intervention in Form eines Seminars aufgezeigt werden. Dabei ist jedoch anzumerken, dass lediglich im Erklären und Vorhersagen die erwartete Annahme der Verbesserung von MZP1 zu MZP2 bestätigt werden kann. Wie nachhaltig der Effekt ist, bleibt offen. Im Beschreiben werden die Personen im Mittelwert nach dem Seminar schlechter.

Teil 3

Diskussion und Ausblick

13 Zusammenfassung der Ergebnisse

Ziel der vorliegenden Arbeit war es, ein videobasiertes Instrument zu entwickeln, das in der Lage ist, die Professionelle Unterrichtswahrnehmung angehender Lehrpersonen hinsichtlich motivationsrelevanter Aspekte im Unterricht zu analysieren und zu fördern. Damit soll ein Beitrag zur Weiterentwicklung der Lehrkräftebildung geleistet werden, indem bereits frühzeitig im Studium notwendige Fähigkeiten erworben werden können. Die große Herausforderung einer Lehrkraft ist es, im komplexen Unterrichtsgeschehen relevante Situationen wahrnehmen, wissensbasiert beschreiben, erklären und Prognosen für den weiteren Verlauf abgeben zu können (Seidel et al. 2010, S. 297; Sherin und van Es 2008, S. 22). Dieser besondere Prozess der Professionellen Unterrichtswahrnehmung (siehe Kapitel 3) ist stark von der Erfahrung einer Lehrperson im Schulunterricht abhängig (Seidel und Prenzel 2007, S. 211–213). Die Professionelle Unterrichtswahrnehmung trägt zum Erfolg des Einsatzes motivationsförderlicher Maßnahmen im Unterricht bei. Es sollte bereits frühzeitig begonnen werden, diese bei den künftigen Lehrpersonen zu fördern, da die Motivation der Schüler:innen einen erheblichen Einfluss auf deren Leistung hat (Steinmayr und Spinath, 2007). Die Lehrpersonen haben die Möglichkeit mit der motivationsförderlichen Gestaltung des Unterrichts, den Lernverlauf positiv zu beeinflussen (Kunter et al. 2011, S. 59) (siehe Kapitel 2 *Motivation und Motivierung im Unterricht*). Aus diesen Gründen bestand das Ziel dieser Arbeit darin, ein valides und standardisiertes Testinstrument zur Analyse der Professionellen Wahrnehmung hinsichtlich der Motivierung zu entwickeln (siehe Kapitel 6 *Zielsetzung und Fragestellung*). Dieses Instrument soll dazu genutzt werden, die aktuellen Fähigkeiten einer Person beschreiben zu können. Als geeignetes Instrument schien ein videobasiertes Tool (siehe Abschnitt 4.1 *Die fünf Heuristiken*). Videovignetten haben den Vorteil, das Unterrichtsgeschehen authentisch abzubilden (Helmke 2015, S. 342; Syring et al. 2015, S. 670). Das bereits an der Universität Koblenz · Landau entwickelte Video-Tool *ViviAn* („Videovignetten zur Analyse von Unterrichtsprozessen“) (siehe Abschnitt 4.3. *Das Videotool ViviAn*) konnte dafür verwendet und um zusätzliche Vignetten zum Thema Motivierung ergänzt werden. Dieser Teil des Instruments wurde hinsichtlich der Erfüllung der verschiedenen *Gütekriterien der Testentwicklung* (siehe Kapitel 5) untersucht. Im Folgenden werden die Ergebnisse dieser Analyse strukturiert entsprechend der Gütekriterien und Forschungsfragen zusammengefasst.

Inhaltliche Validität Stellen die Vignetten repräsentative motivationsförderliche oder -hemmende Situationen dar?

Für das Videotool wurden Filmaufnahmen aus dem Unterricht einer 7. und 8. Klasse zu Videovignetten geschnitten und in das Videovignettenanalyse-Tool (*ViviAn*) eingebettet. Die Validierung der Vignetten, ob sie motivationsförderliche oder -hemmende

Situationen repräsentieren, erfolgte über die aktuelle Motivation der Schüler:innen. Diese konnte mittels wiederholt eingesetzter Fragebögen während des Unterrichts erhoben werden. Die Daten geben Auskunft, inwieweit die Unterrichtsabschnitte mit einer höheren oder geringeren Motivation verbunden sind. Mit der Überprüfung auf signifikante Unterschiede zwischen den einzelnen Messzeitpunkten konnten geeignete Filmaufnahmen gefiltert werden (z.B. intrinsische Motivation der Klassenstufe 8 zum MZP1: $M = 1.57$, $SD = 0.64$; MZP7: $M = 2.43$, $SD = 1.01$; $t = -2.98$, $p = .01$, $d = 1.02$). Diese Daten wurden zudem mit dem Unterrichtsimpuls an die Lehrperson, motivationsförderliche und -hemmende Sequenzen vorzunehmen (z.B. MZP1: Reduktion der Adäquaten Freiräume und des Positiv-konstruktiven Fehlerklimas; MZP7: Geben eines anstrengungsbezogenen Feedbacks), verglichen und auf Kongruenz überprüft. Mit diesem Vorgehen konnten inhaltlich valide Vignetten für das Videotool ausgewählt werden.

Auswertungsobjektivität Ist die Übereinstimmung der Expert:innen bei der Auswertung der Rating-Items ausreichend hoch, um von einer hohen Auswertungsobjektivität ausgehen zu können?

Nach der Validierung der Vignetten erfolgte die Entwicklung von geeigneten Fragen, die der späteren Analyse der Filmsequenzen dienen. Diese Fragen sind entsprechend der Kategorien des Knowledge-based Reasonings als Teilprozess der Professionellen Unterrichtswahrnehmung in drei Kategorien gegliedert: das Beschreiben, Erklären und Vorhersagen. Zur Beurteilung der Ausprägung der Professionellen Unterrichtswahrnehmung der späteren Anwender:innen des Tools werden die Antworten mit denen von Expert:innen verglichen. Diese Expert:innenaussagen dienen auch der Validierung der Analyseaufträge. Zur Erreichung einer hohen Auswertungsobjektivität wird eine hohe Übereinstimmung der Antworten der Expert:innen erwartet. Als Kriterium dient das Fleiss-Kappa. In der Kategorie des Beschreibens wird entsprechend der Einteilung nach Landis und Koch (1977, S. 164–165) ein zufriedenstellendes Ergebnis erreicht (Fleiss' $\kappa = 0.7$). In den Kategorien des Erklärens und Vorhersagens ist die Übereinstimmung gut (Fleiss' $\kappa = 0.78$). Somit kann von einer ausreichend hohen Auswertungsobjektivität ausgegangen werden.

Nützlichkeit Können die Vignetten zum Themenbereich Motivierung bei den Studierenden das Interesse an der Arbeit mit Videos wecken sowie eine hohe Relevanz wahrnehmen lassen?

In einer ersten Pilotierung des entwickelten Tools wurden Studierende mittels einer Fragebogenerhebung nach ihrer Beurteilung hinsichtlich der wahrgenommenen Relevanz, der Interessantheit und der Gestaltung der Lernumgebung gefragt. Als Ergebnis kann festgehalten werden, dass die Studierenden das Tool positiv bewerten. Die Interessantheit erreicht im Mittel einen Summenscore von $M = 9.69$, $SD = 3.17$. Der maximale Score

liegt bei der Beurteilung der Interessantheit und Relevanz bei 15 Punkten. Die Relevanz erreicht im Mittel einen leicht höheren Score ($M = 10.52$, $SD = 2.52$). Die Lernumgebung ViviAn wird ebenfalls als positiv gestaltet beurteilt ($M = 9.56$, $SD = 2.20$). Hier liegt der maximal zu erreichende Score bei 12 Punkten). Die Studierenden schätzen das Tool demnach als relevant, interessant und gut gestaltet ein.

Konstruktvalidität Differenziert das vorliegende Instrument zwischen den drei Phasen des Knowledge-based Reasoning: Beschreiben, Erklären und Vorhersagen?

Das nach der Pilotierung überarbeitete Tool wurde bei Studierenden der Universität Koblenz-Landau am Standort Landau mit dem Schwerpunkt Lehramt ($N = 239$) und bei Lehramtsanwärter:innen ($N = 41$) aus dem staatlichen Studienseminar Landau-Land eingesetzt. Im Rahmen des Validierungsprozesses sollte an der halbierten Stichprobe ($N = 140$) mittels der Raschskalierung in einem ersten Schritt überprüft werden, ob sich die angenommene Modellstruktur abbilden lässt. Entsprechend der Literatur wird der Teilprozess des Knowledge-based Reasonings in drei Phasen gegliedert. Diese drei-dimensionale Struktur weist im Vergleich zur ein- oder zweidimensionalen Struktur bessere AIC- und BIC-Werte auf (siehe Tabelle 43).

Tabelle 43 Modellvergleiche mittels Anova (B = Beschreiben, E = Erklären, V = Vorhersagen, Dim = Dimensionalität, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion)

Model	Dim	AIC	BIC
mod_B/E/V	1-dim	38369.81	39331.73
mod_B+E+V	3-dim	30476.58	31453.21
mod_B+E/V	2-dim	33290.94	34258.74
mod_B/E+V	2-dim	33293.90	34261.70
mod_B/V+E	2-dim	33294.05	34261.85
mod_B+E+V	3-dim	30476.58	31453.21

Hinsichtlich der drei-dimensionalen Struktur kann anhand des vorhandenen Datenmaterials von einer Konstruktvalidität ausgegangen werden. Dabei gilt es, zu berücksichtigen, dass die Ergebnisse nur vergleichend betrachtet werden können und keine absolute Aussage möglich ist.

Interne Konsistenz Bilden die Items die Skalen des Knowledge-based Reasonings reliabel ab?

Für die dreidimensionale Modellstruktur konnten folgende EAP-Reliabilitäten ermittelt werden: Beschreiben $R_{EAP} = 0.756$, Erklären $R_{EAP} = 0.735$, Vorhersagen $R_{EAP} = 0.718$. Diese Werte entsprechen für einen Fähigkeitstest einer guten Reliabilität (Field et al. 2012, S. 806).

Skalierung Zeigen die Items modellkonforme Itemfitwerte?

Die Items werden hinsichtlich ihrer Infit- und Outfit-Werte betrachtet. Die Outfitwerte sollten bei der vorhandenen Stichprobengröße zwischen $0.52 < \text{Outfit}_i < 1.48$ liegen (Smith et al. 1998, S. 78). Streng genommen sollten sie Werte zwischen $0.7 < \text{Outfit}_i < 1.3$ für eine gute Passung erreichen (u.a. Bond und Fox 2012, S. 310). Einschränkungen in der Skalierung kommen bei 11 Items beim Outfit vor, die sich im genannten Toleranzbereich $0.52 < \text{Outfit}_i < 1.48$ befinden. Ein Item erfüllt die Voraussetzungen nicht ($\text{Outfit}_i = 1.52$). Die Infitwerte liegen alle im guten bis sehr guten Bereich.

Konstruktvalidität Lassen sich die Ergebnisse an einer zweiten Stichprobe replizieren?

Zu Beginn muss deutlich gemacht werden, dass die Überprüfung der Ergebnisse der Raschskalierung an einer zweiten Stichprobe (N_2) mit einer Limitation einhergeht. Im Rahmen des iterativen Prozesses mussten einige Items aus dem Modell aufgrund schlechter Kennwerte extrahiert werden. Die Datenauswertung an der zuvor halbierten Stichprobe ($N_2 = 140$) konnte nicht an einer neuen Erhebung mit dem adaptierten Modell vorgenommen werden. Es handelt sich um eine Splittung der ursprünglichen Stichprobe ($N_{\text{gesamt}} = 280$). Somit erfolgt die Auswertung auf Basis der Ergebnisse der Raschskalierung und kann damit nur als Indiz für die Konstruktvalidität gesehen werden. Die Überprüfung der Modellwerte an der zweiten Stichprobe erreichen ähnliche, wenn auch leicht schlechtere Ergebnisse im Vergleich zur ersten Stichprobe (N_1) (siehe Tabelle 44).

Tabelle 44 Kennwerte entsprechend der Stichprobe

Kennwert	Stichprobe N ₁	Stichprobe N ₂
EAP-Reliabilität	Beschreiben $R_{EAP} = 0.756$	Beschreiben $R_{EAP} = 0.721$
	Erklären $R_{EAP} = 0.735$	Erklären $R_{EAP} = 0.714$
	Vorhersagen $R_{EAP} = 0.718$	Vorhersagen $R_{EAP} = 0.694$
WLE-Reliabilität	Beschreiben $R_{WLE} = 0.720$	Beschreiben $R_{WLE} = 0.704$
	Erklären $R_{WLE} = 0.656$	Erklären $R_{WLE} = 0.617$
	Vorhersagen $R_{WLE} = 0.681$	Vorhersagen $R_{WLE} = 0.681$
Standardized-Root-Mean-Residual	SRMR = 0.086	SRMR = 0.087
Lokale stochastische Unabhängigkeit Q3-Statistik	$M_{Q3} = -0.002$	$M_{Q3} = -0.004$
	$Ma_{Q3} = 0.000$	$Ma_{Q3} = 0.000$
Itemfitwerte	$M_{Infit} = 1.001$	$M_{Infit} = 1.007$
Infit und Outfit	$M_{Outfit} = 1.008$	$M_{Outfit} = 1.009$

Zusammenfassend kann gesagt werden, dass die Reliabilitäten für einen Leistungstest akzeptable Werte erreichen. Die Q3-Statistik ist im Mittelwert optimal bis gut. Der SRMR zeigt eine akzeptable Passung. Die Infit- und Outfitmittelwerte stehen ebenfalls für eine gute Passung. Ein Vergleich des Modells zwischen der ersten Stichprobe und der zweiten Stichprobe bestätigt die leicht schlechtere Passung hinsichtlich der informationstheoretischen Indizes AIC und BIC ($\Delta AIC = -79.55$, $\Delta BIC = -79.55$). Die Daten der ersten Stichprobe können das Modell besser erklären ($p = 0.000$).

Theoriegeleitete Konstruktvalidität Sind die angenommenen Schwierigkeitsgrade innerhalb des Reasonings mit dem Instrument überprüfbar?

Studierende an der Universität am Standort Landau und Lehramtsanwärter:innen aus dem Studienseminar ($N_{\text{gesamt}} = 280$) haben die in ViviAn eingebetteten Videos gesehen und die dazugehörigen Analyseaufträge beantwortet. Für die Auswertung wurden nur die Items berücksichtigt, die nach der Raschskalierung im Itempool verblieben sind. Zudem handelt es sich um die zur Raschskalierung verwendeten Stichproben (N_1 und N_2). Daher muss auch hier von einer Limitation des Resultates ausgegangen werden. Bei der Auswertung der Fragen wurde die aus dem Expert:innenrating generierte Musterlösung zu Grunde gelegt. Dabei wurde ein dreistufiges Punkteschema angewandt. Die Teilnehmenden haben bei Übereinstimmung mit den Antworten der Expert:innen zwei Punkte erhalten, bei der richtigen Tendenz einen Punkt (teilrichtige Antwort) und bei einer falschen Antwort null Punkte. Um die Forschungsfrage beantworten und einen Vergleich zwischen dem Beschreiben, Erklären und Vorhersagen ziehen zu können, mussten zunächst die prozentual erreichten Punkte der jeweiligen Dimension berechnet werden. Deskriptiv

lassen sich folgende Mittelwerte abbilden: Im Beschreiben werden die wenigsten Punkte erreicht ($M = 50.8$, $SD = 12.9$), gefolgt vom Vorhersagen ($M = 51.1$, $SD = 9.45$) und Erklären ($M = 53.1$, $SD = 9.28$). Die Varianzanalyse lässt auf signifikante Unterschiede ($F(2,492) = 5.02$; $p = 0.009$, $f = 0.018$) zwischen den drei Kategorien schließen. Es lassen sich adjustierte Unterschiede zwischen dem Beschreiben und Erklären ($p = 0.008$) sowie dem Erklären und Vorhersagen ($p = 0.010$), nicht jedoch zwischen dem Beschreiben und Vorhersagen ($p = 1$) feststellen. Die ursprünglich angenommenen Schwierigkeitsgrade mit abnehmender Punktzahl vom Beschreiben zum Erklären und zum Vorhersagen lassen sich nicht bestätigen.

gruppendifferenzierte Konstruktvalidität Kann mit ViviAn eine Entwicklung in der Ausprägung der Professionellen Unterrichtswahrnehmung zu Beginn des Studiums im Vergleich zum Ende des Studiums erfasst werden?

Zur Beantwortung der Forschungsfrage, ob die Professionelle Unterrichtswahrnehmung hinsichtlich der Motivierung im zeitlichen Zusammenhang mit dem Kompetenzzuwachs im Lauf des Studiums steht, wird mittels einer Anova der Unterschied zwischen den Gruppen überprüft. Zusätzlich zu den Studierenden, die entsprechend ihrem Studium nach Beginn, Mitte und Ende des Studiums eingeteilt wurden, sind die Lehramtsanwärterinnen als ergänzende Gruppe hinzugezogen worden ($N_{\text{gesamt}} = 280$). Die getrennte Analyse entsprechend der drei Dimensionen des Knowledge-based Reasonings zeigt auf, dass sowohl bei der Kategorie des Beschreibens ($F(1,270) = 1.988$, $p = 0.160$) als auch beim Erklären ($F(1,270) = 0.517$, $p = 0.473$) keine Unterschiede erfasst werden können, wohingegen beim Vorhersagen die Ergebnisse darauf hindeuten, dass zwischen den Gruppen signifikante Unterschiede vorhanden sind ($F(1,270) = 5.045$, $p = 0.026$). In der weiteren Analyse mittels paarweisem t-Test mit bonferroni-Adjustierung werden diese Unterschiede zwischen dem Anfang und dem Ende des Studiums aufgedeckt. Entsprechend dieser Ergebnisse ist das Aufzeigen von Unterschieden mit der vorgestellten ViviAn-Erweiterung innerhalb des Knowledge-based Reasonings mit der vorhandenen Stichprobe nur bedingt möglich.

Änderungssensitivität Erweist sich ViviAn als geeignetes Instrument, um eine Entwicklung der Professionellen Wahrnehmung hinsichtlich der Motivierung durch eine Intervention aufzuzeigen?

Die Studierenden am Standort Landau ($N=203$), die im Rahmen des Freien Workloads im Fachbereich der Erziehungswissenschaften das digitale Seminar zur Gestaltung von Lernumgebungen absolviert haben, haben an der Erhebung im Pre-Posttest-Design teilgenommen. Diese Testung beinhaltete die Anwendung des hier vorgestellten Abschnitts von Vivian mit den Vignetten und Analyseaufträgen zum Thema Motivierung im Unterricht. Zwischen dem Pre- und Posttest fand das Seminar statt (eine genaue Beschreibung

der Inhalte kann im Abschnitt 12.2 *Beschreibung des Seminars* nachgelesen werden). Die Analysen beziehen sich alle auf die Verwendung der Items entsprechend den Ergebnissen der Raschskalierung. Es lassen sich Unterschiede in allen drei Dimensionen des Knowledge-based Reasoning zwischen dem Pre- und Posttest feststellen (siehe Tabelle 45). Erwartungsgemäß kann innerhalb des Erklärens und Vorhersagens eine Zunahme der Professionellen Unterrichtswahrnehmung in Form einer höheren Punktzahl von der Pre- zur Posttestung verzeichnet werden. Beim Beschreiben ist ein gegenteiliger Effekt zu erkennen. Die Teilnehmenden haben im Posttest im Mittelwert statistisch geringere Ergebnisse erzielt als im Pretest.

Tabelle 45 Vergleich zwischen der Pre- und Posttestung mit Angabe des t-Test-Ergebnisses und der Effektstärke

Dimension	Pretest	Posttest	t-Test	Effektstärke
Beschreiben	$M_{Pre} = 24.18$ $SD_{Pre} = 6.18$	$M_{Post} = 21.98$ $SD_{Post} = 5.81$	$t(202) = 6.057$ $p = 1$	$d = -0.367$
Erklären	$M_{Pre} = 44.9$ $SD_{Pre} = 7.3$	$M_{Post} = 45.97$ $SD_{Post} = 6.78$	$t(202) = -2.648$ $p = 0.004$	$d = 0.152$
Vorhersagen	$M_{Pre} = 42.94$ $SD_{Pre} = 7.47$	$M_{Post} = 46.59$ $SD_{Post} = 8.07$	$t(202) = -8.310$ $p < 0.001$	$d = 0.469$

14 Diskussion

In diesem Kapitel sollen die Entwicklung des Videotools sowie die zuvor zusammengefassten Ergebnisse (Kapitel 13) analysiert, interpretiert und entsprechend der theoretischen Rahmung diskutiert werden. Dabei wird auf die Limitationen und Implikationen eingegangen.

Einleitung

Der Einsatz von Videos in der Lehrpersonenbildung bietet nachgewiesene Vorteile: die Veranschaulichung der Unterrichtsrealität, die Möglichkeit der Analyse komplexer Unterrichtssituationen sowie die Förderung der Theorie-Praxis-Relationierung im Studium (Bartel und Roth 2017, S. 43–44; Blomberg et al. 2013, S. 106; Luo et al. 2021, S. 1392–1393; van Es 2009, S. 101; Ertl-Schmuck 2019, S. 48–49). Dies führt zu einer Steigerung der Motivation der Studierenden, sich mit professionsrelevanten Aspekten vertieft auseinanderzusetzen (Luo et al. 2021, S. 1392–1393). Bereits im Studium können Videos verwendet werden, um Unterrichtssequenzen systematisch zu beobachten und zu beurteilen (Seidel und Prenzel 2007, S. 214). Es zeigt sich, dass die Einbettung von Videos in ein online-Tool hilfreich ist, die Professionelle Unterrichtswahrnehmung zu analysieren und zu fördern (Roth 2017; Seidel et al. 2010). Ausgehend von diesem Potenzial stellt sich die Frage, inwiefern das vorliegende videobasierte Instrument geeignet ist, die Professionelle Unterrichtswahrnehmung differenziert zu erfassen und zu fördern. Bisherige Videotools legen den Fokus auf Unterrichtsaspekte wie Klassenführung oder Lernunterstützung (Frommelt et al. 2016, S. 365; Gold und Hasselhorn 2013, S. 149; Hellermann et al. 2015, S. 104; Santagata und Guarino 2011, S. 142–143; Sunder et al. 2016, S. 9). Diese Arbeit hingegen ergänzt das Portfolio um Vignetten zum Thema Motivation und Motivierung. Das Themenfeld ist insbesondere wichtig, da die Motivation einen bedeutsamen Effekt auf die Lern- und Leistungsentwicklung der Schüler:innen hat (Deci und Ryan 1993; Schiefele und Streblow 2006; Spinath 2005). Dadurch ist es erforderlich, angehende Lehrpersonen darauf vorzubereiten, dass sie nicht nur ihr theoretisches Wissen zur Motivierung im Unterricht adäquat anzuwenden lernen, sondern auch geschult werden, wie sie die Motivation der Lernenden wahrnehmen können.

Entwicklung und Validierung der Videovignetten

Als theoretisches Rahmenmodell wurde bei der Erstellung der Videosequenzen das IBAS-Modell (Benning et al. 2019) zugrunde gelegt (siehe Abschnitt 2.2 Motivierung durch Zielorientierung im Lernkontext). Dieses bietet mit der Einteilung der vier Dimensionen unterrichtlichen Handelns - Inhalt, Bewertung, Autonomie und Soziales - mit den dazugehörigen Facetten einen sehr strukturierten, komplexen, aber auch übersichtlichen Rahmen. Die vorliegende Arbeit ist aufgrund der Fokussierung auf eine beschränkte Auswahl

an Facetten des motivationsförderlichen Handelns limitiert und fokussiert die Analyse auf sechs Aspekte: Leistungsdifferenzierung, individuelle Bezugsnormorientierung, Relevanz, adäquate Freiräume, adäquate Strukturierung, positiv-konstruktives Fehlerklima. Daher kann nicht von der Entwicklung und Validierung eines einzigen Instruments zur Analyse der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung gesprochen werden, sondern nur von mehreren motivationsrelevanten Aspekten.

Die Entwicklung und empirische Validierung geeigneter Videosequenzen stellt einen ersten grundlegenden Meilenstein der vorliegenden Forschungsarbeit dar. Mit der Erhebung und Auswertung der tatsächlichen, aktuellen Motivation der Schüler:innen in der Unterrichtseinheit (Krapp und Hascher 2014a, S. 234) können gezielt Videosequenzen ausgewählt werden, die von den Lernenden selbst als motivierend wahrgenommen wurden. Im Unterschied dazu, verwenden andere Forschungsarbeiten Expert:innenratings zur inhaltlichen Validierung der Vignetten (Oser et al. 2010; u.a. Meschede et al. 2015; Seidel et al. 2010; Star und Strickland 2008). Der Untersuchungsfokus dieser Arbeit liegt auf dem hoch-inferenten Konstrukt der Motivation (Büttner 2008, S. 288) und stellt damit die besondere Herausforderung dar, diesen intrinsischen Prozess der Schüler:innen von außen adäquat zu beobachten und zu bewerten (Florio-Hansen 2014, S. 57; Krapp und Hascher 2014a, S. 234). Um der empirischen Beurteilung gerecht zu werden, wird der alternative, an Individuen orientierte Zugang über eine Selbsteinschätzung gewählt. Das für die Erhebung ausgewählte Instrument - Aktuelle Lernmotivation von Prenzel et al. (1996) - hat sich als geeignet bewährt. Es erwies sich in der vorliegenden Studie beim Einsatz im Unterrichtskontext als praktikables und effizientes Instrument, mit dem Unterschiede in der Wahrnehmung der Motivation zwischen einzelnen Unterrichtssequenzen empirisch gesichert dargestellt werden konnten. Diese Forschungsergebnisse erweitern die bestehende Forschung um eine zusätzliche methodische Alternative der inhaltlichen Validierung von Videosequenzen. Nicht auszuschließen ist allerdings, dass sich ein zusätzliches Rating durch eine Expert:innengruppe förderlich für die Auswahl geeigneter Szenen erweisen könnte. Eine Kombination aus Befragung und Rating wäre für künftige Forschungen eine denkbare Vorgehensweise.

Die entwickelten Videovignetten erfassen eine Schulform, ein Unterrichtsfach, eine Lehrperson und jeweils eine Klasse der Jahrgangstufen sieben und acht einer Realschule plus. Damit sind die Ergebnisse der Validierung der Vignetten durch die aktuelle Motivation der Lernenden möglicherweise nicht generalisierbar. Die spezifische Zusammensetzung der Klasse, beispielsweise hinsichtlich sozialer Herkunft oder Lernvoraussetzungen, könnten ebenfalls die Ergebnisse beeinflusst haben. Zudem befanden sich die gefilmten Lernenden auf Grund der Altersstruktur vermutlich in einer Phase großer Veränderungen (Pubertät), die sich auf die Motivation, das Interesse und das Lernverhalten ausgewirkt haben können. So könnten beispielsweise hormonelle oder soziale Veränderungen in der Peer-Group Einfluss auf die Motivation genommen haben. Diese Annahmen decken sich auch mit der Beschreibung der Motivation als komplexem Prozess von

Rheinberg und Vollmeyer (2019, S. 261), der nicht nur von den Zielen der Lernenden, der persönlichen Eigenschaften wie der Selbstwahrnehmung und der eigenen Werte ausgeht, sondern auch den Einfluss neuro-hormoneller Strukturen auf die Motivation beschreibt. Zusätzliche Einblicke aus der Erhebung der Lern- und Leistungsmotivation der Schüler:innen mittels der SELMO-Befragung deuten darauf hin, dass beide Klassen im Gesamteindruck im Bereich der Lernzielorientierung unterhalb der Normstichprobe liegen. Da Lernzielorientierungen im engen Zusammenhang mit dem schulischen Engagement und dem Leistungserfolg stehen (Spinath et al. 2012, S. 52), kann möglicherweise bei der vorliegenden Stichprobe das schulische Interesse im Allgemeinen geringer ausgeprägt gewesen sein. Anhand der eingesetzten Fragebögen lässt sich zudem ein geringes Fachinteresse am Geografie-Unterricht beider Klassen im Mittelwert aufdecken. Dies könnte die Beurteilung der aktuellen Motivation gemindert haben. Trotz geringerer Ausprägung der Lernzielorientierung und des Fachinteresses lassen sich Unterschiede zwischen den einzelnen Messzeitpunkten aufzeigen. Es ist naheliegend, dass diese Unterschiede tatsächlich im Zusammenhang mit dem motivationsförderlichen oder -hemmenden Verhalten der Lehrperson zu diesen Zeitpunkten standen.

Auf Grund des hohen zeitlichen Aufwands der Erhebung im Unterrichtsalltag konnte für die Forschungsarbeit für den gesuchten Zeitraum nur eine Lehrperson akquiriert werden. Eine größere Bandbreite an Lehrpersonen hätte das Material attraktiver und umfassender gestalten können, indem unter anderem vielseitige erfolgreiche Motivationsstrategien identifiziert worden wären. Damit hätten zusätzliche und aussagekräftigere Erkenntnisse generiert werden können. Möglicherweise übt auch die Disposition der Lehrperson, wie von Blömeke et al. (2015) in ihrem Kompetenzmodell beschrieben, beispielsweise die Persönlichkeit, eigene Werthaltungen und Überzeugungen einen erheblichen Einfluss auf die Motivation der Lernenden aus. Infolge dieses Ansatzes wäre ein Sample aus unterschiedlichen Lehrpersonen wünschenswert gewesen. Die genannten Einschränkungen der Auswahl am Videomaterial könnten Auswirkungen auf die Motivation der Studierenden oder Lehramtsanwärter:innen gehabt haben, da es eventuell zu wenig Abwechslung bietet. Zukünftige Studien sollten verschiedene Schulformen, Fächer und Altersgruppen vergleichen, um ein umfassenderes Bild von den Einflussfaktoren auf die Lernmotivation zu erhalten und vielseitigeres Material von Videoszenen anbieten zu können.

Auf Basis der zuvor beschriebenen Auswahl der Unterrichtsszenen unter Berücksichtigung der Rahmenbedingungen des Settings und der Stichproben wurden anschließend die Videovignetten geschnitten. Zur kognitiven Entlastung wurde die Dauer der Videosequenz deutlich reduziert. Um dennoch das komplexe Unterrichts-geschehen im Video festhalten zu können, ermöglichen, wie Meschede und Steffensky (2018, S. 26) aufzeigen, mehrperspektivische Darstellungen, den gesamten Unterrichtsprozess weitgehend nachvollziehbar zu machen. Eine ganze Unterrichtssequenz kann aufgrund der Komplexität dennoch zur Überforderung führen (Syring et al. 2015, S. 680). Kurze

Videsequenzen bieten den Vorteil, das Material zu strukturieren und damit den Fokus auf ausgewählte Aspekte zu lenken (Meschede und Steffensky 2018, S. 26). Dennoch bilden sie das Geschehen authentisch ab (Riegel 2013, S. 13). Dies kann mit den Ergebnissen der Evaluation bestätigt werden (siehe Abschnitt 9.3 Ergebnisse der Pilotierungsstudie). In anderen Videostudien zeigte sich bereits, dass eine Länge von 1-4 Minuten geeignet ist, wie zum Beispiel beim OBSERVER (Seidel et al. 2010; Jahn 2014) oder bei weiteren Kolleg:innen (Gold et al. 2013; Hellermann et al. 2015; Meschede et al. 2015) (siehe Abschnitt 3.4.1 Forschungslage zur Erfassung und Förderung der Professionellen Unterrichtswahrnehmung).

In der vorliegenden Arbeit haben die Videosequenzen eine maximale Länge von zwei Minuten, was von der deutlichen Mehrheit der Studierenden als „nicht zu kurz“ gewertet wird. Vor diesem Hintergrund ist davon auszugehen, dass eine Überbelastung vermieden werden kann. Die weiteren Ergebnisse der Evaluation deuten darauf hin, dass das motivationsförderliche Handeln der Lehrperson und die Reaktion der Lernenden gebündelt dargestellt werden können. Vergleichsweise findet man diese Form der Reduktion als „segmenting“ (das Video in kleine Einheiten aufteilen) oder „weeding“ (das Weglassen nicht relevanter Inhalte) (Ibrahim et al. 2012, S. 220).

Eine alternative Methode, die Komplexität und die damit einhergehende kognitive Reizüberflutung zu reduzieren, könnte das „signaling“ darstellen. Mit Hilfe von Hervorhebungen kann der Aufmerksamkeitsfokus der Anwender:innen auf das Wesentliche gelenkt werden (Mayer und Moreno 2003, S. 48). Eine lernunterstützende Aufbereitung der Videos durch signaling oder segmenting wirkt sich positiv auf den Lernerfolg von Noviz:innen aus (Ibrahim et al. 2012, S. 230–231). In der vorliegenden Arbeit zeigte sich, dass die Proband:innen vor allem im Bereich des Beschreibens der Situation weniger Punkte als erwartet erreichten, daher könnte neben der Segmentierung das signaling beim Beschreiben der Situation unterstützend wirken. Dies wäre ein Ansatz für nachfolgende Forschungen mit dem Videotool ViviAn.

Gestaltung der Analyseaufträge

In dieser Forschungsarbeit wurden, wie in früheren Videostudien erfolgreich eingesetzt (Seidel et al. 2010; Star und Strickland 2008; Meschede et al. 2015), geschlossene Items verwendet, um die Datenerhebung zu strukturieren, ökonomisch zu gestalten und eine höhere Auswertungsobjektivität zu erreichen. Diese Vorgehensweise bringt jedoch einige Einschränkungen mit sich. Geschlossene Items bieten nur eine begrenzte Möglichkeit zur differenzierten Erfassung der Professionellen Unterrichtswahrnehmung (von Aufschneider et al. 2017, S. 93). Die vorgegebenen Antwortmöglichkeiten können dazu führen, dass komplexe Gedanken nicht adäquat abgebildet werden und lediglich das vorhandene Wissen abgerufen wird, statt eine Eigenleistung zu fördern. Die Antworten sind stark von den vorgegebenen Optionen abhängig. Die Verwendung geschlossener Items bietet in dieser

Forschungsarbeit zwar die oben genannten praktischen Vorteile, bringen aber auch Limitationen mit sich.

Wie in vielzähligen Forschungsarbeiten belegt, dient der Ansatz der Expert:innenratings neben der Validierung der Videovignetten auch der Validierung der Analyseaufträge sowie der Erstellung einer Musterlösung und trägt damit zu einer hohen Auswertungsobjektivität bei (Oser et al. 2010; u.a. Meschede et al. 2015; Seidel et al. 2010; Star und Strickland 2008). Dieses Verfahren wurde demnach auch für die vorliegende Auswertung der Analyseaufträge angewandt. Insgesamt kann eine zufriedenstellende bis gute Übereinstimmung der Meinungen der Expert:innen erreicht werden. Das Ergebnis kann sich damit mit anderen Forschungsarbeiten, zum Beispiel mit dem OBSERVER Videotool (Seidel et al. 2010, S. 303) und der Arbeit von Meschede et al. (2015, S. 325) messen. Hinsichtlich des Expert:innenratings zur Einigung und Überprüfung der richtigen Antworten muss beachtet werden, dass die Auswahl dieser Personen nicht zwangsläufig repräsentativ und hinsichtlich der Qualifikation richtig sein muss (Kubinger 2009, S. 56).

Einsatz, Potenzial und Grenzen des Videotools

Die entwickelten und validierten Videovignetten und Analyseaufträge sowie zusätzliche Informationen zur Klasse und Unterrichtseinheit wurden in das Video-Tool *ViviAn* („Videovignetten zur Analyse von Unterrichtsprozessen“, abrufbar über die Homepage <https://vivian-training.de>) implementiert. Damit wird den Studierenden entsprechend den Ergebnissen der Pilotierungsstudie zur Akzeptanz des Tools (Kapitel 9) eine interessante, relevante und gut gestaltete Lernumgebung geboten, mit der sie ihre Professionelle Unterrichtswahrnehmung hinsichtlich motivationsrelevanter Aspekte im Unterricht analysieren und fördern können. Dies steht im Einklang mit der bisherigen Evaluation des Tools *ViviAn* von Bartel und Roth (2020). Somit erhalten die Studierenden einen Lernort für „eine berufsspezifische kompetenzorientierte Ausbildung, die situativ eingebettet ist und die Komplexität des Unterrichtshandelns thematisiert“ (Oser et al. 2010, S. 24). Die Studierenden haben mit dem Tool die Möglichkeit, authentische und realitätsnahe Videoausschnitte aus dem Unterrichtsalltag zu sehen und zu analysieren.

Im Rahmen der Erhebung hatten die Teilnehmenden in der Lernumgebung *ViviAn* vor der Beantwortung der Analysefragen die Möglichkeit, jedes Video einmal zu betrachten. Dieses Vorgehen soll der realen Unterrichtssituation entsprechen, in der die Lehrperson in Echtzeit die relevanten Ereignisse wahrnehmen, beschreiben, erklären und Auswirkungen auf den weiteren Lehr-Lern-Prozess bestimmen muss. Wie von Aufschnaiter et al. (2017, S. 92) darlegen, soll das einmalige Analysieren der Videosequenz die tatsächlichen Fähigkeiten widerspiegeln. Die Entscheidung gegen das wiederholte Betrachten des Videoausschnitts wurde bewusst wegen dieses Arguments getroffen, auch wenn zum einen eine nachgewiesene kognitive Überforderung durch das Videoformat auftreten kann (Syring et al. 2015, S. 680), zum anderen Kontexterfahrungen mit der fremden

Klasse fehlen. Denn auch die fehlenden Informationen können die Analyse beeinflussen. Um diese Einschränkung abmildern zu können, wurden im Videotool Zusatzinformationen eingebettet. Dadurch können die Studierenden im Vorfeld unter anderem alle Materialien, Informationen zur Unterrichtseinheit oder die Zusammensetzung der Klasse ohne Zeitvorgabe einsehen. Dies scheint, wie bisherige Arbeiten aufgezeigt haben, eine sinnvolle Ergänzung darzustellen (Bartel und Roth 2017, S. 45; von Aufschneider et al. 2017, S. 95). Es muss jedoch berücksichtigt werden, dass dies das vollständige Erfahrungswissen der Lehrperson über ihre eigene Klasse nicht ersetzen kann.

Ein nächster bedeutsamer Schritt bestand in der Anwendung des Videotools ViviAn mit den validierten Vignetten und den Analysefragen (raschskalierten Items) im Rahmen eines Seminars mit Studierenden an der Universität sowie im Studienseminar für das Lehramt an Realschulen plus. Entsprechend der Forschungslage war anzunehmen, dass die Fähigkeit hinsichtlich der Professionellen Unterrichtswahrnehmung innerhalb der drei Teilprozesse des Knowledge-based Reasonings bei Noviz:innen abnimmt (Stürmer et al. 2015, S. 352; Schäfer und Seidel 2015, S. 50–51; Meschede et al. 2015, S. 328; Vgl. Frommelt et al. 2016, S. 369). Dementsprechend sollen die Personen im Beschreiben den höchsten Fähigkeitswert erreichen, den niedrigsten im Vorhersagen, und im mittleren Bereich liegt das Erklären. In der vorliegenden Arbeit erreichen die Studierenden und Lehramtsanwärter:innen allerdings im Beschreiben den niedrigsten Score. Dieses Ergebnis deckt sich nicht mit den gängigen Forschungsarbeiten zum Einsatz von Videotools zur Förderung der Professionellen Unterrichtswahrnehmung: Bei Frommelt et al. (2016, S. 369) zeigt sich wie in anderen OBSERVER-Studien (Schäfer und Seidel 2015, S. 50–51; Stürmer et al. 2015, S. 352), dass die höchsten Fähigkeitswerte im Beschreiben erreicht werden. Allerdings gibt es Gründe, die das vorliegende Phänomen dieser Arbeit erklären können.

- 1) Im Zuge der Expertiseforschung konnte aufgezeigt werden, dass es vor allem für Noviz:innen eine Herausforderung ist, Unterrichtssituationen genau wahrzunehmen und zu beschreiben (Seidel et al. 2010, S. 297–298). Der Unterrichtsalltag ist hochkomplex mit vielen gleichzeitigen, aber auch unvorhersehbaren Ereignissen (Oser et al. 2010, S. 6), was eine ständige Anpassung erfordert. Dies kann insbesondere bei unerfahrenen Lehrpersonen dazu führen, dass sie Situationen nicht präzise genug beschreiben können. Die vorliegenden Ergebnisse stützen diese Annahme. Die Fähigkeit zum wissensbasierten Beschreiben verbessert sich jedoch laut Forschungslage mit zunehmender Berufstätigkeit (Seidel und Prenzel 2007, S. 211–213). Da das Beschreiben die Grundlage für die weiteren Reasoning-Prozesse bildet, sollte es bereits im Studium stärker gefördert werden. Das unterstreicht die Bedeutung von videobasierten Trainings wie dem ViviAn-Tool.
- 2) Zusätzlich lag der Fokus der Intervention eher auf dem Erklären und Vorhersagen. Das Beschreiben von Situationen wurde weniger geübt und auf Grund des corona-

bedingten online-Formats auch nicht adäquat genug mit Feedback versehen. Diese Interpretation erfordert eine tiefergehende Analyse des Interventionsmaterials und gegebenenfalls Anpassung des Materials und Formats. Ziel der Arbeit war es primär, die Sensitivität des Instruments zu überprüfen. Ansatzpunkte für die Entwicklung einer geeigneten Intervention zur Förderung der Professionellen Unterrichtswahrnehmung könnten dennoch abgeleitet werden. Dabei wäre es sinnvoll, das Videotool selbst in die Intervention einzubinden und nicht nur als Analyseinstrument zu nutzen.

- 3) Zudem konnte in bisherigen Videoforschungen nachgewiesen werden, dass sich Noviz:innen leichter von unwesentlichen Faktoren ablenken lassen und damit den Fokus auf relevante Aspekte verlieren (Keskin et al. 2024; Sherin 2009). Auch hier können mit zunehmender Erfahrung Lehrpersonen Ablenkungen im Unterricht eher ignorieren oder sich schneller wieder fokussieren (Huang et al. 2023, S. 131). Daher brauchen Lehrpersonen eine gute Selektionskompetenz. Das Noticing als Wahrnehmung der relevanten Situationen für das Lehren und Lernen (u.a. Seidel et al. 2010, S. 297; Sherin 2009, S. 384) ist eine Fähigkeit, die eng mit dem Beschreiben von Situationen im Zusammenhang steht. Denn das Beschreiben ist unmittelbar auf die Wahrnehmung angewiesen und möglicherweise deshalb in der vorliegenden Untersuchung schwächer ausgeprägt.

Im Vergleich zum Beschreiben konnte gezeigt werden, dass den Studierenden das wissensbasierte Erklären von Ereignissen im Unterricht leichter fällt ($p = 0.008$). Eine Begründung dafür kann darin liegen, dass die Vermittlung von Theorien und Evidenzbasierung im Studium eine zentrale Rolle einnimmt (Bresges et al. 2019, S. 4). Studierende können demnach erlernte Konzepte, Modelle und Theorien auf Situationen anwenden und diese damit erklären. Im Vorhersagen über den weiteren Verlauf haben die Studierenden erwartungsgemäß entsprechend der Studienlage (Schäfer und Seidel 2015, S. 50–51; Stürmer et al. 2015, S. 352) einen geringeren Fähigkeitswert als im Erklären. Die Ergebnisse deuten darauf hin, dass beim Vorhersagen antizipierende und zukunftsgerichtete Kompetenzen notwendig sind, die wieder in einem stärkeren Zusammenhang mit der Erfahrung einer Person im Unterrichten stehen. Konsequenzen für den weiteren Unterrichtsverlauf abzuleiten, stellt scheinbar eine Herausforderung dar. Zwar stützen diese Befunde die theoretische Annahme, dass die Professionelle Unterrichtswahrnehmung von der Lehrerfahrung abhängig ist (Berliner 2001, 1987; Carter et al. 1988), explizit untersucht wurde dies mit der vorliegenden Arbeit nicht. Ein Vergleich zwischen Noviz:innen und Expert:innen konnte aber bereits in anderen Forschungen aufdecken, dass deutliche Unterschiede zwischen erfahrenen Lehrpersonen und Studierenden zu verzeichnen sind (Seidel und Prenzel 2007, S. 211–213). Davon abgeleitet war es denkbar, dass die Professionelle Unterrichtswahrnehmung bereits im Verlauf des Studiums zunimmt. Deswegen wurde dieser Frage nachgegangen. Hierbei konnte aufgezeigt werden, dass die Studierenden am Ende des Studiums im Vergleich mit den Studienanfänger:innen

ausschließlich in der Dimension des Vorhersagens höhere Fähigkeitswerte erreicht haben. Im Beschreiben und Erklären blieben die Studierenden auf dem gleichen Niveau. Dass die Erlangung der Fähigkeit systematischen Beurteilens einem Prozess unterliegt (Seidel und Prenzel 2007, S. 214), wird durch diese Befunde bestätigt. Weitere Untersuchungen, die klären, weshalb die Studierenden innerhalb des Studiums besser Vorhersagen für den weiteren Unterrichtsverlauf treffen können, müssen angeschlossen werden. Ein möglicher Grund kann in der zunehmenden Unterrichtserfahrung durch das Absolvieren von Praktika oder PES-Tätigkeiten und damit verbesserten situativen Antizipationskompetenz liegen.

Der gesamte Prozess der Professionellen Unterrichtswahrnehmung trägt entscheidend zum erfolgreichen, pädagogischen Handeln bei (Kersting et al. 2010; Schäfer und Seidel 2015; Seidel et al. 2010). In der Folge stellt sich die Frage, ob die Expertise nur durch die praktische Erfahrung im Unterrichten (Berliner) angeeignet werden kann oder ob Tools wie ViviAn zur Förderung der Professionellen Wahrnehmung im Studium beitragen können. Bisherige Forschungsarbeiten wie die von Gold und Hasselhorn (2013, S. 149) bestätigen die Eignung von Videotools für diesen Zweck. Bezogen auf die Zielsetzung dieser Arbeit, ein valides videobasiertes Instrument zu entwickeln, wurde ViviAn zusätzlich im Rahmen eines Seminars in einem Pre-Post-Design eingebettet. Es bestand die Annahme, dass durch den Einsatz einer Intervention, im vorliegenden Fall durch ein Seminar mit verschiedenen selbstgesteuerten Lernelementen der Wissensvermittlung, die Teilnehmenden sich in ihren Fähigkeiten im Bereich des Reasonings verbessern. Entscheidend für die Höhe des Lernzuwachses scheint die Dauer der Intervention zu sein. Frommelt et al. (2016, S. 368) konnten in ihrer Studie aufzeigen, dass bereits 15 Stunden Intervention ausreichen, um die Professionelle Unterrichtswahrnehmung mittels Videos zu fördern. Da für die vorliegende Erhebung aufgrund der Corona-Pandemie eine digital-asynchrone Veranstaltungsform gewählt wurde, ist nicht beurteilbar, wie intensiv und lange sich die Studierenden mit dem Thema beschäftigt haben. In einer Präsenzveranstaltung hätte nicht nur der zeitliche Aspekt kontrolliert, sondern der Lernprozess direkt beeinflusst und die Maßnahmen reflektiert werden können. Dennoch lassen die vorliegenden Ergebnisse zumindest beim Erklären und Vorhersagen Rückschlüsse darauf zu, dass ein Lernzuwachs auch unter den gegebenen Umständen erreicht und mit dem Videotool messbar gemacht werden konnte. Dement-sprechend kann gezeigt werden, dass eine gezielte Intervention hilfreich sein könnte, um die Professionelle Unterrichtswahrnehmung zu fördern. Interessanterweise konnte dieses Ziel nicht für die Kategorie des Beschreibens erreicht werden. Mögliche Gründe, warum hier eine gegenteilige Entwicklung zu verzeichnen war, können in der Fokussierung der Seminaufgaben auf das Erklären und Vorhersagen liegen oder in der vereinfachten Anforderung durch die Vorgabe des zu beobachtenden Handelns der Lehrperson in den Analyseaufträgen (siehe Abschnitt 12.5 Interpretation und Schlussfolgerungen). Zudem könnte sich das fehlende direkte Feedback zu den Beobachtungen negativ auf das Beschreiben ausgewirkt haben.

Dies würde sich mit den Annahmen von Enenkiel (2022, S. 303) decken. Auf Grund der Tatsache, dass sich der negative Befund für das Beschreiben ebenfalls im Abbilden der Schwierigkeitsgrade gezeigt hat, lassen sich eventuell Rückschlüsse auf die fehlende Eignung dieser Items in dieser Kategorie schließen. Eine Überarbeitung bzw. eine Ergänzung der Items ist erforderlich.

Eine wesentliche Limitation dieser Forschungsarbeit liegt in der Auswertung der Validierung des Videotools zur Erfassung der Professionellen Unterrichtswahrnehmung (siehe Kapitel 11) sowie der Förderung der Professionellen Unterrichtswahrnehmung (siehe Kapitel 12). Die Auswertung der Daten basiert auf den raschskalierten Ergebnissen anhand der gleichen Stichprobe. Diese Methodik wurde gewählt, um eine effiziente Analyse zu ermöglichen und erste Tendenzen innerhalb der erhobenen Daten aufzuzeigen. Dennoch bringt dieses Verfahren einige bedeutende Einschränkungen mit sich, die die Validität und Verallgemeinerbarkeit der Ergebnisse reduzieren. Es wäre sinnvoll gewesen, nach der Raschskalierung die weiteren Forschungsfragen an einer zusätzlichen Stichprobe zu beantworten. Dadurch wären eine systematische Überprüfung sowie Testung der Ergebnisse auf ihre Stabilität möglich gewesen. Kritisch zu betrachten ist zudem, dass die Ergebnisse dieser Studie an einer neu akquirierten Stichprobe hätten überprüft werden sollen. Eine solche Validierung wäre entscheidend gewesen, um sicherzustellen, dass die gefundenen Tendenzen nicht nur spezifisch für die ursprüngliche Stichprobe sind, sondern auch auf eine weitere übertragbar wären. Durch das Fehlen einer solchen Überprüfung bleibt unklar, inwieweit die Ergebnisse generalisierbar sind oder ob sie möglicherweise durch spezifische Merkmale der untersuchten Gruppe beeinflusst wurden. Die gewählte Vorgehensweise kann zwar erste Tendenzen aufzeigen, diese sind jedoch zwingend an einer Folgestudie zu überprüfen.

Das Potenzial des entwickelten Videotools zum Einsatz in der Lehrpersonenbildung bleibt dennoch bestehen. Die gezielte Förderung der Professionellen Unterrichtswahrnehmung der Studierenden hinsichtlich motivationsrelevanter Aspekte kann einen wesentlichen Beitrag zur Befriedigung der Grundbedürfnisse der Lernenden entsprechend der Selbstbestimmungstheorie nach Deci und Ryan (1993) nach Autonomie, Kompetenzerleben und sozialer Eingebundenheit im Unterricht leisten. Bereits in der Lehramtsausbildung sollte verstärkt der Fokus auf die systematische Analyse von Unterrichtssituationen gelegt werden, um nachhaltig die Motivation der Schüler:innen positiv beeinflussen zu können. Basierend auf dem Konzept der Professionellen Unterrichtswahrnehmung (Sherin und van Es 2008; Seidel et al. 2010) wird deutlich, dass diese relevant für ein erfolgreiches Handeln im Unterricht ist. In der Lehrpersonenbildung kann durch den Einsatz eines videobasierten Tools gezielt an der Förderung des wissensbasierten Beschreibens, Erklärens und Vorhersagens angesetzt werden. Wie sich in der Literatur (Frommelt et al. 2016, S. 368; Sunder et al. 2016, S. 9; Gold und Hasselhorn 2013, S. 149) sowie in dieser Dissertation gezeigt hat, eignen sich solche Tools, um die Professionelle

Unterrichtswahrnehmung analysieren und fördern zu können. Es wäre sinnvoll, videobasierte Reflexionen als festen Bestandteil in der Lehrpersonenbildung zu implementieren.

Wie Blömeke et al. (2015) in ihrem PID-Modell beschreiben, setzt sich die Kompetenz der Lehrperson aus der Disposition, den situationsspezifischen Fähigkeiten und der tatsächlichen Performanz zusammen. Die Chance, im Studium Lerngelegenheiten zu schaffen, in denen das professionelle Wissen, z.B. zu motivationsförderlichem Handeln, in authentischen Unterrichtssituationen, z.B. durch Videovignetten, beschrieben, erklärt und kritisch reflektiert werden kann, um Entscheidungen für das konkrete Handeln ableiten zu können, wäre sinnvoll. Dadurch kann zudem eine Relationierung von Theorie und Praxis zur Förderung einer professionellen Handlungskompetenz angebahnt werden.

Methodische Reflexion: Raschskalierung und Grenzen der Validierung

Der Fokus der methodischen Reflexion liegt auf der Interpretation der Auswertung des raschskalierten Modells.

Das vorliegende Instrument deutet auf die erwartete Konstruktvalidität eines dreidimensionalen Modells hin, konsistent mit früheren Studien zur Professionellen Unterrichtswahrnehmung (u.a. Seidel und Stürmer 2014, S. 759). Damit kann die Eindimensionalität von Meschede et al. (2015, S. 326), wie in ihrer Arbeit formuliert, ebenfalls nicht bestätigt werden. Unter den gegebenen Bedingungen ermöglicht ViviAn eine Differenzierung hinsichtlich der Professionellen Unterrichtswahrnehmung zwischen den drei Dimensionen Beschreiben, Erklären und Vorhersagen im Rahmen des Knowledge-based Reasonings. Damit ist es wie der OBSERVER - als einer der bekanntesten und wiederholt evaluierten Vertreter eines videobasierten Instruments zur Förderung der Professionellen Unterrichtswahrnehmung - in der Lage, zwischen den drei Dimensionen zu unterscheiden (Seidel und Stürmer 2014, S. 759; Jahn 2014, S. 176). Es ist jedoch wichtig zu betonen, dass eine absolute Aussage über die Struktur des Modells nicht getroffen werden kann, vielmehr handelt es sich um ein vergleichendes Urteil (Bühner 2006, S. 352). Damit kann sich die vorliegende Arbeit den bisherigen Erkenntnissen anschließen.

Die EAP-Reliabilitäten weisen für einen Fähigkeitstest mit Werten $\alpha > 0.7$ auf eine ausreichend gute interne Konsistenz hin (Field et al. 2012, S. 806). Damit schließen die Ergebnisse an andere Videostudien an (Stürmer und Seidel 2015, S. 60). Auf Grund dieser Reliabilität kann das vorliegende Instrument für die Analyse der Ausprägung der Fähigkeit zur Professionellen Unterrichtswahrnehmung verwendet werden.

Insgesamt wurden 162 Items in einem iterativen Prozess analysiert. Sukzessive mussten Items ausgeschlossen werden, die den Standards nicht entsprechen. Im abschließenden Modell weisen die noch verbleibenden 104 Items eine geordnete Struktur hinsichtlich der Schwellenparameter auf. Dies bedeutet, dass die erwartete Reihenfolge der Schwellen eingehalten wird und sie voneinander trennbar sind. Es kann davon ausgegangen werden, dass mit Zunahme der Kategorie das latente Merkmal, also die Fähigkeit der

Professionellen Unterrichtswahrnehmung, stärker ausgeprägt ist. Damit kann diese Modellannahme bestätigt werden.

Die punktbiseriale Korrelation als ergänzender Fit-Indikator kann als weiteres Kriterium der Trennschärfe gesehen werden und sollte ebenfalls aufsteigend sein (Wu et al. 2016, S. 85). Das finale Modell zeigt jedoch bei 16 Items keinen Anstieg über alle drei Kategorien, was auf eine unzureichende Differenzierung zwischen Personen mit unterschiedlichen Fähigkeiten hinweist. Diese Items tragen möglicherweise nicht zur Trennschärfe bei. Das könnte damit im Zusammenhang stehen, dass die mittlere Kategorie beim vorliegenden Partial-Credit-Modell bereits die richtige Tendenz aufzeigt und die Kategoriendifferenzierung zur obersten Kategorie zu gering ist. Eine Überprüfung der Reliabilität durch den Ausschluss dieser Items ergab, dass die Items zur internen Konsistenz beitragen. Daher werden diese Items sowohl wegen möglicher Kompensationseffekte als auch aus konzeptuellen Gründen (unterrepräsentative Anzahl an Items pro Kategorie) trotz der anzunehmenden Trennschärfeprobleme beibehalten.

Bei vier Itempaaren zeigt sich im finalen Modell eine hohe stochastische Abhängigkeit. Diese Items hängen hinsichtlich des Antwortverhaltens stark zusammen, womöglich weil sich die Inhalte sehr ähneln oder logisch aufeinander aufbauen. Folglich kann die Reliabilität überschätzt werden. Die Überprüfung der lokalen stochastischen Unabhängigkeit zeigt im Mittelwert gute bis optimale Ergebnisse. Daher werden die vier Itempaare beibehalten (ähnlich wie in anderen Studien durch die Metaanalyse von Christensen et al. (2017, S. 178) aufgezeigt).

Die Ergebnisse aus der Skalierung können folgendermaßen interpretiert werden. Alle Items weisen gute bis sehr gute Infitwerte auf. Die Items zeigen kein auffällig abweichendes Antwortverhalten. Das bedeutet, dass sie zwischen den Personenfähigkeiten differenzieren können. Es kann davon ausgegangen werden, dass die Daten gut zum theoretischen Modell passen. Die Kriteriumkennwerte werden in der Literatur unterschiedlich streng festgelegt. Grundsätzlich wird in der vorliegenden Arbeit bei der Beurteilung der Fitwerte die strenge Variante bei der Auswertung berücksichtigt. Abgesehen von zwölf Items erfüllen alle diese Outfit-Richtwerte von $0.7 < \text{Outfit}_i < 1.3$ (u.a. Bond und Fox 2012, S. 310). Bei elf Items wird eine auf Stichprobengröße wissenschaftlich angepasste Annahme von $0.52 < \text{Outfit}_i < 1.48$ (Smith et al. 1998, S. 78) als Toleranzbereich definiert, also als akzeptabel interpretiert. Bei einem Item sind die Outfit-Werte außerhalb des Toleranzbereichs (Outfit = 1.52). Es unterschätzt möglicherweise die Vorhersagekraft und sollte noch einmal genau überprüft werden. Vorerst wird dieses Item für das vollständige Abbilden der Itemkategorie beibehalten. Insgesamt bleibt die Modellpassung trotz der Limitation als gut zu bewerten.

In Bezug auf die Validität ist die Entscheidung, Items mit weniger gut erfüllter Skalierung beizubehalten, kritisch zu betrachten. Es wurden einige Items aus inhaltlichen Überlegungen nicht ausgeschlossen, was dazu führt, dass einige Kennwerte der Raschskalierung nur bedingt erfüllt sind. Diese Vorgehensweise beeinträchtigt die Qualität der

Ergebnisse, da diese Items nicht den strengen Anforderungen an die Modellkonformität genügen. Es fand ein Abwägen zwischen der Beibehaltung dieser Items statt, um wertvolle Einsichten zu bekommen und die Dimensionen gut abbilden zu können. Aber andererseits besteht das Risiko, dass diese Items aufgrund ihrer unzureichenden Skalierbarkeit zu eingeschränkten Ergebnissen führen.

Zusammenfassend können die Ergebnisse der Raschskalierung als zufriedenstellend beurteilt werden. Die Reliabilitäten liegen im guten Bereich für einen Leistungstest. Damit scheint die interne Konsistenz gegeben zu sein. Die Trennschärfe ist weitestgehend angemessen sowie die stochastische Unabhängigkeit. Die Fit-Werte sind für nahezu alle Items im definierten Bereich und bestätigen damit die Modellannahme. Dennoch muss kritisch angemerkt werden, dass insgesamt zu wenige Items nach der Raschskalierung pro Itemkategorie erhalten bleiben konnten. Dies führt dazu, dass einige relevante Aspekte möglicherweise nicht ausreichend abgedeckt werden. Eine unzureichende Anzahl an Items kann die Validität und Reliabilität der Ergebnisse beeinträchtigen und somit die Aussagekraft der gesamten Studie einschränken. Eine Ergänzung mit zusätzlichen Items und erneuter Raschskalierung wäre sinnvoll.

Abschließend wird die Übertragbarkeit der Ergebnisse auf eine zweite Stichprobe interpretiert. Der zweite Teil der zu Beginn halbierten Stichprobe wird dafür genutzt. Die AIC- und BIC-Werte dienen hierbei als zentrales Kriterium. Diese Indizes fallen bei der zweiten Stichprobe geringfügig schlechter aus ($\Delta AIC = -79.55$, $\Delta BIC = -79.55$). Dabei wird die Differenz als moderat eingestuft. Der Unterschied zwischen beiden Modellen ist so gering, dass eine vergleichbare Erklärungskraft vorliegt. Eine Übertragung auf eine andere Stichprobe aus dem gleichen Setting kann grundlegend angenommen werden, bleibt jedoch mit Einschränkungen verbunden. Die Passung des Modells mit der ersten Stichprobe ist entsprechend der Anova signifikant besser. Dieses Ergebnis deutet darauf hin, dass das erste Modell eindeutig die Datenstruktur besser darstellt. Es kann als logische Konsequenz betrachtet werden, da das Modell auf Basis der ersten Stichprobe geschätzt und optimiert wurde. Ein limitierender Faktor bleibt die Anwendung des Split-Half Verfahrens. Diese Methodik kann zwar nützlich sein, um die Reliabilität der Ergebnisse zu überprüfen, hat jedoch auch signifikante Auswirkungen auf die Stichprobengröße und damit auf die Validität der Raschskalierung. Durch die Teilung der Stichprobe wird die Anzahl der Daten deutlich reduziert, was potenziell zu einer geringeren statistischen Power führt. Eine kleinere Stichprobengröße kann dazu beitragen, dass die Schätzungen der Parameter weniger stabil sind. Infolgedessen können die aus der Raschskalierung abgeleiteten Ergebnisse lediglich Tendenzen widerspiegeln, anstatt robuste und verlässliche Schlussfolgerungen zu ermöglichen. Dieser Limitation kann entgegengesetzt werden, dass die zweite Hälfte der Stichprobe wertvolle Informationen zur Passung des Datenmodells liefern kann. Es konnte festgestellt werden, dass das gewählte Modell den Daten tatsächlich gerecht wird und die Tendenzen damit validieren. Dennoch sollte eine Forschung

darauf abzielen, ausreichend große und repräsentative Stichproben zu verwenden und systematisch vorzugehen.

Im Rahmen dieser Forschungsarbeit konnte nur ein Teil der Gütekriterien abgedeckt werden und dies zum Teil mit Einschränkungen. Für die Aussagekraft der Ergebnisse wäre es zudem sinnvoll gewesen, ein zweites Instrument zur Messung desselben Konstrukts heranzuziehen und mittels Korrelation auf Übereinstimmung zu testen. Zusätzliche Erhebungen im Rahmen der Teilnehmendenbefragung, wie die Erfassung des pädagogisch-psychologischen Wissens, hätten positiv zur Kriteriumsvalidität beitragen können. Zudem muss kritisch hinterfragt werden, ob bei den Teilprozessen des Reasonings mit der wissensbasierten Beschreibung, wissensbasierten Erklärung und wissensbasierten Vorhersage nicht das Wissen einen essenziellen Anteil des Projektes hätte darstellen müssen. Für folgende Studien wäre zudem eine Analyse des Wissenszuwachses nach der Anwendung des Tools ein lohnendes Untersuchungsfeld.

In einem Längsschnittdesign mittels zusätzlicher, nachgeschalteter Testung hätte mit einem weiteren Kriterium die prädiktive Validität überprüft werden können (Döring und Bortz 2016, S. 447). Damit hätte der Frage, inwieweit die Ergebnisse des Tests, in diesem Fall die Ausprägung der Professionellen Unterrichtswahrnehmung hinsichtlich der Motivierung, tatsächlich Einfluss auf die Unterrichtsqualität einer künftigen Lehrperson hat, nachgegangen werden können. Eine solche Korrelation könnte die Relevanz für Studierende und Lehramtsanwärter:innen unterstreichen. In einer Folgestudie könnte überprüft werden, ob die Person mit einer hohen Professionellen Unterrichtswahrnehmung im Bereich der Motivierung auch tatsächlich von den Lernenden als motivierender beurteilt wird. Alternativ kann die intrinsische Motivation der Lernenden herangezogen werden.

15 Perspektiven

Das hier vorgestellte und weiterentwickelte Videotool ViviAn um den Themenkomplex motivationsrelevanter Aspekte im Unterricht ermöglicht eine differenzierte Erfassung der drei Teilbereiche des Knowledge-based Reasonings mit dem Beschreiben, Erklären und Vorhersagen. Es ist sinnvoll, bereits im Studium Beschreibungsfähigkeiten durch gezielte Beobachtungsaufgaben zu fördern. Insbesondere, da die Studierenden, wie in der vorliegenden Arbeit aufgezeigt, den geringsten mittleren Fähigkeitsscore in dieser Kategorie aufweisen. Neben dem Beschreiben spielen die wissensbasierte Erklärung von relevanten Unterrichtssituationen sowie die Beurteilung der Auswirkungen auf den weiteren Lehr-Lernverlauf eine wichtige Rolle für ein erfolgreiches Unterrichten und Lernen. Auch hierfür stellt das Videotool einen Ansatz dar, diese Fähigkeiten bereits in der ersten Phase der Lehrpersonenbildung zu fördern. Die Arbeit bietet damit einen relevanten Beitrag zur Bildungsforschung hinsichtlich der Professionellen Unterrichtswahrnehmung im Zusammenhang mit motivationsrelevanten Aspekten. Sie bestärkt Hinweise für gelingende konzeptuelle Lehr-Lernformate, die eine Theorie-Praxis-Relationierung ermöglichen. Das Ziel eines solchen Projektes ist die Stärkung einer evidenzbasierten Lehrpersonenbildung durch datenbasierte Rückmeldung und Diagnostik. Es zeigt sich, dass ViviAn als reflexive Lernumgebung genutzt werden kann. Durch den Einsatz der Vignetten im Studium kann perspektivisch mit gemeinsamen Feedback- und Reflexionszeiten die Professionellen Unterrichtswahrnehmung in Bezug auf allgemein didaktische Merkmale unter Einbezug des pädagogisch-psychologischen Wissens gefördert werden. Dies kann durch eine bewusste Auseinandersetzung mit der Analyse von Videos erreicht werden. Dabei sollte ViviAn so im Studium implementiert werden, dass die Vignetten pausiert und die einzelnen Mikroszenen gemeinsam dreiteilig strukturiert analysiert (Beschreiben, Erklären, Vorhersagen) werden können. Neben der Lehre kann es als Fortbildungs- und Assessment-Tool eingesetzt werden. ViviAn ist so gestaltet, dass es geringe Zugangshürden hat, leicht anzuwenden ist und einen hohen Transferwert besitzt.

Die Herangehensweise zur Entwicklung und Validierung von ViviAn zeigt große methodische Übereinstimmungen an bisherige Videoforschungen. Eine Übertragung der motivationsrelevanten Aspekte auf andere Fächer ist möglich und erwünscht. Motivation und Motivierung ist ein fächerübergreifendes Thema und in allen Disziplinen und Schulformen relevant, bis in die Erwachsenenbildung hinein. Die Methode der inhaltlichen Validierung der Vignetten über Schüler:innenurteile, stellt einen innovativen Ansatz dar und wird dem Konstrukt der Motivation gerecht. Hier bieten das Vorgehen und die Ergebnisse interessante Anknüpfungspunkte für weitere Forschungen in diesem Bereich.

Literaturverzeichnis

Adams, Raymond J.; Wu, Margaret L.; Wilson, Mark (2012): The Rasch Rating Model and the Disordered Threshold Controversy. In: *Educational and Psychological Measurement* 72 (4), S. 547–573. DOI: 10.1177/0013164411432166.

Alboukadel, Kassambara (2023): rstatix. Pipe-Friendly Framework for Basic Statistical Tests. Version R package version 0.7.2.

Altmann, Anna F.; Kändler, Celia (2015): Videobasierte Instrumente zur Testung und videobasierte Trainings zur Förderung von Kompetenzen bei Lehrkräften. In: Timo Leuders, Matthias Nückles, Silke Mikelskis-Seifert und Kathleen Philipp (Hg.): *Pädagogische Professionalität in Mathematik und Naturwissenschaften*, Bd. 27. Wiesbaden: Springer Fachmedien, S. 39–67.

Ames, Carole (1992a): Achievement Goals and the Classroom Motivational Climate. In: Dale H. Schunk und Judith L. Meece (Hg.): *Student Perceptions in the Classroom*. Hillsdale, N.J.: L. Erlbaum, S. 327–348.

Ames, Carole (1992b): Classrooms: Goals, Structures, and Student Motivation. In: *Journal of Educational Psychology* 84, S. 261–271.

Artelt, Cordula; Gräsel, Cornelia (2009): Diagnostische Kompetenz von Lehrkräften. In: *Zeitschrift für Pädagogische Psychologie* 23 (34), S. 157–160. DOI: 10.1024/1010-0652.23.34.157.

Assor, Avi; Kaplan, Haya; Roth, Guy (2002): Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. In: *The British Journal of Educational Psychology* 72 (2), S. 261–278. DOI: 10.1348/000709902158883.

Bandura, Albert (1997): *Self-efficacy. The exercise of control*. New York: W.H. Freeman and Company.

Bartel, Marie-Elene; Roth, Jürgen (2017): Diagnostische Kompetenz von Lehramtsstudierenden fördern. In: Juliane Leuders, Timo Leuders, Susanne Prediger und Silke Ruwisch (Hg.): *Mit Heterogenität im Mathematikunterricht umgehen lernen. Konzepte und Perspektiven für eine zentrale Anforderung an die Lehrerbildung*. Wiesbaden: Springer Spektrum (Konzepte und Studien zur Hochschuldidaktik und Lehrerbildung Mathematik), S. 43–52.

Bartel, Marie-Elene; Roth, Jürgen (2020): Video- und Transkriptvignetten aus dem Lehr-Lern-Labor – die Wahrnehmung von Studierenden. In: Burkhard Priemer und Jürgen Roth (Hg.): *Lehr-Lern-Labore*, Bd. 223. Berlin, Heidelberg: Springer, S. 299–315.

- Baumert, Jürgen; Kunter, Mareike (2011): Das Kompetenzmodell von COACTIV. In: Mareike Kunter, Jürgen Baumert, Werner Blum, Uta Klusmann, Stefan Krauss und Michael Neubrand (Hg.): Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV. Münster [u.a.]: Waxmann, S. 29–53.
- Benjamini, Yoav; Hochberg, Yosef (1995): Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 57 (1), S. 289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.
- Benning, Katharina; Praetorius, Anna-Katharina; Janke, Stefan; Dickhäuser, Oliver; Dresel, Markus (2019): Das Lernen als Ziel: Zur unterrichtlichen Umsetzung einer Lernzielstruktur. In: *Unterrichtswissenschaft* 47 (4), S. 523–545. DOI: 10.1007/s42010-019-00054-7.
- Beret, Ann-Kathrin; Lengnink, Katja; von Aufschnaiter, Claudia (2017): Diagnostische Kompetenz gezielt fördern – Videoeinsatz im Lehramtsstudium Mathematik und Physik. In: Christoph Selter, Stephan Hußmann, Corinna Höble, Christine Knipping, Katja Lengnink und Julia Michaeli (Hg.): Diagnose und Förderung heterogener Lerngruppen. Theorien, Konzepte und Beispiele aus der MINT-Lehrerbildung. 1. Auflage. Münster: Waxmann, S. 149–168.
- Berliner, David C. (1987): Der Experte im Lehrerberuf: Forschungsstrategien und Ergebnisse. In: *Unterrichtswissenschaft* (3), S. 295–305.
- Berliner, David C. (2001): Learning about and learning from expert teachers. In: *International Journal of Educational Research* 35 (5), S. 463–482. DOI: 10.1016/S0883-0355(02)00004-6.
- Berliner, David C. (2004): Describing the behavior and documenting the accomplishments of expert teachers. In: *Bulletin of Science Technology & Society* 25 (3), S. 1–13.
- Bieg, Sonja; Mittag, Waldemar (2011): Leistungsverbesserungen durch Förderung der selbstbestimmten Lernmotivation. In: Markus Dresel und Lena Lämmle (Hg.): Motivation, Selbstregulation und Leistungsexzellenz. Berlin [u.a.]: Lit (Talentförderung, Expertiseentwicklung, Leistungsexzellenz, 9), S. 219–236.
- Bischoff, Sonja; Brühwiler, Christian; Baer, Matthias (2005): Videotest zur Erfassung «adaptiver Lehrkompetenz». In: *Beiträge zur Lehrerinnen- und Lehrerbildung* 23 (3), S. 382–397.
- Blomberg, Geraldine; Renkl, Alexander; Gamoran Sherin, Miriam; Borko, Hilda; Seidel, Tina (2013): Five research-based heuristics for using video in pre-service teacher education. In: *Journal for Educational Research Online* 5 (1), S. 90–114.

- Blömeke, Sigrid; Gustafsson, Jan-Eric; Shavelson, Richard J. (2015): Beyond Dichotomies. Competence Viewed as a Continuum. In: *Zeitschrift für Psychologie* 223 (1), S. 3–13. DOI: 10.1027/2151-2604/a000194.
- Bond, Trevor G.; Fox, Christine M. (2012): Applying the Rasch model. Fundamental measurement in the human sciences. 2. Aufl. New York: Routledge.
- Bortz, Jürgen; Schuster, Christof (2010): Statistik für Human- und Sozialwissenschaftler. Berlin, Heidelberg: Springer.
- Bresges, André; Harring, Marius; Kauertz, Alexander; Nordmeier, Volkhard; Parchmann, Ilka (2019): Die Theorie-Praxis-Verzahnung in der Lehrerbildung – eine Einführung in die Thematik. In: Bundesministerium für Bildung und Forschung (BMBF) (Hg.): Verzahnung von Theorie und Praxis im Lehramtsstudium. Erkenntnisse aus Projekten der „Qualitätsoffensive Lehrerbildung“. Frankfurt a. M.: Druck- und Verlagshaus Zarbock, S. 4–7.
- Brünken, Roland; Münzer, Stefan; Spinath, Birgit (2019): Pädagogische Psychologie - Lernen und Lehren. Göttingen: Hogrefe (Bachelorstudium Psychologie).
- Brunner, Martin; Anders, Yvonne; Hachfeld, Axinja; Krauss, Stefan (2011): Diagnostische Fähigkeiten von Mathematiklehrkräften. In: Mareike Kunter, Jürgen Baumert, Werner Blum, Uta Klusmann, Stefan Krauss und Michael Neubrand (Hg.): Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV. Münster [u.a.]: Waxmann, S. 215–234.
- Brunner, Martin; Kunter, Mareike; Krauss, Stefan; Baumert, Jürgen; Blum, Werner; Dubberke, Tamar et al. (2006): Welche Zusammenhänge bestehen zwischen dem fachspezifischen Professionswissen von Mathematiklehrkräften und ihrer Ausbildung sowie beruflichen Fortbildung? In: *Zeitschrift für Erziehungswissenschaft* 9 (4), S. 521–544. DOI: 10.1007/s11618-006-0166-1.
- Bühner, Markus (2006): Einführung in die Test- und Fragebogenkonstruktion. 2., aktualisierte Auflage. München [u.a.]: Pearson Studium (PS Psychologie).
- Bühner, Markus (2011): Einführung in die Test- und Fragebogenkonstruktion. 3., aktualisierte und erw. Auflage. München, Boston [u.a.]: Pearson Studium (Psychologie).
- Bühner, Markus (2021): Einführung in die Test- und Fragebogenkonstruktion. 4., korrigierte und erw. Auflage. München: Pearson Studium (PS Psychologie).
- Bürger, Kathrin (2013): Multiple State- und Trait-Ziele im Kontext Unterricht. Eine interaktionistische Perspektive.
- Büttner, Gerhard (2008): Fragebögen und Ratingskalen. In: Wolfgang Schneider und Marcus Hasselhorn (Hg.): Handbuch der pädagogischen Psychologie. Göttingen [u.a.]: Hogrefe (Handbuch der Psychologie, 10), S. 282–290.

- Carlson, Robert E. (1990): Assessing teachers' pedagogical content knowledge: Item development issues. In: *Journal of Personnel Evaluation in Education* 4 (2), S. 157–163. DOI: 10.1007/BF00126124.
- Carter, Kathy; Cushing, Katherine; Sabers, Donna; Stein, Pamela; Berliner, David (1988): Expert-Novice Differences in Perceiving and Processing Visual Classroom Information. In: *Journal of Teacher Education* 39 (3), S. 25–31. DOI: 10.1177/002248718803900306.
- Cerasoli, Christopher P.; Nicklin, Jessica M.; Ford, Michael T. (2014): Intrinsic motivation and extrinsic incentives jointly predict performance: a 40-year meta-analysis. In: *Psychological Bulletin* 140 (4), S. 980–1008. DOI: 10.1037/a0035661.
- Christensen, Karl Bang; Makransky, Guido; Horton, Mike (2017): Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. In: *Applied Psychological Measurement* 41 (3), S. 178–194. DOI: 10.1177/0146621616677520.
- Cohen, Jacob (1977): *Statistical power analysis for the behavioral sciences*. Rev. ed. Hillsdale, NJ: Erlbaum.
- Cohen, Jacob (1988): *Statistical power analysis for the behavioral sciences*. 2. Auflage. New York: Lawrence Erlbaum Associates.
- Deci, Edward L.; Ryan, Richard M. (1993): Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. In: *Zeitschrift für Pädagogik* 39 (2), S. 223–238, zuletzt geprüft am 12.07.2016.
- Digel, Sabine (2013): Netzgestützte Videofallarbeit. Ein didaktisches Konzept zur Kompetenzentwicklung von Lehrenden. In: *Magazin erwachsenenbildung.at* 20, S. 1–9.
- Digel, Sabine; Baust, Christina; Schrader, Josef (2014): Videofallbasiertes Lernen. Ein innovatives Konzept zur Weiterbildung von Hochschullehrenden. In: *Hochschule und Weiterbildung* 2, S. 40–45.
- Döring, Nicola; Bortz, Jürgen (2016): *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Unter Mitarbeit von Sandra Pöschl-Günther. 5. vollständig überarbeitete, aktualisierte und erweiterte Auflage. Berlin, Heidelberg: Springer (Springer-Lehrbuch). Online verfügbar unter <http://dx.doi.org/10.1007/978-3-642-41089-5>.
- Eid, Michael; Gollwitzer, Mario; Schmitt, Manfred (2017): *Statistik und Forschungsmethoden*. 5., korrigierte Auflage. Weinheim, Basel: Beltz.
- Eid, Michael; Schmidt, Katharina (2014): *Testtheorie und Testkonstruktion*. Göttingen [u.a.]: Hogrefe (Bachelorstudium Psychologie).

- Enenkiel, Patrizia (2022): Diagnostische Fähigkeiten mit Videovignetten und Feedback fördern. Gruppenarbeitsprozesse zur Bestimmung von Längen, Flächen- und Rauminhalten. Wiesbaden: Springer Spektrum.
- Enenkiel, Patrizia; Roth, Jürgen (2019): Der Einfluss von Feedback auf die Entwicklung diagnostischer Fähigkeiten von Mathematiklehramtsstudierenden. In: Andreas Frank, Stefan Krauss und Karin Binder (Hg.): Beiträge zum Mathematikunterricht 2019. Münster: WTM-Verlag, S. 1087–1090.
- Epstein, Joyce L. (1988): Effective schools or effective students: Dealing with diversity. In: Ron Haskins und Duncan MacRae (Hg.): Policies for America's public schools. Teachers, equity, and indicators. Norwood, N.J.: Ablex Publ. Co (Child and family Policy, 6), S. 89–126.
- Ertl-Schmuck, Roswitha (2019): Das Theorie-Praxis-Verhältnis als produktive Irritation in der Lehrerbildung für berufsbildende Schulen. In: Bundesministerium für Bildung und Forschung (BMBF) (Hg.): Verzahnung von Theorie und Praxis im Lehramtsstudium. Erkenntnisse aus Projekten der „Qualitätsoffensive Lehrerbildung“. Frankfurt a. M.: Druck- und Verlagshaus Zarbock, S. 48–53.
- Field, Andy; Miles, Jeremy; Field, Zoë; Field, Andy P. (2012): Discovering statistics using R. Los Angeles [u.a.]: Sage.
- Fleiss, Joseph L. (1971): Measuring nominal scale agreement among many raters. In: *Psychological Bulletin* 76 (5), S. 378–382. DOI: 10.1037/h0031619.
- Florio-Hansen, Inez de (2014): Lernwirksamer Unterricht. Eine praxisorientierte Anleitung. Darmstadt: WBG.
- Freiberger, Verena; Spinath, Birgit (2014): Children's intrinsic motivation to learn: Does it decline over time and if so, why? In: Susan H. Landry und Cary L. Cooper (Hg.): Wellbeing: A Complete Reference Guide. Wellbeing in children and families. Oxford, UK: John Wiley & Sons Ltd (Volume I), S. 73–88.
- Frommelt, Manueala; Furrer Auf der Maur, Gabriela; Biaggi, Sandro; Hugener, Isabelle; Krammer, Kathrin (2016): Videos in der Ausbildung von Lehrkräften: Förderung der professionellen Unterrichtswahrnehmung durch die Analyse von eigenen bzw. fremden Videos. In: *Unterrichtswissenschaft* (04), S. 357–373.
- Geiser, Christian; Eid, Michael (2010): Item-Response-Theorie. In: Christof Wolf und Henning Best (Hg.): Handbuch der sozialwissenschaftlichen Datenanalyse. 1. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 311–332.

- Gold, Andreas; Hasselhorn, Marcus (2013): Pädagogische Psychologie. Erfolgreiches Lernen und Lehren. 3. vollständig überarbeitete und erweiterte Auflage. Hg. v. Marcus Hasselhorn, Herbert Heuer und Sylvia Schneider. Stuttgart: Kohlhammer (Standards Psychologie).
- Gold, Bernadette; Förster, Stephan; Holodyski, Manfred (2013): Evaluation eines videobasierten Trainingsseminars zur Förderung der professionellen Wahrnehmung von Klassenführung im Grundschulunterricht*. In: *Zeitschrift für Pädagogische Psychologie* 27 (3), S. 141–155. DOI: 10.1024/1010-0652/a000100.
- Goodwin, Charles (1994): Professional Vision. In: *American Anthropologist* 96 (3), S. 606–633.
- Gottfried, Adele Eskeles; Fleming, James S.; Gottfried, Allen W. (2001): Continuity of academic intrinsic motivation from childhood through late adolescence. A longitudinal study. In: *Journal of Educational Psychology* 93 (1), S. 3–13. DOI: 10.1037/0022-0663.93.1.3.
- Hanisch, Anna (2015): Kognitive Aktivierung im Rechtschreibunterricht – Konzept und Operationalisierung. In: Katrin Liebers, Brunhild Landwehr, Anne Marquardt und Kezia Schlotter (Hg.): *Lernprozessbegleitung und adaptives Lernen in der Grundschule*. Wiesbaden: Springer Fachmedien, S. 187–192.
- Hartig, Johannes; Goldhammer, Frank (2010): Modelle der Item-Response-Theorie. 1. Aufl.: Juventa Verlag.
- Hattie, John (2018): *Lernen sichtbar machen*. 4. unveränderte Auflage. Baltmannsweiler: Schneider Verlag.
- Heckhausen, Jutta; Heckhausen, Heinz (Hg.) (2018): *Motivation und Handeln*. 5., überarbeitete und erweiterte Auflage. Berlin, Germany: Springer (Springer-Lehrbuch).
- Hellermann, Christina; Gold, Bernadette; Holodyski, Manfred (2015): Förderung von Klassenführungsfähigkeiten im Lehramtsstudium. In: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 47 (2), S. 97–109. DOI: 10.1026/0049-8637/a000129.
- Helmke, Andreas (2015): *Unterrichtsqualität und Lehrerprofessionalität. Diagnose Evaluation und Verbesserung des Unterrichts*. Franz Emanuel Weinert gewidmet. [Betrachtet die Hattie-Studien]. 6. Aufl. Seelze-Velber: Kallmeyer (Schule weiterentwickeln, Unterricht verbessern : Orientierungsband).
- Heymann, Hans Werner (2015): Warum sollte Unterricht "kognitiv aktivieren"? Anregung von vertiefendem, verstehendem, vernetzendem Lernen. In: *Pädagogik (Weinheim)* 67 (5), S. 6–9.

- Hofmann, Rita; Roth, Jürgen (2020): Arbeiten mit Funktionsgraphen – Zur Diagnose von Fehlern und Fehlvorstellungen beim Funktionalen Denken. In: *mathematica didactica* 43 (1).
- Hu, Li-tze; Bentler, Peter M. (1999): Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. In: *Structural Equation Modeling: A Multidisciplinary Journal* 6 (1), S. 1–55. DOI: 10.1080/10705519909540118.
- Huang, Yizhen; Miller, Kevin F.; Cortina, Kai S.; Richter, Dirk (2023): Teachers' professional vision in action. In: *Zeitschrift für Pädagogische Psychologie* 37 (1-2), S. 122–139. DOI: 10.1024/1010-0652/a000313.
- Ibrahim, Mohamed; Antonenko, Pavlo D.; Greenwood, Carmen M.; Wheeler, Denna (2012): Effects of segmenting, signalling, and weeding on learning from educational video. In: *Learning, Media and Technology* 37 (3), S. 220–235. DOI: 10.1080/17439884.2011.585993.
- Ingenkamp, Karlheinz; Lissmann, Urban (2008): Lehrbuch der pädagogischen Diagnostik. 6., neu ausgestattete Aufl. Weinheim [u.a.]: Beltz (Beltz Pädagogik).
- Jacobs, Victoria R.; Lamb, Lisa L. C.; Philipp, Randolph A. (2010): Professional Noticing of Children's Mathematical Thinking. In: *JRME* 41 (2), S. 169–202. DOI: 10.5951/jresmetheduc.41.2.0169.
- Jahn, Gloria; Stürmer, Kathleen; Seidel, Tina; Prenzel, Manfred (2014): Professionelle Unterrichtswahrnehmung von Lehramtsstudierenden. In: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 46 (4), S. 171–180. DOI: 10.1026/0049-8637/a000114.
- Jahn, Gloria Katharina (2014): Studien zur Überprüfung der Validität eines Instruments zur Erfassung professioneller Unterrichtswahrnehmung von Lehramtsstudierenden. Dissertation.
- Kaiser, Lisa-Maria; Besa, Kris-Stephen; Wilde, Matthias; Großmann, Nadine (2021): Eine mehrdimensionale Betrachtung des Druckerlebens von Schüler*innen der Sekundarstufe II aus der Perspektive der Selbstbestimmungstheorie. In: *Zeitschrift für Erziehungswissenschaft* 24 (6), S. 1401–1427. DOI: 10.1007/s11618-021-01056-x.
- Kelava, Augustin; Moosbrugger, Helfried (2020a): Deskriptivstatistische Itemanalyse und Testwertbestimmung. In: Helfried Moosbrugger und Augustin Kelava (Hg.): *Testtheorie und Fragebogenkonstruktion*. Berlin, Heidelberg: Springer, S. 143–158.
- Kelava, Augustin; Moosbrugger, Helfried (2020b): Einführung in die Item-Response-Theorie (IRT). In: Helfried Moosbrugger und Augustin Kelava (Hg.): *Testtheorie und Fragebogenkonstruktion*. Berlin, Heidelberg: Springer, S. 369–409.

- Kelava, Augustin; Moosbrugger, Helfried (2020c): Testwertverteilung. In: Helfried Moosbrugger und Augustin Kelava (Hg.): Testtheorie und Fragebogenkonstruktion. Berlin, Heidelberg: Springer, S. 159–169.
- Kennedy, Cathleen A.; Wilson, Mark (2007): Using Progress Variables to Interpret Student Achievement and Progress (BEAR Technical Report No. 2006-12-01). University of California. Berkeley.
- Kersting, Nicole B.; Givwin, Karen B.; Sotelo, Francisco L.; Stigler, James W. (2010): Teachers' Analyses of Classroom Video Predict Student Learning of Mathematics. Further Explorations of a Novel Measure of Teacher Knowledge. In: *Journal of Teacher Education* 61 (1-2), S. 172–181. DOI: 10.1177/0022487109347875.
- Keskin, Özün; Seidel, Tina; Stürmer, Kathleen; Gegenfurtner, Andreas (2024): Eye-tracking research on teacher professional vision: A meta-analytic review. In: *Educational Research Review* 42, S. 100586. DOI: 10.1016/j.edurev.2023.100586.
- KMK (2019): Standards für die Lehrerbildung: Bildungswissenschaften, vom Beschluss der Kultusministerkonferenz vom 16.12.2004 i.d.F vom 16.05.2019.
- Kramer, Klaudia (2002): Die Förderung von motivationsunterstützendem Unterricht - Ansatzpunkte und Barrieren. Christian-Albrechts-Universität Kiel. Online verfügbar unter http://macau.uni-kiel.de/receive/dissertation_diss_00000752, zuletzt geprüft am 22.09.2016.
- Krammer, Kathrin (2014): Fallbasiertes Lernen mit Unterrichtsvideos in der Lehrerinnen- und Lehrerbildung. In: *Beiträge zur Lehrerinnen- und Lehrerbildung* 32, S. 164–175.
- Krammer, Kathrin; Reusser, Kurt (2005): Unterrichtsvideos als Medium der Aus- und Weiterbildung von Lehrpersonen. In: *Beiträge zur Lehrerinnen- und Lehrerbildung* 23 (1), S. 35–50.
- Krapp, Andreas; Hascher, Tina (2014a): Die Erforschung menschlicher Motivation. In: Lieselotte Ahnert (Hg.): *Theorien in der Entwicklungspsychologie*. Berlin: Springer VS (Springer Lehrbuch), S. 234–251.
- Krapp, Andreas; Hascher, Tina (2014b): Theorien der Lern- und Leistungsmotivation. In: Lieselotte Ahnert (Hg.): *Theorien in der Entwicklungspsychologie*. Berlin: Springer VS (Springer Lehrbuch), S. 252–281.
- Kubinger, Klaus D. (2009): *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. 2., überarb. und erw. Aufl. Göttingen: Hogrefe Verlag GmbH & Co. KG.

- Kunter, Mareike; Kleickmann, Thilo; Klusmann, Uta; Richter, Dirk (2011): Die Entwicklung professioneller Kompetenz von Lehrkräften. In: Mareike Kunter, Jürgen Baumert, Werner Blum, Uta Klusmann, Stefan Krauss und Michael Neubrand (Hg.): Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV. Münster [u.a.]: Waxmann, S. 55–68.
- Lamprianou, Iasonas (2020): Applying the Rasch Model in Social Sciences Using R and BlueSky Statistics. New York, Abingdon: Routledge.
- Landis, J. Richard; Koch, Gary G. (1977): The Measurement of Observer Agreement for Categorical Data. In: *Biometrics* 33 (1), S. 159–174. DOI: 10.2307/2529310.
- Leuders, Timo; Dörfler, Tobias; Leuders, Juliane; Philipp, Kathleen (2018): Diagnostic Competence of Mathematics Teachers: Unpacking a Complex Construct. In: Timo Leuders, Kathleen Philipp und Juliane Leuders (Hg.): Diagnostic Competence of Mathematics Teachers. Unpacking a Complex Construct in Teacher Education and Teacher Practice. Cham: Springer International Publishing, S. 3–32.
- Lienert, Gustav A.; Raatz, Ulrich (1998): Testaufbau und Testanalyse. 6. Aufl., Studienausg // 6. Auflage. Weinheim: Beltz, Psychologie-Verl.-Union; Beltz (Grundlagen Psychologie).
- Lindmeier, Anke (2013): Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In: Ulrich Riegel und Klaas Macha (Hg.): Videobasierte Kompetenzforschung in den Fachdidaktiken. Münster [u.a.]: Waxmann (Fachdidaktische Forschungen, 4), S. 45–61.
- Luhmann, Maïke (2015): R für Einsteiger. Einführung in die Statistiksoftware für die Sozialwissenschaften. 4., vollst. überarb. und erw. Aufl. Weinheim [u.a.]: Beltz.
- Luo, Yumei; Lin, Jinping; Yang, Yi (2021): Students' motivation and continued intention with online self-regulated learning: A self-determination theory perspective. In: *Zeitschrift für Erziehungswissenschaft* 24 (6), S. 1379–1399. DOI: 10.1007/s11618-021-01042-3.
- Markus, Stefan; Schwab, Susanne (2021): Zusammenhänge von sozialen Beziehungen mit schulischem Wohlbefinden und emotionalem Erleben von Grundschüler*innen: Soziale Eingebundenheit. Sozialbeziehungen im Fokus von Schule und LehrerInnenbildung. Hg. v. Gerda Hagenauer und Diana Raufelder: Waxmann, S. 351–366.
- Masters, Geoff N. (1982): A Rasch model for partial credit scoring. In: *Psychometrika* 47 (2), S. 149–174. DOI: 10.1007/BF02296272.
- Mayer, Richard E.; Moreno, Roxana (2003): Nine Ways to Reduce Cognitive Load in Multimedia Learning. In: *Educational Psychologist* 38 (1), S. 43–52. DOI: 10.1207/S15326985EP3801_6.

- Meschede, Nicola; Steffensky, Mirjam (2018): Methodologische Perspektive: Audiovisuelle Daten als Lerngelegenheiten in der Lehrer/innenbildung. In: Magdalena Sonnleitner, Stefan Prock, Astrid Rank und Petra Kirchhoff (Hg.): Video- und Audiografie von Unterricht in der LehrerInnenbildung. Planung und Durchführung aus methodologischer, technisch-organisatorischer, ethisch-datenschutzrechtlicher und inhaltlicher Perspektive. Opladen & Toronto: Barbara Budrich, S. 21–36.
- Meschede, Nicola; Steffensky, Mirjam; Wolters, Marco; Möller, Kornelia (2015): Professionelle Wahrnehmung der Lernunterstützung im naturwissenschaftlichen Grundschulunterricht. In: *Unterrichtswissenschaft* (04), S. 317–336.
- Meyer, Hilbert (2020): Was ist guter Unterricht? 15. Aufl. Berlin: Cornelsen Scriptor (Praxisbuch Praxisbuch).
- Moosbrugger, Helfried; Kelava, Augustin (2020a): Qualitätsanforderungen an Tests und Fragebogen („Gütekriterien“). In: Helfried Moosbrugger und Augustin Kelava (Hg.): Testtheorie und Fragebogenkonstruktion. Berlin, Heidelberg: Springer, S. 13–38.
- Moosbrugger, Helfried; Kelava, Augustin (Hg.) (2020b): Testtheorie und Fragebogenkonstruktion. Berlin, Heidelberg: Springer.
- Moosbrugger, Helfried; Schermelleh-Engel, Karin; Gädde, Jana C.; Kelava, Augustin (2020): Testtheorien im Überblick. In: Helfried Moosbrugger und Augustin Kelava (Hg.): Testtheorie und Fragebogenkonstruktion. Berlin, Heidelberg: Springer, S. 251–273.
- Oser, Fritz; Heinzer, Sarah; Salzmann, Patrizia (2010): Die Messung der Qualität von professionellen Kompetenzprofilen von Lehrpersonen mit Hilfe der Einschätzung von Filmvignetten. In: *Unterrichtswissenschaft* (1), S. 5–28.
- Paradies, Liane; Linser, Hans Jürgen (2009): Differenzieren im Unterricht. 4. Aufl. Berlin: Cornelsen Scriptor.
- Pauli, Christine; Reusser, Kurt (2006): Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. In: *Zeitschrift für Pädagogik* 52 (6), S. 774–798.
- Pospeschill, Markus (2010): Testtheorie, Testkonstruktion, Testevaluation. Mit 77 Fragen zur Wiederholung. München [u.a.]: Reinhardt (UTB, 3431 : Psychologie).
- Prenzel, Manfred; Kristen, Alexandra; Dengler, Petra; Ettle, Roland; Beer, Thomas (1996): Selbstbestimmt motiviertes und interessiertes Lernen in der kaufmännischen Erstausbildung. In: Klaus Beck (Hg.): Lehr-Lern-Prozesse in der kaufmännischen Erstausbildung. Wissenserwerb Motivierungsgeschehen und Handlungskompetenzen. Stuttgart: Steiner (Zeitschrift für Berufs- und Wirtschaftspädagogik : Beihefte, 13), S. 108–127.

- R Core Team (2019): R: A language and environment for statistical computing. Wien: R Foundation for Statistical Computing. Online verfügbar unter <https://www.r-project.org/>.
- R Core Team (2023): stats: The R Stats Package. Version 4.2.3.
- Rakoczy, Katrin (2006): Motivationsunterstützung im Mathematikunterricht. Zur Bedeutung von Unterrichtsmerkmalen für die Wahrnehmung von Schülerinnen und Schülern. In: *Zeitschrift für Pädagogik* 52 (6), S. 822–843.
- Rakoczy, Katrin (2008): Motivationsunterstützung im Mathematikunterricht. Unterricht aus der Perspektive von Lernenden und Beobachtern. Münster [u.a.]: Waxmann (Pädagogische Psychologie und Entwicklungspsychologie, 65).
- Rakoczy, Katrin; Buff, Alex; Lipowsky, Frank (2005): Befragungsinstrumente. In: Eckhard Klieme, Christine Pauli und Kurt Reusser (Hg.): Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie. "Unterrichtsqualität, Lernverhalten und mathematisches Verständnis", Bd. 1. Frankfurt am Main: GFPF (Materialien zur Bildungsforschung, 13).
- Rasch, Björn; Friese, Malte; Hofmann, Wilhelm Johann; Naumann, Ewald (2014a): Quantitative Methoden 2. Einführung in die Statistik für Psychologen und Sozialwissenschaftler. 4., überarb. Aufl. Berlin, Heidelberg: Springer (Springer-Lehrbuch).
- Rasch, Björn; Friese, Malte; Naumann, Ewald; Hofmann, Wilhelm (2014b): Quantitative Methoden 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler. 4., überarbeitete Auflage. Berlin, Heidelberg: Springer.
- Rauch, Dominique; Hartig, Johannes (2020): Interpretation von Testwerten in der Item-Response-Theorie (IRT). In: Helfried Moosbrugger und Augustin Kelava (Hg.): Testtheorie und Fragebogenkonstruktion. Berlin, Heidelberg: Springer, S. 411–424.
- Reeve, Johnmarshall; Jang, Hyungshim (2006): What teachers say and do to support students' autonomy during a learning activity. In: *Journal of Educational Psychology* 98 (1), S. 209–218. DOI: 10.1037/0022-0663.98.1.209.
- Rehm, Markus; Bölsterli, Katrin (2014): Entwicklung von Unterrichtsvignetten. In: Dirk Krüger, Ilka Parchmann und Horst Schecker (Hg.): Methoden in der naturwissenschaftsdidaktischen Forschung. Berlin [u.a.]: Springer Spektrum (Lehrbuch), S. 213–225.
- Reinmann, Gabi; Mandl, Heinz (2006): Unterrichten und Lernumgebungen gestalten. In: Andreas Krapp und Bernd Weidenmann (Hg.): Pädagogische Psychologie. Ein Lehrbuch. 5., vollst. überarb. Aufl., [Nachdr.]. Weinheim: Beltz, PVU, S. 613–658.

- Rentzsch, Katrin; Leplow, Bernd; Schütz, Astrid (2009): Psychologische Diagnostik. Grundlagen und Anwendungsperspektiven. 1. Aufl. Stuttgart [u.a.]: Kohlhammer (Grundriss der Psychologie, 16).
- Reusser, Kurt (2005): Situiertes Lernen mit Unterrichtsvideos. In: *Journal für Lehrerinnen- und Lehrerbildung* 2, S. 8–18.
- Revelle, William (2020): psych: Procedures for Personality and Psychological Research. Version R package version 2.0.9. Online verfügbar unter <https://CRAN.R-project.org/package=psych>.
- Rheinberg, Falko; Vollmeyer, Regina (2019): Motivation. 9., erw. und überarb. Aufl. Stuttgart: Kohlhammer (Grundriss der Psychologie, 6).
- Rheinberg, Falko; Vollmeyer, Regina; Burns, Bruce D. (2001): FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen. Online verfügbar unter <http://www.psych.uni-potsdam.de/people/rheinberg/messverfahren/FAM-Langfassung.PDF>, zuletzt geprüft am 12.07.2016.
- Richardson, Michelle; Abraham, Charles; Bond, Rod (2012): Psychological correlates of university students' academic performance: a systematic review and meta-analysis. In: *Psychological Bulletin* 138 (2), S. 353–387. DOI: 10.1037/a0026838.
- Riegel, Ulrich (2013): Videobasierte Kompetenzforschung in den Fachdidaktiken. In: Ulrich Riegel und Klaas Macha (Hg.): Videobasierte Kompetenzforschung in den Fachdidaktiken. Münster [u.a.]: Waxmann (Fachdidaktische Forschungen, 4), S. 9–24.
- Riegel, Ulrich; Macha, Klaas (Hg.) (2013): Videobasierte Kompetenzforschung in den Fachdidaktiken. Münster [u.a.]: Waxmann (Fachdidaktische Forschungen, 4).
- Robitzsch, Alexander; Kiefer, Thomas; Wu, Margaret (2020a): Package 'TAM'. Online verfügbar unter <https://cran.r-project.org/web/packages/TAM/TAM.pdf>.
- Robitzsch, Alexander; Kiefer, Thomas; Wu, Margaret (2020b): TAM: Test Analysis Modules. Version R package version 3.5-19. Online verfügbar unter <https://CRAN.R-project.org/package=TAM>.
- Rost, Jürgen (2004): Lehrbuch Testtheorie - Testkonstruktion. 2., vollst. überarb. und erw. Aufl. Bern [etc.]: Verlag Hans Huber (Aus dem Programm Huber: Psychologie Lehrbuch).
- Roth, Erwin (Hg.) (1995): Sozialwissenschaftliche Methoden. Lehr- und Handbuch für Forschung und Praxis. 4., durchges. Aufl. München: Oldenbourg (Lehr- und Handbücher der Sozialwissenschaften).

- Roth, Jürgen (2017): Videovignetten zur Analyse von Unterrichtsprozessen (ViviAn) – Ein Entwicklungs-, Forschungs- und Lehrprogramm. In: Ulrich Kortenkamp und Ana Kuzle (Hg.): Beiträge zum Mathematikunterricht 2017. Münster: WTM-Verlag, S. 1277–1280.
- Roth, Jürgen (2020): Theorie-Praxis-Verzahnung durch Lehr-Lern-Labore – Das Landauer Konzept der mathematikdidaktischen Lehramtsausbildung. In: Burkhard Priemer und Jürgen Roth (Hg.): Lehr-Lern-Labore. Berlin, Heidelberg: Springer, S. 59–83.
- RStudio Team (2019): RStudio: Integrated Development for R. Version Version 1.2.5001. Boston: R Studio, PBC. Online verfügbar unter URL <http://www.rstudio.com/>.
- Ryan, Richard M.; Deci, Edward L. (2000): Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. In: *The American psychologist* 55 (1), S. 68–78.
- Santagata, Rossella; Guarino, Jody (2011): Using video to teach future teachers to learn from teaching. In: *ZDM Mathematics Education* 43 (1), S. 133–145. DOI: 10.1007/s11858-010-0292-3.
- Schäfer, Stefanie; Seidel, Tina (2015): Noticing and reasoning of teaching and learning components by pre-service teachers. In: *Journal for Educational Research Online* 7 (2), S. 34–58.
- Schecker, Horst (2014): Überprüfung der Konsistenz von Itemgruppen mit Cronbachs α . In: Dirk Krüger, Ilka Parchmann und Horst Schecker (Hg.): Methoden in der naturwissenschaftsdidaktischen Forschung. Berlin [u.a.]: Springer Spektrum (Lehrbuch), online Zusatzmaterial. Online verfügbar unter <http://static.springer.com/sgw/documents/1426184/application/pdf/Cronbach+Alpha.pdf>.
- Schiefele, Ulrich (2014): Förderung von Interessen. In: Gerhard W. Lauth (Hg.): Interventionen bei Lernstörungen. Förderung, Training und Therapie in der Praxis. 2., überarb. und erw. Aufl. Göttingen [u.a.]: Hogrefe, S. 251–261.
- Schiefele, Ulrich; Schaffner, Ellen (2011): Lernmotivation. In: Sandra Brandt (Hg.): Lehren und Lernen im Unterricht. Baltmannsweiler: Schneider Verlag (Professionswissen für Lehrerinnen und Lehrer, / Hans Ulrich Grunder ... (Hrsg.) ; Bd. 2), S. 11–40.
- Schiefele, Ulrich; Schaffner, Ellen (2015): Motivation. In: Elke Wild und Jens Möller (Hg.): Pädagogische Psychologie. 2. Aufl. 2015. vollst. überarb. u. aktualisierte. Berlin, Heidelberg: Springer (Springer-Lehrbuch), S. 154–175.
- Schiefele, Ulrich; Schaffner, Ellen (2020): Motivation. In: Elke Wild und Jens Möller (Hg.): Pädagogische Psychologie. Third edition. Berlin, Heidelberg: Springer, S. 164–185.
- Schiefele, Ulrich; Streblov, Lilian (2006): Motivation aktivieren. In: Heinz Mandl (Hg.): Handbuch Lernstrategien. Göttingen [u.a.]: Hogrefe, S. 232–247.

- Schrader, Friedrich-Wilhelm (2009): Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. In: *Zeitschrift für Pädagogische Psychologie* 23 (34), S. 237–245. DOI: 10.1024/1010-0652.23.34.237.
- Sedlmeier, Peter; Renkewitz, Frank (2008): *Forschungsmethoden und Statistik in der Psychologie*. München: Pearson Studium.
- Seidel, Tina; Blomberg, Geraldine; Stürmer, Kathleen (2010): "OBSERVER" - Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE. In: Eckhard Klieme (Hg.): *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes*. Weinheim [u.a.]: Beltz (*Zeitschrift für Pädagogik* : Beiheft, 56), S. 296–306.
- Seidel, Tina; Prenzel, Manfred (2007): Wie Lehrpersonen Unterricht wahrnehmen und einschätzen. Die Erfassung pädagogisch-psychologischer Kompetenzen mit Videosequenzen. In: Manfred Prenzel (Hg.): *Kompetenzdiagnostik*. Wiesbaden: VS Verlag für Sozialwissenschaften (*Zeitschrift für Erziehungswissenschaft* : Sonderheft, 8), S. 201–216.
- Seidel, Tina; Stürmer, Kathleen (2014): Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. In: *American Educational Research Journal* 51 (4), S. 739–771. DOI: 10.3102/0002831214531321.
- Sherin, Miriam (2009): The development of teacher's professional vision in video clubs. In: Ricki Goldman (Hg.): *Video research in the learning sciences*. Reprint. Mahwah, NJ: Erlbaum, S. 383–395.
- Sherin, Miriam Gamoran (2002): When Teaching Becomes Learning. In: *Cognition and Instruction* 20 (2), S. 119–150. DOI: 10.1207/S1532690XCI2002_1.
- Sherin, Miriam Gamoran; Han, Sandra Y. (2004): Teacher learning in the context of a video club. In: *Teaching and Teacher Education* Vol. 20 (2), S. 163–183.
- Sherin, Miriam Gamoran; van Es, E. A. (2008): Effects of Video Club Participation on Teachers' Professional Vision. In: *Journal of Teacher Education* 60 (1), S. 20–37. DOI: 10.1177/0022487108328155.
- Smith, Richard M.; Schumacker, Randall E.; Bush, M. Joan (1998): Using Item Mean Squares to Evaluate Fit to the Rasch Model. In: *Journal of Outcome Measurement* 2 (1), S. 66–78.
- Spinath, Birgit (2005): Motivation als Kompetenz: Wie wird Motivation lehr- und lernbar? In: Regina Vollmeyer und Joachim Brunstein (Hg.): *Motivationspsychologie und ihre Anwendung*. 1. Aufl. Stuttgart: Kohlhammer, S. 203–219.

- Spinath, Birgit; Kriegbaum, Katharina; Stiensmeier-Pelster, Joachim; Schöne, Claudia; Dickhäuser, Oliver (2016): Negative Veränderungen von Zielorientierungen über Schülergenerationen hinweg. Ein 10-Jahresvergleich der SELLMO. In: *Zeitschrift für Pädagogische Psychologie* 30 (4), S. 271–278. DOI: 10.1024/1010-0652/a000185.
- Spinath, Birgit; Stiensmeier-Pelster, Joachim; Schöne, Claudia; Dickhäuser, Oliver (2012): SELLMO. Skalen zur Erfassung der Lern- und Leistungsmotivation. Manual. 2., überarb., neu normierte Auflage. Göttingen: Hogrefe Verlag GmbH & Co. KG.
- Stangl, Werner (2003): Von der Unmöglichkeit zur Motivation in der Schule. In: *Schulmagazin 5-10. Impulse für kreativen Unterricht* 71 (1), S. 9–12. Online verfügbar unter <https://www.stangl-taller.at/PSYCHOLOGIE/Motivation.pdf>.
- Star, Jon R.; Strickland, Sharon K. (2008): Learning to observe. Using video to improve preservice mathematics teachers' ability to notice. In: *Journal of Mathematics Teacher Education* 11 (2), S. 107–125. DOI: 10.1007/s10857-007-9063-7.
- Steffensky, Mirjam; Kleinknecht, Marc (2016): Wirkungen videobasierter Lernumgebungen auf die professionelle Kompetenz und das Handeln (angehender) Lehrpersonen. Ein Überblick zu Ergebnissen aus aktuellen (quasi-)experimentellen Studien. In: *Unterrichtswissenschaft* (04), S. 305–322.
- Steinmayr, Ricarda; Spinath, Birgit (2007): Predicting School Achievement from Motivation and Personality. In: *Zeitschrift für Pädagogische Psychologie* 21 (3/4), S. 207–216. DOI: 10.1024/1010-0652.21.3.207.
- Steyer, Rolf; Eid, Michael (2001): Messen und Testen. 2. Auflage. Berlin, Heidelberg: Springer.
- Strobl, Carolin (2012): Das Rasch-Modell. Eine verständliche Einführung für Studium und Praxis. 2., erweiterte Auflage. Mering: Rainer Hampp Verlag (Sozialwissenschaftliche Forschungsmethoden).
- Stürmer, Kathleen; Könings, Karen D.; Seidel, Tina (2013): Declarative knowledge and professional vision in teacher education: effect of courses in teaching and learning. In: *The British Journal of Educational Psychology* 83 (Pt 3), S. 467–483. DOI: 10.1111/j.2044-8279.2012.02075.x.
- Stürmer, Kathleen; Seidel, Tina (2015): Assessing Professional Vision in Teacher Candidates. In: *Zeitschrift für Psychologie* 223 (1), S. 54–63. DOI: 10.1027/2151-2604/a000200.
- Stürmer, Kathleen; Seidel, Tina; Kunina-Habenicht, Olga (2015): Unterricht wissensbasiert beobachten. Unterschiede und erklärende Faktoren bei Referendaren zum Berufseinstieg. In: *Zeitschrift für Pädagogik* 61 (3), S. 345–360.

- Sunder, Cornelia; Todorova, Maria; Möller, Kornelia (2016): Kann die professionelle Unterrichtswahrnehmung von Sachunterrichtsstudierenden trainiert werden? – Konzeption und Erprobung einer Intervention mit Videos aus dem naturwissenschaftlichen Grundschulunterricht. In: *ZfDN* 22, S. 1–12. DOI: 10.1007/s40573-015-0037-5.
- Syring, Marcus; Bohl, Thorsten; Kleinknecht, Marc; Kuntze, Sebastian; Rehm, Markus; Schneider, Jürgen (2015): Videos oder Texte in der Lehrerbildung? Effekte unterschiedlicher Medien auf die kognitive Belastung und die motivational-emotionalen Prozesse beim Lernen mit Fällen. In: *Zeitschrift für Erziehungswissenschaft* 18 (4), S. 667–685. DOI: 10.1007/s11618-015-0631-9.
- Tepner, Oliver; Dollny, Sabrina (2014): Entwicklung eines Testverfahrens zur Analyse fachdidaktischen Wissens. In: Dirk Krüger, Ilka Parchmann und Horst Schecker (Hg.): *Methoden in der naturwissenschaftsdidaktischen Forschung*. Berlin [u.a.]: Springer Spektrum (Lehrbuch), S. 311–323.
- Trendtel, Matthias; Pham, Giang; Yanagida, Takuya (2016): Skalierung und Linking. In: Simone Breit und Claudia Schreiner (Hg.): *Large-Scale Assessment mit R. Methodische Grundlagen der österreichischen Bildungsstandardüberprüfung*. Wien: Facultas, S. 185–224.
- van Es, Elizabeth A.; Sherin, Miriam G. (2002): Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. In: *Journal of Technology and Teacher Education* 10 (4), S. 571–596.
- van Es, Elizabeth A. (2009): Participants' Roles in the Context of a Video Club. In: *The Journal of the Learning Sciences* Vol. 18 (1), S. 100–137.
- Vock, Miriam; Gronostaj, Anna (2017): *Umgang mit Heterogenität in Schule und Unterricht*. 1. Auflage. Berlin: Friedrich-Ebert-Stiftung, Abt. Studienförderung (Schriftenreihe des Netzwerk Bildung, 40).
- von Aufschnaiter, Claudia; Selter, Christoph; Michaelis, Julia (2017): Nutzung von Vignetten zur Entwicklung von Diagnose- und Förderkompetenzen - Konzeptionelle Überlegungen und Beispiele aus der MINT-Lehrerbildung. In: Christoph Selter, Stephan Hußmann, Corinna Höhle, Christine Knipping, Katja Lengnink und Julia Michaelis (Hg.): *Diagnose und Förderung heterogener Lerngruppen. Theorien, Konzepte und Beispiele aus der MINT-Lehrerbildung*. 1. Auflage. Münster, New York: Waxmann, S. 85–105.
- Walz, Moritz (2020): *Das Interventionsverhalten von Studierenden mit divergierender prozessdiagnostischer Fähigkeit "Deuten"*. Dissertation. Universität Koblenz-Landau.

- Walz, Moritz; Roth, Jürgen (2019): Interventionen in Schülergruppenarbeitsprozesse und Reflexion von Studierenden – Einfluss diagnostischer Fähigkeiten. In: Andreas Frank, Stefan Krauss und Karin Binder (Hg.): Beiträge zum Mathematikunterricht 2019. Münster: WTM-Verlag, S. 1099–1102.
- Weinert, Franz E. (2000): Lehren und Lernen für die Zukunft - Ansprüche an das Lernen in der Schule. In: *Pädagogische Nachrichten Rheinland-Pfalz* (2), S. 1–16.
- Wettstein, Alexander; Raufelder, Diana (2021): Beziehungs- und Interaktionsqualität im Unterricht: Soziale Eingebundenheit. Sozialbeziehungen im Fokus von Schule und LehrerInnenbildung. Hg. v. Gerda Hagenauer und Diana Raufelder: Waxmann, S. 17–28.
- Wild, Elke; Möller, Jens (2009): Pädagogische Psychologie. Mit 27 Tabellen. Heidelberg: Springer.
- Wild, Elke; Möller, Jens (Hg.) (2015): Pädagogische Psychologie. 2. Aufl. 2015. vollst. überarb. u. aktualisierte. Berlin, Heidelberg: Springer (Springer-Lehrbuch).
- Wild, Elke; Möller, Jens (Hg.) (2020): Pädagogische Psychologie. Third edition. Berlin, Heidelberg: Springer.
- Wilson, Mark; Masters, Geoffrey N. (1993): The partial credit model and null categories. In: *Psychometrika* Vol. 58 (1), S. 87–99.
- Wirtz, Markus Antonius; Nachtigall Christof (2008): Deskriptive Statistik: Statistische Methoden für Psychologen. Teil 1. 5., überarb. Aufl. Weinheim [u.a.]: Juventa Verlag.
- Witte, Kerstin (2019): Angewandte Statistik in der Bewegungswissenschaft (Band 3). Berlin, Heidelberg: Springer.
- Wright, Benjamin D.; Masters, Geoffrey N. (1982): Rating scale analysis. Chicago: Mesa Press.
- Wu, Margaret; Tam, Hak Ping; Jen, Tsung-Hau (2016): Educational Measurement for Applied Researchers. Singapore: Springer Singapore.
- Yadav, Aman; Phillips, Michael M.; Lundeberg, Mary A.; Koehler, Matthew J.; Hilden, Katherine; Dirkin, Kathryn H. (2011): If a picture is worth a thousand words is video worth a million? Differences in affective and cognitive processing of video and text cases. In: *Journal of Computing in Higher Education* 23 (1), S. 15–37. DOI: 10.1007/s12528-011-9042-y.
- Zöfel, Peter (2003): Statistik für Psychologen. Im Klartext. München: Pearson, Higher Education (PS - Psychologie).
- Züll, Cornelia; Menold, Natalja (2014): Offene Fragen. In: Nina Baur und Jörg Blasius (Hg.): Handbuch Methoden der empirischen Sozialforschung, Bd. 17. Wiesbaden: Springer Fachmedien, S. 713–719.

Abbildungsverzeichnis

<i>Abbildung 1 Teilprojekte von DiAmant.....</i>	<i>3</i>
<i>Abbildung 2 Selbstbestimmungstheorie.....</i>	<i>6</i>
<i>Abbildung 3 Regulationsformen.....</i>	<i>7</i>
<i>Abbildung 4 Ergebnisse der PLS-SEM.....</i>	<i>9</i>
<i>Abbildung 5 TARGET und IBAS – Strukturen</i>	<i>11</i>
<i>Abbildung 6 IBAS-Modell.....</i>	<i>12</i>
<i>Abbildung 7 Professionelle Kompetenz von Lehrpersonen.....</i>	<i>24</i>
<i>Abbildung 8 Diagnoseprozess</i>	<i>25</i>
<i>Abbildung 9 Situationsbezogene Fähigkeiten im Modell „Kompetenz als Kontinuum“</i>	<i>26</i>
<i>Abbildung 10 Professionelle Unterrichtswahrnehmung.....</i>	<i>27</i>
<i>Abbildung 11 Professionelle Unterrichtswahrnehmung mit den Teilprozessen.....</i>	<i>29</i>
<i>Abbildung 12 Das Videotool OBSERVER.....</i>	<i>46</i>
<i>Abbildung 13 Erweiterung des OBSERVERS zum Extended Research Tool</i>	<i>48</i>
<i>Abbildung 14 Startseite der Homepage zum Videotool ViviAn</i>	<i>57</i>
<i>Abbildung 15 Abrufen von Zusatzinformationen am Beispiel "Thema und Ziele"</i>	<i>58</i>
<i>Abbildung 16 Unterrichtsplanung der Lehrperson mit ergänzter Instruktion</i>	<i>86</i>
<i>Abbildung 17 Auszug aus dem Fragebogen zur Erfassung der aktuellen Motivation</i>	<i>89</i>
<i>Abbildung 18 Standbild aus einer Videovignette.....</i>	<i>92</i>
<i>Abbildung 19 Videoschnitt-Plan</i>	<i>101</i>
<i>Abbildung 20 Beschreibung der Vignetten 1 und 2.....</i>	<i>103</i>
<i>Abbildung 21 Beschreibung der Vignetten 4 und 5.....</i>	<i>104</i>
<i>Abbildung 22 Beschreibung der Vignetten 6 und 7.....</i>	<i>105</i>
<i>Abbildung 23 Analyseaufträge zum Beschreiben</i>	<i>107</i>
<i>Abbildung 24 Analyseaufträge zum Erklären.....</i>	<i>108</i>
<i>Abbildung 25 Analyseaufträge zum Vorhersageng.....</i>	<i>108</i>
<i>Abbildung 26 Erklärungen für ein nicht gezeigtes, aber motivationsförderliches Lehrpersonenhandeln</i>	<i>109</i>
<i>Abbildung 27 Lernumgebung ViviAn am Beispiel eines Videos zum Themenbereich Motivierung im Unterricht.....</i>	<i>113</i>
<i>Abbildung 28 Einschätzung der Relevanz an der Arbeit mit Videos von den Studierenden des Seminars</i>	<i>117</i>

<i>Abbildung 29 Beurteilung der Interessantheit an der Arbeit mit Videos von den Studierenden des Seminars</i>	<i>117</i>
<i>Abbildung 30 Einschätzung der Gestaltung der Lernumgebung von den Studierenden des Seminars</i>	<i>117</i>
<i>Abbildung 31 Category Characteristic Curve</i>	<i>124</i>
<i>Abbildung 32 ungeordnete Schwellenparameter</i>	<i>125</i>
<i>Abbildung 33 Übersicht der Items des finalen Modells</i>	<i>144</i>
<i>Abbildung 34 Studiendesign</i>	<i>158</i>
<i>Abbildung 35 mittlerer Summenscore pro Dimension im Vergleich zwischen Pre- und Posttestung</i>	<i>160</i>
<i>Abbildung 36 Beispiel für eine Aufgabenstellung im Rahmen des Seminars</i>	<i>161</i>

Tabellenverzeichnis

<i>Tabelle 1 Instrumente zur Erfassung der Motivation im Lehr-Lernkontext.....</i>	17
<i>Tabelle 2 Übersicht über ausgewählte Studien zur Erforschung der Professionellen Unterrichtswahrnehmung mittels Videos</i>	33
<i>Tabelle 3 Beispielitems aus dem OBSERVER.....</i>	46
<i>Tabelle 4 Vergleich der EAP-Reliabilitäten zwischen dem OBSERVER und der erweiterten Version</i>	48
<i>Tabelle 5 Klassifikation der Referendar:innen nach Fähigkeit.....</i>	49
<i>Tabelle 6 Legende zur Abweichung der Mittelwerte pro Faktor vom Gesamtmittelwert</i>	90
<i>Tabelle 7 Dokumentation der Videoanalyse.....</i>	91
<i>Tabelle 8 Reliabilität der Skalen.....</i>	92
<i>Tabelle 9 Auswertung der aktuellen Lernmotivation der Klassenstufe 7.....</i>	92
<i>Tabelle 10 Instruktionen der Lehrperson, Schwerpunkte motivationsfördernder und -hemmender Aspekte für die Unterrichtseinheit.....</i>	94
<i>Tabelle 11 Testung auf Normalverteilung Klassenstufe 7.....</i>	95
<i>Tabelle 12 Mittelwertvergleich zwischen der intrinsischen Motivation zu verschiedenen Messzeitpunkten – Klassenstufe 7</i>	96
<i>Tabelle 13 Auswertung der aktuellen Lernmotivation der Klassenstufe 8</i>	97
<i>Tabelle 14 Instruktionen der Lehrperson, Schwerpunkte motivationsfördernder und -hemmender Aspekte für die Unterrichtseinheit.....</i>	97
<i>Tabelle 15 Testung auf Normalverteilung Klassenstufe 8</i>	99
<i>Tabelle 16 Mittelwertvergleich zwischen der intrinsischen Motivation zu verschiedenen Messzeitpunkten – Klassenstufe 8</i>	100
<i>Tabelle 17 Interpretation Fleiss-Kappa.....</i>	112
<i>Tabelle 18 Kurzbeschreibung des Fragebogens</i>	116
<i>Tabelle 19 Beurteilung der Vignetten und Analyseaufträge.....</i>	118
<i>Tabelle 20 beispielhafte Anpassung der Items</i>	119
<i>Tabelle 21 Verteilung nach Semestern – Gesamtstichprobe</i>	120
<i>Tabelle 22 Verteilung nach Semestern - Teilstichproben N1 und N2.....</i>	121
<i>Tabelle 23 aufgestellte Lavaan-Modelle für Modellvergleiche</i>	128
<i>Tabelle 24 Deskriptive Statistik).....</i>	133
<i>Tabelle 25 Modellvergleiche mittels Anova).....</i>	134
<i>Tabelle 26 Schwellenparameter</i>	135

<i>Tabelle 27 Zusammenfassung der Q3-Statistik)</i>	136
<i>Tabelle 28 Q3 Statistik „model1“</i>	137
<i>Tabelle 29 kritische Outfitwerte der Items im „model 1“</i>	137
<i>Tabelle 30 Trennschärfe der Items der Dimension des Beschreibens mit Angabe der Itemschwierigkeit und der Häufigkeit der Antworten pro Kategorie</i>	138
<i>Tabelle 31 Trennschärfe der Items der Dimension des Erklärens mit Angabe der Itemschwierigkeit und der Häufigkeit der Antworten pro Kategorie</i>	139
<i>Tabelle 32 Trennschärfe der Items der Dimension des Vorhersagens mit Angabe der Itemschwierigkeit und der Häufigkeit der Antworten pro Kategorie</i>	140
<i>Tabelle 33 ausgeschlossene Items</i>	141
<i>Tabelle 34 kritische Outfitwerte der Items im finalen Modell</i>	142
<i>Tabelle 35 kritische Q3-Werte im finalen Model</i>	142
<i>Tabelle 36 Trennschärfe pro Kategorie</i>	143
<i>Tabelle 37 Anova zwischen dem "model1" und dem finalen Modell „finalmod“</i> ...	143
<i>Tabelle 38 Modellvergleich zwischen Stichprobe N1 und N2 mittels Anova</i>	145
<i>Tabelle 39 Verteilung nach Semestern – Gesamtstichprobe</i>	150
<i>Tabelle 40 Mittelwerte der normierten erreichten Punktzahl mittels Prozenttransformation über alle Proband:innen mit Standardabweichung</i>	152
<i>Tabelle 41 Verteilung nach Semestern – Stichprobe Pre- und Posttest</i>	156
<i>Tabelle 42 Effektstärke nach Cohen</i>	159
<i>Tabelle 43 Modellvergleiche mittels Anova</i>	166
<i>Tabelle 44 Kennwerte entsprechend der Stichprobe</i>	168
<i>Tabelle 45 Vergleich zwischen der Pre- und Posttestung mit Angabe des t-Test- Ergebnisses und der Effektstärke</i>	170

Anhang

- A. Fragebogenanalyse: Erhebung der Motivation
- B. Fragebogen Aktuelle Lernmotivation
- C. Feldforschung - Dokumentation
- D. Hospitationsprotokoll – Ein Beispiel
- E. Dokumentation der Videoanalyse
- F. ViviAn – Anleitung für Studierende
- G. Itembeschreibung mit Musterlösung
- H. Itemkennwerte
- I. Selbstständigkeitserklärung

A. Fragebogenanalyse: Erhebung der Motivation

Testgliederung	Antwortskala	Beispiel	Gütekriterien	Anmerkung
SELLMO Skalen zur Erfassung der Lern- und Leistungsmotivation - Spinath (2002)				
31 Items, 4 Fakt.: - Lernziele - Annäherungs-Leistungsziele - Vermeidungs-Leistungsziele - Arbeitsvermeidung Bearbeitungsdauer: 8-15 Min.	5-stufig - Stimmt gar nicht - Stimmt eher nicht - Weder/noch - Stimmt eher - Stimmt genau	In der Schule geht es mir darum... - neue Ideen zu bekommen (LZ) - zu zeigen, dass ich bei einer Sache gut bin (ALZ) - dass andere Schülerinnen und Schüler nicht denken, ich sei dumm (VLZ) - keine schwierigen Tests oder Arbeiten zu haben (AV)	- Split-Half-Reliabilität.: 0.73 - 0.78 - Retest-R.: 0.54 - 0.63 - Faktorenanalyse: Vierfaktorenlösung - Konstruktvalidität nachgewiesen - Normierung: Getrennt für Klassenstufe 4-6 und 7-10	+ Vergleich mit Norm, Einstufung der Klasse zu Beginn möglich - Zu viele Items für Befragung innerhalb des Unterrichts
Videostudie "Unterrichtsqualität, Lernverhalten und mathematisches Verständnis" Zwischenerhebung: Aktuelle Motivation nach Video - Rakoczy (2005)				
15 Items, 5 Fakt.: - amotiviert - external - introhiziert - identifiziert - intrinsich .	4-stufig - Stimmt gar nicht - Stimmt eher - Stimmt genau	In diesen Mathematikstunden... - machte das Lernen/Arbeiten Spaß (Intri) - habe ich mitgearbeitet, weil ich die Sache später brauchen kann (id) - habe ich nur das getan, wozu mich der Lehrer aufgefordert hat (ex) - war mir alles egal (am) - war ich aufmerksam, weil ich immer aufpasse (ij)	Cronbachs Alpha: 0.61 - 0.73 Exploratorische Faktorenanalyse: Nach Kaiser-Kriterium nur 4 Faktoren (intrinsisch und identifiziert zusammen). Aus inhaltlichen Gründen werden die Skalen für weitere Analysen dennoch getrennt	+ Differenzierung d. Motivationsarten + Formulierung gut geeignet für Befragung innerhalb des Unterrichts + allgemeine Aussagekraft zur Motivation entsprechend der Phase sollte vergleichbar sein + Vergleich mit Kramer möglich
Multiple Stait- und Trait-Ziele - Bürger (2013)				
11 Items, 3 Fakt.: Bedürfnisse - Autonomie - Kompetenz - Soziale Eingebundenheit 4 Items: Intrinsische Motivation Individuelle Lernvoraussetzungen 12 Items, 2 Fakt.: - Interesse - Kompetenzüberzeugung	4-stufig - Nie - Selten - Manchmal - Häufig 4-stufig: - Stimmt gar nicht bis - Stimmt genau	Im Mathematikunterricht... - Kann ich selber entscheiden wie ich arbeiten will (A) - Werde ich für gute Leistungen gelobt (K) - Habe ich das Gefühl dazuzugehören (S) In diesen Mathematikstunden... verging die Zeit wie im Flug (Intr) Interesse + Mathe ist spannend - Mathe ist langweilig Kompetenzüberzeugung + Mathe liegt mir - Ich bin unbegabt für Mathe	- Bedürfnisse: Cronbachs Alpha: .62 - .78 - Intrinsische Motivation: Cronbachs Alpha .72 - Faktorenanalyse - Individuelle Lernvoraussetzungen: Cronbachs Alpha: .91	+ Aufteilung in Bedürfnisse + Interesse und Kompetenzüberzeugungen: wenig Items, gute Reliabilität, schnell zu überprüfen
FAM - Rheinberg				
12 Items, 4 Fakt.: - Lernziele (MAS) - Annäherungs-Leistungsziel (PAP)	4-stufig - Trifft gar nicht zu - Trifft völlig zu	In der Schule verfolge ich im Allgemeinen das Ziel... - Viele neue Fähigkeiten zu erwerben (MAS)	- Cronbachs Alpha: .73 - .87 - Exploratorische Faktorenanalyse	+ Faktoren z.T. wie beim Sellmo + Affiliationsziele ähnlich wie

<ul style="list-style-type: none"> - Arbeitsvermeidungs-Ziel (WOA) - Affiliations-Ziel (AFL) 		<ul style="list-style-type: none"> - Arbeiten besser zu schaffen als andere (PAP) - Nicht so schwer zu arbeiten (WOA) - Mich mit Klassenkameraden auszutauschen (AFL) 		<p>Bedürfnis nach soz. Eingebundenheit</p> <p>+ Differenzierung zw. Merkmalen der Schüler:innenpersönlichkeit und Unterricht</p> <p>- Zu spezifische Fragestellungen (direkter Aufgabenbezug)</p> <p>- Skalierung zu feinstufig? Entscheidungsphase länger</p> <p>Fazit: eher ungeeignet für geplante Unterrichtsphasen</p>
<p>18 Items, 4 Fakt.:</p> <ul style="list-style-type: none"> - Herausforderung - Interesse - Erfolgswahrscheinlichkeit - Misserfolgsbefürchtung 	<p>7 stufig:</p> <ul style="list-style-type: none"> - trifft nicht zu - bis trifft zu 	<p>Meine momentane Einstellung..</p> <ul style="list-style-type: none"> - Ich mag solche Rätsel und Knocheleien (I) - Ich glaube, dass kann jeder schaffen (E) - Es ist mir etwas peinlich, hier zu versagen (M) - Die Aufgabe ist eine richtige Herausforderung für mich (H) 		

Interventionsstudie motivationsunterstützender Unterricht in der kaufmännischen Erstausbildung

- Kramer (2002)

<p>38 Items, 6 Fakt.:</p> <ul style="list-style-type: none"> - Wahrgenommene Unterrichtsbedingungen der Lernsituation - Inhaltliche Relevanz - Instruktionsqualität (Klarheit +Überforderung) - Soziale Eingebundenheit - Inhaltliches Interesse beim Lehrenden - Kompetenzunterstützung - Autonomieunterstützung 	<p>3 stufig</p> <ul style="list-style-type: none"> - 0 kam nicht bzw. Nie vor - 1 kam zum Teil bzw. Gelegentlich vor - 2 kam sehr stark bzw. sehr häufig vor 	<p>In dieser Unterrichtsstunde...</p> <ul style="list-style-type: none"> - Wurde ich in Situationen gebracht, wo ich selbst merken konnte, wie wichtig der Stoff ist (IR) - Ging mir alles zu schnell (Ü) - Habe ich einen Überblick über die geplante Vorgehensweise erhalten (KL) - Hat mich die Begeisterung der Lehrkraft richtig angesteckt (II) - War die Atmosphäre freundlich und entspannt (SE) - Wurde ich sachlich über meine Fortschritte informiert (KO) 	<p>In früheren Untersuchungen wurden die Gütekriterien überprüft und der Fragebogen angepasst (Prenzel et al., 1996)</p> <ul style="list-style-type: none"> - Wahrgenommene Unterrichtsbedingungen Cronbachs Alpha: .65 - .87 - Varianten der Lernmotivation Cronbachs Alpha: .68 - .78 - Emotionale Empfindungen Cronbachs Alpha: .81 - .89 	<ul style="list-style-type: none"> - Skalierung auf 3 Stufen, zu große Abstände - 38 Items aus 46 bei Prenzel extrahiert, aber im Fragebogen enthalten →Transparenz fehlt + Differenzierung nach Motivationsarten möglich, Faktoren wie bei Rakoczy, aber höhere Reliabilität, da mehr Items? + 3 Bedürfnisse auch abgefragt
<p>36 Items, 6 Fakt.:</p> <p>Lernmotivation</p> <ul style="list-style-type: none"> - Amotiviert - External - Intrinsisch - Identifiziert - Introjiert - interessiert 				
<p>16 Items, 3 Fakt.:</p> <p>Emotionale Empfindungen</p> <ul style="list-style-type: none"> - Wichtigkeits- - Positive und - Negative Empfindungen 				

B. Fragebogen Aktuelle Lernmotivation

Kramer, K. (2002). Die Förderung von motivationsunterstützendem Unterricht - Ansatzpunkte und Barrieren. Retrieved from http://macau.uni-kiel.de/receive/dissertation_diss_00000752

Prenzel, M., Kristen, A., Dengler, P., Ettl, R., & Beer, T. (1996). Selbstbestimmt motiviertes und interessiertes Lernen in der kaufmännischen Erstausbildung. In K. Beck (Ed.), *Zeitschrift für Berufs- und Wirtschaftspädagogik : Beihefte: Vol. 13. Lehr-Lern-Prozesse in der kaufmännischen Erstausbildung. Wissenserwerb Motivierungsgeschehen und Handlungskompetenzen* (pp. 108–127). Stuttgart: Steiner.

Skaleninstruktion	In diesem Fragebogen geht um deine Motivation in der Phase des Unterrichts, die du gerade hattest. Deine Angaben werden vollkommen anonym behandelt. Bitte lies dir alle Fragen sorgfältig durch und kreuze das an, was du denkst und für dich am besten passt. Denke bitte bei allen Fragen nur an die letzte Phase innerhalb der Stunde.
Itemstamm	Beim Lernen/Arbeiten in dieser Unterrichtsphase
Skalierung	(1) stimmt gar nicht, (2) stimmt eher nicht, (3) teils teils (4) stimmt eher, (5) stimmt genau
Variable	Text
<i>intri</i>	Aktuelle Lernmotivation - intrinsisch
<i>id</i>	Aktuelle Lernmotivation - identifiziert
<i>ex</i>	Aktuelle Lernmotivation - external
<i>am</i>	Aktuelle Lernmotivation - amotiviert
<i>intro</i>	Aktuelle Lernmotivation – introjiziert
	Verwendete Items
<i>intri1</i>	...machte das Lernen/Arbeiten richtig Spaß.
<i>intri2</i>	...war ich neugierig/wissbegierig
<i>intri3</i>	...verging die Zeit wie im Flug
<i>id1</i>	...wollte ich selbst den Stoff verstehen
<i>id2</i>	...war mir klar, dass ich das später brauchen kann/können muss
<i>id3</i>	...habe ich mich eingesetzt, weil ich meinen eigenen Zielen ein Stück näher kommen konnte
<i>ex1</i>	...habe ich mich nur angestrengt/nur mitgemacht, damit ich keinen Ärger bekomme
<i>ex2</i>	...habe ich nur das getan, wozu mich die Lehrkraft aufgefordert hat
<i>ex3</i>	...habe ich nur das getan/gelernt, was von mir verlangt wurde
<i>am1</i>	...war mir alles egal
<i>am2</i>	...war ich mit meinen Gedanken woanders
<i>am3</i>	...hatte ich keine Lust, mich zu beteiligen
<i>intro1</i>	...habe ich mich angestrengt, wie sich das für ordentliche Schüler/innen gehört
<i>intro2</i>	...versuchte ich, alles so zu erledigen, wie es von mir erwartet wird
<i>intro3</i>	...war ich aufmerksam, um alles möglichst richtig/gut zu machen

	Anmerkungen
	1. Ausgeschlossene Items
<i>Amotiviert</i>	...versuchte ich mich zu drücken
<i>External</i>	...hätte ich ohne Druck von außen nichts getan
<i>Introjiert</i>	...habe ich mich selbst unter Druck gesetzt, um alles möglichst richtig/gut zu machen
	2. Originalitems vor Änderung
<i>id1</i>	...wollte ich selbst den Stoff verstehen/beherrschen
<i>id2</i>	...war mir klar, dass ich das für meinen Beruf können muss
<i>ex1</i>	...habe ich mich nur angestrengt, damit ich keinen Ärger bekomme
<i>ex3</i>	...habe ich nur das getan/gelernt, was ausdrücklich von mir verlangt wurde
<i>intro1</i>	...habe ich mich angestrengt, wie sich das für ordentliche Auszubildende gehört
	3. neu entwickelte Items
<i>ex3</i>	...habe ich nur das getan/gelernt, was von mir verlangt wurde
<i>intro3</i>	...war ich aufmerksam, um alles möglichst richtig/gut zu machen
<i>am3</i>	...hatte ich keine Lust, mich zu beteiligen

C. Feldforschung - Dokumentation

Feldforschungstagebuch I

PROTOKOLL DER BEOBACHTUNG

[. .2017] [VON BIS UHR] [ORT:]

Protokollführer/in:

Teilnehmende:

BEOBACHTUNGSNOTIZEN (W-Fragen, Normen, Interaktionen)	Theoretische Notizen (Hypothesen, Interpretationen, Schlussfolgerungen)	Methodische Notizen (Rolle, Verhalten, Identität des/r Beobachters/in)

Feldforschungstagebuch II

PROTOKOLLBLATT FÜR ALLTAGDOKUMENTE

[. .2017] [VON BIS UHR] [ORT:]

Protokollführer/in:

Teilnehmende:

CHECKLISTE

ERGÄNZUNGEN

- Raumskizze
- Umgebungsbeschreibung
- Sitzplan
- Tafelbilder
- Dokumente
- Mitschriften
-

D. Hospitationsprotokoll – Ein Beispiel

Hospitationsprotokoll

DATUM	31.08.2017	KLASSE	8, 21 SuS
STARTZEIT	7:55 Uhr	FACHLEHRERIN	Frau xx
ENDZEIT	9:26 Uhr	FACH	Erdkunde
		HOSPITANTIN	Frau Treiber

Thema der Stunde: Platten und Plattentektonik

Hospitationsschwerpunkte

(1) DIDAKTISCHE KRITERIEN

Unterrichtsverfahren

- Frontalunterricht

Lehrmethode

- Lehrvortrag
- Vor allem Lehrgespräch
 - Bei der Besprechung der Aufgaben: SuS melden sich und werden dann aufgerufen, selten auch ohne Meldung (vor allem, wenn jm. noch nichts im Unterricht gesagt hat)

Lernformen

- Einzelarbeit und Schülervortrag
- Partnerarbeit, Methode: Murrelgespräch
- Zahlenroulette: durchzählen, jeder merkt sich seine Zahl, danach neuen Sitzplatz suchen, Nr.17 beginnt (von Hospitantin zufällig ausgewählt) und soll mindestens einen Satz zur Theorie von Wegner sagen, Regeln müssen beachtet werden (keine Zahl direkt davor oder danach, kein Satz doppelt benennen)

Lernziele

- Werden nicht klar transparent gemacht

Strukturierung

- Klar erkennbare Struktur
- Zu Beginn Schülervortrag
- Am Ende des Vortrages erinnert die Lehrperson wer in der nächsten Woche dran ist
- Folgender Schwerpunkt: Vergleich von Aufgaben aus der Vertretungsstunde der letzten Woche sowie vorangegangener Hausaufgaben (die zu bearbeitenden Aufgaben werden an der Tafelinnenseite notiert: LB S.xy, Nr. x,y)
 - Alle Aufgaben werden schrittweise besprochen und verglichen, Ergebnissicherung über Tafelbilder, Folien, Arbeitsblätter
- Gibt Hinweise, was besonders wichtig ist
 - „ihr könnt euch merken“, weist dabei auf einen Test hin
- Lehrkraft notiert „Hausaufgabe: Lernen“
- Kein offensichtliches Stundenende, die Klasse wird sehr unruhig, SuS packen die Sachen ein, dann kommt ein kurzes „dürft in die Pause gehen“

(2) FÜHRUNGSSTIL

Direkt vs. indirekt

- Festgelegter und gesteuerter Ablauf der Stunde, alles vorstrukturiert (Lernprozesse, Zeitplan)
- Zu Beginn der Stunde stehen alle SuS und werden von der Lehrperson begrüßt, die SuS begrüßen die Lehrperson mit einem monotonen und langgezogenen „Guten Morgen Frau xx“
- Die Hospitantin wird kurz vorgestellt und darf sich selbst kurz äußern, sodass die SuS wissen, wieso noch jemand im Unterricht sitzt
- Die Lehrperson fragt die SuS, was sie in der letzten Woche gemacht haben
 - die Lehrperson notiert die Aufgaben an der Tafel
- Beim Vergleichen der Lösungen: die Lehrperson fordert die SuS direkt mit Namen auf
 - „xx, bitte Vorlesen“
 - Gibt eine kurze Rückmeldung, ob die Antwort „richtig“ ist oder fragt „Haben das alle von euch so?“, „Wo ist der Unterschied bei dir?“ oder sagt „bitte korrigiert es“ (und legt dazu eine Folie auf), „bitte ergänzen“
- Anpassung und Flexibilität an Bedürfnisse und Wünsche der SuS wird nicht ersichtlich, haben scheinbar kein Einfluss auf Unterrichtsverlauf

Lernberatung

- Unterstützung beim Ausführen von Lernhandlungen beim Schülervortrag
 - Vorhandenes Wissen soll abgerufen werden: zwei Vortragsweisen (Müller 1 & Müller 2)
 - Affektive Unterstützung während des Vortrags: jeder ist aufgeregt, Ruhe bewahren, einen guten Freund ansehen, sich wohlfühlen, durchatmen
- Vorschläge zur Lerntechnik: Eingehen auf verschiedene Lerntypen (lesen, markieren, zeichnen)
 - Lehrperson geht genauer auf das Zeichnen ein
 - Nachfrage „Warum solltet ihr was abzeichnen?“
 - geht auf Unterschied zwischen Einkleben eines Ausdrucks und selber Erstellen einer Skizze ein, dabei kritisiert sie das Vorgehen des Einklebens
- Hilfestellung bei Lernvorbereitung: Arbeitstechnik
 - Zusammenfassung in farbigen Rahmen setzen bzw. farbig hervorheben, damit wird deutlich, dass der Teil eine wichtige Grundlage bildet
 - Lehrperson empfiehlt zweifarbig zu arbeiten, weil es dadurch übersichtlicher wird
- Herausfordern von verfügbarem oder tragem Wissen
 - Aufforderungen vorhandenes Wissen wieder zu aktivieren
 - Einfordern von Erklärungen
- Lernhandlungen reflektieren: Aufforderung zur metakognitiven Reflexion
 - Die Lehrperson fordert auf zu reflektieren, warum ein Plakat eingesetzt wurde, gibt aber den Lernenden keine Möglichkeit zu antworten, sondern nutzt die Frage eher als Überleitung zu Vorschlägen, wie man es einbinden könnte

Wertschätzung/Caring

- ermutigt bei Vorträgen
 - gibt Tipps, Jeder ist aufgeregt, es wird keiner ausgelacht, hält während des Vortrags Blickkontakt und lächelt
- ist höflich, freundlich, offen, ruhig, lächelt
- Lobt bei richtigen Antworten „richtig“
- spricht die SuS immer mit Namen an
 - weiß allerdings ein paar wenige Namen nicht sofort, guckt dann auf den Sitzplan
- Hört zu, wenn sich SuS im Unterricht melden und geht sofort auf die SuSmeldungen und Zwischenfragen ein
 - Rückfrage zum Vortrag, ob eine PPT-Präsentation gemacht werden darf: ja
 - Schülerfrage nach Material für Poster: gerne, aber rechtzeitig ansprechen

- Rückmeldung
 - Während des Schülerreferats füllt die Lehrperson ein Bewertungsprotokoll aus und gibt es im Anschluss dem Schüler mit der Aufforderung zu überprüfen, ob alles Angemerkte verstanden wurde (Note steht mit auf dem Zettel), Lehrperson fragt am Ende der Stunde noch einmal nach „Bist du mit dem Bewertungsbogen klar gekommen?“
 - Gibt dem Vortragenden ein Feedback: erst Positives (Fachbegriffe wurden in eigenen Worten erklärt), dann Punkte, die noch nicht so gut waren und verbessert werden können (Betonung ausbauen, Plakat einbinden)
 - „bitte korrigiert es“ (legt dazu eine Folie auf), „Haben das alle so?“, „Wir lassen das so stehen“
 - Gibt nicht immer Rückmeldung, ob eine Antwort falsch oder richtig ist bzw. geht gar nicht auf das Gesagte ein
- S merkt RS Fehler am Tafelbild an, die Lehrkraft bedankt sich und korrigiert sich
- Lernunterstützung und Lernhilfen bei Lernschwierigkeiten nicht offensichtlich

(3) KOMMUNIKATION

Verbales und vokales Verhalten

Sprache

- flüssige Sprache ohne unnötige Pausen
- klare und verständliche Sprache, nicht kompliziert, neue Begriffe/Fremdwörter werden erklärt
- vielgestaltige Sprache
- Suggestivfrage: „Das Plakat dient wozu?“ wird sofort von der Lehrperson beantwortet.

Sprechverhalten

- Variationsreiche Sprache hinsichtlich der Lautstärke, Betonung und Tonfall
- angemessene Geschwindigkeit
- enthusiastische Ausführungen
- „Ruhe jetzt“ etwas lauter und bestimmter
- Lehrperson beobachtet, dass S kein Heft hat, sondern lose Blättersammlung und fragt nach „Wo ist denn dein Heft?“, „Das haut mich um, du musst das nachtragen Mäuschen“
 - Spricht S später noch einmal an: Auftrag für nächste Woche ist alles zu übertragen
 - S antwortet plump
 - Lehrperson spaßig „Sonst rufe ich zu Hause mal an und Frage, was da so im Hause XY los ist. Willst doch keine Schande über deine Familie bringen.“
 - Ironie, alle lachen
 - Äußerungen und Reaktionen der Lehrperson sind nicht immer eindeutig interpretierbar
- Humor und Ironie
 - „ach Leute, ihr wisst, dass ihr die Zeichen machen könnt wie ihr wollt. Ihr könnt auch Totenköpfe zeichnen, aber ob ihr dann fertig werdet?!“
 - „jetzt endlich wach?“

Zusammenfassungen und Erklärungen

- das Wichtigste wird hervorgehoben
 - Das solltet ihr wissen/kennen/euch merken
 - Das kann abgefragt werden
- Ergebnisse einzelner Lernabschnitte werden zusammengefasst
 - „eine Menge Informationen“ deswegen wird ein zusammenfassendes Tafelbild erstellt mit der Aufforderung mitzuschreiben
 - Visualisierung an der Tafel: gut lesbar, große Schrift, übersichtlich, spricht während des Anschreibens nicht
- Lehrperson gibt z.T. Zusatzklärungen oder vervollständigt

Nonverbales Verhalten

- bei einzelnen Übungen verstärkt, z.B. beim Zahlenroulette: lächelt, nickt
- sicherer Stand
- angemessener Wechsel zwischen Sitzen und Stehen je nach Lehr-Lernform
- verbale und nonverbale Kommunikation sind kongruent
- freundlich, offen, Augenkontakt
- empfänglich für SuS-Reaktionen (Unruhe, Unkonzentriertheit) mit Anpassung (Blickkontakt, Aufforderung, Bewegung im Raum, mit Namen ansprechen)

(4) DENKFÖRDERUNG

Affektive Voraussetzungen

- die Lehrperson beobachtet alle Lernenden und hört aktiv zu
- die Lehrperson erfasst die Aktionen und Reaktionen der Lernenden schnell und geht wirksam auf sie ein
- spricht die Lernenden mit Namen an
- hält Blickkontakt
- versucht auch SuS anzusprechen, die sich nicht freiwillig melden „Den XY habe ich heute noch gar nicht gehört“
- bei fehlerhaften Antworten geht die Lehrperson meistens nicht drauf ein, lässt die Antwort stehen und fragt, wer eine andere Lösung hat
→ kaum Erklärungen und Unterstützungen bei Fehlern

Kognitive Voraussetzungen

- Lehrkraft fragt „Gibt es Fragen?“
- Vorwissen soll aktiviert werden „Wer möchte zusätzlich was sagen?“, „Worum ging es?“
- fordert selten Begründungen auf „Weißt du genaueres?“
- überprüft weniger, ob Zusammenhänge verstanden wurden
- meist fehlendes Auswerten falscher Antworten
- Beibehalten der Unterrichtsstruktur
- gibt wenig Anstöße zu divergentem Denken, Antworten meistens klar in eine Richtung gefordert
- zeigt eine kurze Animation über die Plattenbewegung am Whiteboard
→ fordert SuS auf darüber nachzudenken „Wer hat eine Ahnung, worum es geht?“

Metakognitive Voraussetzungen

- unterstützt bei Steuerung des Lernens (siehe Scaffolding)
- keine Reflexion der Denkprozesse oder –strategien beobachtbar oder auch Diskussion von Fehlern

Organisatorische Voraussetzungen

- sorgt für Interaktion unter den Lernenden z.B. durch Methodenwahl
- hält den Lernprozess durch Hinweise und Nachfragen in Gang
- visualisiert Ergebnisse (Folie, Tafelbild, Animation)

(5) AFFEKTIVE ASPEKTE

Soziale Integration

- SuS-Rückmeldung zum Schülervortrag: wertschätzendes Feedback, SuS sprechen sich direkt mit Namen an, SuS äußern Kritik (Lehrperson wartet den Prozess ab, gibt keine Zwischenrückmeldung)

- Einschalten und Einrichten des Whiteboards soll von S übernommen werden
 - lässt etwas lauter werdenden Austausch unter SuS über den Prozess zu (geht nicht darauf ein)
 - lässt sich bei der Bedienung von SuS helfen bzw. fragt nach
- gegenseitige Kontrolle zwischen Banknachbar:innen, ob Anweisungen der Lehrperson befolgt wurden
- Methode: Murrelphase
 - die SuS haben 2-3 Minuten Zeit sich die Theorie von Wegner (siehe Heft) anzusehen und sollten sich wesentliche Punkt einprägen

Empathie

- während des Schülervortrags sitzt die Lehrperson hinten in der letzten Reihe, schreibt Protokoll, Mimik und Gestik ist stets offen, freundlich, hält Blickkontakt, lächelt
- die Lehrkraft fragt bei SoS nach „Bist du noch müde?“, nach Beobachtung der nonverbalen Kommunikation → „JA“
- Echtheit: Die Lehrkraft korrigiert einen RS-Fehler auf ihrer Folie und kommentiert „ihr seht, ich mache auch Fehler“.

Fairness

- Bewertungsprotokoll, um allen SuS eine gleiche Benotungsgrundlage zu geben
- Demonstration der Lehrperson wie der Schüler betont
 - soll dem Schüler verdeutlichen, wie er spricht (er fragt nach, was die Lehrkraft damit meint, weil er ja betone)
 - kann aber auch als lächerlich machen wirken
- Lehrkraft warten, wenn SuS sagen, dass sie noch nicht fertig seien
 - erlaubt auch S, die nicht so schnell sind, ein Foto zu machen, falls sie es nicht in der Zeit schaffen das Tafelbild abzuschreiben

Regeln

- die Lehrkraft gibt eine Liste aus, auf der sich die SuS eintragen müssen, wenn sie Materialien oder Hausaufgaben vergessen haben
 - ein S war krank und hatte die Aufgaben nicht: hätten nachgeholt werden müssen
 - Lehrperson trägt noch nach, wenn sie merkt, dass jemand doch noch was vergessen hat (z.B. Heft)
- Lehrperson schaltet während der Doppelstunde eine kurze Trinkpause ein
- Tadel: ein S kommentiert „Bei dir läuft`s“ als ein anderer S sich die Nase putzen will
 - Lehrperson „Jetzt endlich wach“, dann wird der S zum Tafelwischen aufgefordert
- Verwarnungen: es wird sehr unruhig in der Klasse
 - Lehrperson restriktiv: „Jetzt ist es so weit“ und schreibt das Wort VERWARNUNGEN an die rechte Tafelseite
 - sofortige Ruhe
- während des Abschreibens des Tafelbilds soll Ruhe herrschen
 - wird es unruhig, spricht die Lehrperson entsprechenden S an und fragt, was los sei oder dreht sich einfach nur um und guckt in die Klasse
 - sofort wieder Ruhe

E. Dokumentation der Videoanalyse

Tabelle 46 Übersicht der Videoanalyse für das Schneiden der Vignetten

IBAS - Anpassung	Datum	Zeiten	Inhalt
Strukturiertheit ↓	25.01.18 I	00:25 – 05:10	
Strukturiertheit ↑	19.01.18 I	00:10 – 00:22	Rückblick
		00:50 – 01:39	Rückblick
		02:15 – 03:07	Überblick und Überleitung zur aktuellen UE
		03:08 – 03:46	Fahrplan der Stunde, Überleitung zum Video
		22:59 – 23:03	Überleitung und Rückblick
		23:18 – 23:21	Fahrplan Verweis
		23:40 – 23:45	Ausblick nächste Stunde
Strukturiertheit ↑	01.02.18 II	00:34 – 01:52	Orientierung, Thema einbindung, Stundeninhalt/ Vorgehensweise, Plakatregeln
Freiräume ↑	09.02.18 II	01:38 – 01:59	Wahlaufgaben
		02:38 – 03:49	Wahlaufgaben
		03:50 – 03:52	Wahl der Methode Einzel- oder Partnerarbeit
		04:23 – 04:48	Zusatzaufgaben
		04:52 – 05:09	Wahlfreiheit am S-Beispiel erklärt
		08:35 – 08:50	Wahlfreiheit
		10:28 – 11:30	Zusatzaufgaben -> Tafel notiert
		19.01.18 II	00:18 – 01:20 02:53 03:45
Freiräume ↓	02.11.17 I	00:39 – 01:10	Hausaufgaben Kontrolle -> Liste
		01:11 – 01:18	Aufforderung eines S bei MitSuS zu kontrollieren
		01:56	S soll sich in die Liste eintragen
		02:45 – 02:59	Ermahnung des S, hat Aufgabe (01:11) nicht gemacht S-Aufruf ohne Meldung
		08:17 – 08:20	S-Aufruf ohne Meldung
		09:22 – 09:39	

IBAS - Anpassung	Datum	Zeiten	Inhalt
Leistungs- differenzierung ↓	09.02.18 I	04:39 – 05:08	Test-Anmerkung, keine Unterschiede
Leistungs- differenzierung ↑	09.02.18 II	04:23 – 04:48	Zusatzaufgaben
		05:34 – 06:01	Hilfestellung
		10:28 – 11:30	Zusatzaufgaben an Tafel notiert
		21:51	S-Hilfestellung
		24:22	Zeit verlängert
		24:40 – 24:47	Zusatzaufgaben
		30:28	S-Hilfestellung
Anstrengungs-be- zogenes Feedback ↓	09.02.18 I	00:56 – 01:30	Aufforderung sich im Test anzustrengen, ordentlich zu schreiben, Mühe zu geben
Kognitive Aktivierung ↑	17.11.17 II	01:34 – 02:44	Thema Stockwerkbau – Übertrag tropischer Regenwald
		03:39 – 03:54	Folie Faultier
		04:24 – 04:49	Faultier und Stockwerkbau verknüpfen
		07:12 – 07:29	Faultier und Stockwerkbau verknüpfen
Interessantheit ↑	01.02.18 I	00:34 – 01:10	Vorstellung des Themas und der Methode
		25:30 – 32:16	Gruppenarbeitsphase
Individuelle Bezugsnorm- orientierung ↑	19.01.18 I	01:57 – 02:11	Klassenlob für Mitarbeit
		13:00 – 13:30	Klassenlob für Interesse, Vergleich letzter Film
		23:04 – 23:14	Klassenlob für Mitarbeit, S-Lob Vergleich zu sonst
Relevanz ↓	25.01.18 I	12:17 – 12:54	Vorstellung der Leitfragen – ohne Relevanz
Relevanz ↑	25.01.18 II	00:21 – 01:20	Relevanz des Themas, an S gerichtete Frage
		01:21 – 01:46	Vorwissen, Erfahrungen zum Thema
		03:44 – 04:25	u.a. Relevanz durch LP zusammengefasst
Relevanz ↑	19.01.18 II	34:10 – 37:30	Relevanz der Methode
Fehlerklima ↑	25.01.18 II	46:03 – 46:26	Fehler nicht schlimm
		50:30 – 51:43	Ermutigt, wiederholt Fehler nicht schlimm
		52:12 – 53:08	Diskussion, Hinterfragen, Hilfestellung positiv
Fehlerklima ↓	02.11.17 I	01:25 – 01:50	Fehler -> negative Konsequenzen
		08:21 – 08:28	S kann nicht antworten -> schlecht

F. ViviAn – Anleitung für Studierende

Das Videovignetten-Analysetool kurz ViviAn

I. Zur Anmeldung:

Sie finden das Tool unter der Homepage: https://dms.dienst.uni-landau.de/dms/vivian_neu/wordpress/

Eine ausführliche Anleitung zur Anmeldung gibt es unter folgendem Link: https://dms.dienst.uni-landau.de/dms/vivian_neu/wordpress/wp-content/uploads/2020/04/Wie_melde_ich_mich_bei-Vi-viAn_an.pdf

Die Anleitung kurz und knapp:

1. Registrierung auf der Plattform (Wer bereits registriert ist, geht weiter zu Schritt 3)
2. Datenschutzerklärung ausfüllen und hochladen unter „Mein Profil“
3. Kursanmeldung:
 Der Kurs = Ihre Gruppe
 Dozentin = Treiber
 Semester = SoSe20

Achtung: Die Anmeldung ist ab dem 06.05.2020 möglich! Ich muss Sie nach erfolgreicher Anmeldung und Überprüfung der Datenschutzerklärung erst noch freischalten (Ich werde ab dem 06.05. mindestens 2mal täglich die Anmeldungen bearbeiten.). Sobald Sie ein grünes Häkchen sehen, haben Sie Zugriff auf die Vignetten und können Vignette 1-7 bearbeiten.

II. Zum Tool:

Die Lernumgebung sieht folgendermaßen aus:



Bevor Sie mit dem Video starten, beachten Sie folgende Hinweise:

Das Video können Sie nur einmal sehen, weder vor- noch zurückspulen und auch nicht pausieren (so wie es im Unterricht auch ist). Die Sequenzen sind alle kürzer als 3 Minuten. Die Materialien in der Lernumgebung können Sie sich bereits vor dem Start des Videos ansehen und auch während des Videos darauf zugreifen. Am Ende des Videos klicken Sie bitte auf den Button „Analyseaufträge“, um mit der Bearbeitung dieser (nach unten scrollen) beginnen zu können.

Als Tipp: Am stabilsten funktioniert der Internet Explorer oder Google Chrome.

Lernumgebung ansehen:

Video starten:

Aufträge bearbeiten:

III. Zu den Analyseaufträgen:

Es gibt drei Arten von Analyseaufträgen.

1. Beschreiben: Hier beschreiben Sie die Komponenten des Unterrichts (durch ankreuzen).
2. Erklären: Einzelne Unterrichtshandlungen aus dem Video sollen auf Basis wissenschaftlicher Theorien und Befunde erklärt werden.
3. Vorhersagen: Die Wirkungen dieser Unterrichtshandlungen auf den weiteren Lehr-Lernprozess sollen vorhergesagt werden.

Alle Fragen (Items) werden über eine 4-stufige Antwortskala beantwortet. Sie haben die Möglichkeit zwischen stimmt genau, stimmt eher, stimmt eher nicht und stimmt gar nicht zu wählen.

Ein Beispiel:

1. Beschreiben: Die Lehrperson lässt Fehler analysieren.
2. Erklären: um den Schülerinnen und Schülern die Möglichkeit zu geben, aus den Fehlern zu lernen.
3. Vorhersagen: deshalb machen die Schülerinnen und Schüler später weniger Fehler.

Eine Lösung und Ihre Antworten erhalten Sie am Ende des Workshops. Bitte gedulden Sie sich.

Viel Spaß und Erfolg bei der Bearbeitung!

G. Itembeschreibung mit Musterlösung

Tabelle 47 Item-Beschreibung mit Musterlösung (Antwortmöglichkeiten der Rating-Items: 4 - stimmt genau, 3 - stimmt eher, 2 - stimmt eher nicht, 1 - stimmt gar nicht) durch Zuordnung der Partial-Credit-Punkte (0 - falsch, 1 - richtige Tendenz, 2 - richtig), Items mit * wurden nach der Raschskalierung entfernt

Vig	Code	Item	Pkt	Antwort
Dimension Beschreiben				
Itemstamm: Entscheiden Sie inwieweit die BESCHREIBUNGEN auf die gesehene Vignette zutreffen.				
Vignette 1	V1BS1*	Die Lehrperson unterstützt Schülerinnen und Schüler, wenn sie Hilfe benötigen.	0 1 2	1 oder 2 4 3
	V1BS2	Die Lehrperson lobt die Schülerinnen und Schüler, wenn sie sich anstrengen.	0 1 2	3 oder 4 2 1
	V1BS3	Die Lehrperson erläutert Fachbegriffe.	0 1 2	3 oder 4 2 1
	V1BS4*	Die Lehrperson bietet nach Schwierigkeitsgrad gestaffelte Zusatzaufgaben an.	0 1 2	1 oder 2 3 4
	V1BS5	Die Schülerinnen und Schüler erarbeiten sich einen Inhalt selbstständig.	0 1 2	1 oder 2 3 4
	V1BS6*	Die Lehrperson passt die Arbeitszeit an die Schülerinnen und Schüler an.	0 1 2	1 oder 2 3 4
Vignette 2	V2BS1*	Die Lehrperson betont, dass die Aufgaben von einigen Schülerinnen und Schülern gut bewältigt werden können.	0 1 2	1 oder 2 3 4
	V2BS2	Der Test beinhaltet differenzierte Aufgabenformulierungen.	0 1 2	3 oder 4 2 1
	V2BS3	Die Leistungserwartungen werden sowohl mündlich als auch schriftlich vermittelt.	0 1 2	3 oder 4 2 1
	V2BS4	Die Schülerinnen und Schülern werden über die Lernziele der einzelnen Aufgaben informiert.	0 1 2	3 oder 4 2 1
	V2BS5*	Die Lehrperson beschreibt ihr zugrundeliegendes soziales Bezugssystem der Leistungsbeurteilung.	0 1 2	1 oder 2 4 3
	V2BS6*	Die Lehrperson geht auf die Rückfragen der Schülerinnen und Schüler ein.	0 1 2	3 oder 4 1 2
	V2BS7*	Im Klassenraum herrscht eine gepflegte und aufgeräumte Ordnung.	0 1 2	3 oder 4 1 2
Vignette 3	V3BS1*	Die Lehrperson stellt einen Bezug zum Alltag der Schülerinnen und Schüler her.	0 1 2	1 oder 2 3 4
	V3BS2*	Die Fragen der Lehrperson sind klar und präzise formuliert.	0 1 2	1 oder 2 4 3
	V3BS3*	Der Unterrichtseinstieg ist hinsichtlich des Stundenablaufs informierend.	0 1 2	1 oder 2 4 3
	V3BS4*	Die Lehrperson lässt unterschiedliche Meinungen vergleichen.	0 1 2	3 oder 4 2 1
	V3BS5*	Der Umgangston ist respektvoll und wertschätzend.	0 1 2	3 oder 4 1 2

Vig	Code	Item	Pkt	Antwort
Vignette 4	V4BS1*	Die Fragen der Lehrperson sind klar und präzise formuliert.	0 1 2	1 oder 2 4 3
	V4BS2*	Die Lehrperson stellt einen Bezug zum Alltag der Schülerinnen und Schüler her.	0 1 2	1 oder 2 4 3
	V4BS3	Die Lehrperson hat nur einen geringen Sprechanteil.	0 1 2	3 oder 4 1 2
	V4BS4	Die Lehrperson fasst die Relevanz des Themas zusammen.	0 1 2	1 oder 2 3 4
	V4BS5*	Der Unterrichtseinstieg ist hinsichtlich des Stundenablaufs informierend.	0 1 2	1 oder 2 4 3
	V4BS6	Die Lehrperson lässt unterschiedliche Meinungen vergleichen.	0 1 2	3 oder 4 1 2
Vignette 5	V5BS1*	Die Schülerinnen und Schüler wählen für die Bearbeitung der Aufgaben zwischen zwei Sozialformen.	0 1 2	1 oder 2 3 4
	V5BS2*	Die Aufgaben und Materialien sind vorstrukturiert.	0 1 2	1 oder 2 4 3
	V5BS3	Die Lehrperson beschreibt die Lernziele.	0 1 2	3 oder 4 2 1
	V5BS4*	Die Auswahlmöglichkeiten sind begrenzt.	0 1 2	1 oder 2 4 3
	V5BS5	Das Vorwissen der Schülerinnen und Schüler wird aktiviert.	0 1 2	3 oder 4 2 1
	V5BS6	Die Lehrperson bietet den Schülerinnen und Schülern individuelle Hilfestellungen an.	0 1 2	3 oder 4 2 1
	V5BS7	Die Lehrperson lobt die Schülerinnen und Schüler, wenn sie sich trauen schwierige Aufgaben zu lösen.	0 1 2	3 oder 4 2 1
	V5BS8*	Die Schülerinnen und Schüler entscheiden frei und uneingeschränkt, welches Problem sie in der Stunde bearbeiten möchten.	0 1 2	3 oder 4 1 2
Vignette 6	V6BS1*	Die Lehrperson benennt das Thema der Stunde.	0 1 2	1 oder 2 4 3
	V6BS2	Die Schülerinnen und Schüler bringen sich aktiv mit eigenen Ideen ein.	0 1 2	3 oder 4 2 1
	V6BS3	Es werden Kommunikationsregeln für eine erfolgreiche Gruppenarbeit wiederholt.	0 1 2	3 oder 4 2 1
	V6BS4*	Die Lehrperson benennt die geplante Methode für die Bearbeitung der Aufgabe.	0 1 2	1 oder 2 3 4
	V6BS5*	Die Gruppenzusammensetzung wird durch die Lehrperson vorgegeben.	0 1 2	3 oder 4 2 1
	V6BS6	Die Schülerinnen und Schüler werden zum Nachdenken angeregt.	0 1 2	3 oder 4 2 1

Vig	Code	Item	Pkt	Antwort
Vignette 6	V6BS7	Die Lehrperson beschreibt die Lernziele der Unterrichtseinheit.	0 1 2	3 oder 4 2 1
	V6BS8*	Die Lehrperson gestaltet die Leistungserwartung transparent.	0 1 2	1 oder 2 4 3
	V6BS9*	Der Stundeneinstieg ist logisch aufgebaut.	0 1 2	1 oder 2 4 3
Vignette 7	V7BS1*	Das Vorwissen der Schülerinnen und Schüler wird aktiviert.	0 1 2	3 oder 4 1 2
	V7BS2	Die Fehler der Lehrperson werden tabuisiert.	0 1 2	3 oder 4 2 1
	V7BS3*	Die Lehrperson schaut bei Fehlern genervt.	0 1 2	3 oder 4 2 1
	V7BS4	Die Lehrperson nimmt auf gemachte Aussagen Bezug.	0 1 2	3 oder 4 1 2
	V7BS5*	Fehler werden als Teil des Lernprozesses gesehen.	0 1 2	1 oder 2 3 4
	V7BS6	Die Lehrperson korrigiert falsche Aussagen der Schülerinnen und Schüler sofort.	0 1 2	3 oder 4 2 1
	V7BS7*	Die Lehrperson formuliert die Fragen genau.	0 1 2	3 oder 4 1 2
Dimension Erklären				
Itemstamm: Beurteilen Sie, welche der folgenden ERKLÄRUNGEN bezüglich der Motivationsförderung theoretisch angemessen sind.				
Vignette 1	Die Lehrperson passt die Arbeitszeit an die Schülerinnen und Schüler an ,			
	V1E11	um den Schülerinnen und Schülern die Erreichung des Lernziels zu ermöglichen.	0 1 2	1 oder 2 3 4
	V1E12*	um den Arbeitsprozess der Schülerinnen und Schüler zu strukturieren.	0 1 2	3 oder 4 1 2
	V1E13	um jeder Schülerin und jedem Schüler die Möglichkeit zu geben, in ihrem oder seinem eigenen Tempo zu arbeiten.	0 1 2	1 oder 2 3 4
	V1E14*	um eine Differenzierung zu ermöglichen.	0 1 2	1 oder 2 3 4
	Die Lehrperson bereitet zusätzliche Hilfestellungen (für schwierige Aufgaben vor) ,			
	V1E21	um das Lernen zu erleichtern.	0 1 2	1 oder 2 3 4
	V1E22	um einen Anreiz zum selbstorganisiertem Lernen zu geben.	0 1 2	1 oder 2 4 3
	V1E23	um die Lernziele zu verdeutlichen.	0 1 2	3 oder 4 1 2
	Die Lehrperson staffelt Aufgaben nach unterschiedlichem Schwierigkeitsgrad ,			
V1E31*	um auf die individuellen Leistungen der Schülerinnen und Schüler einzugehen.	0 1 2	1 oder 2 3 4	

Vig	Code	Item	Pkt	Antwort
Vignette 1	V1E32	um das Lernmaterial lebendiger zu gestalten.	0 1 2	3 oder 4 2 1
	V1E33	um die Schülerinnen und Schüler kognitiv anzuregen, die Aufgaben entsprechend ihres Lernstandes zu bearbeiten.	0 1 2	1 oder 2 3 4
	V1E34	um ein angstfreies Arbeitsklima zu schaffen.	0 1 2	1 oder 2 4 3
	Die Lehrperson vergibt an die schnelleren Schülerinnen und Schüler nach Beendigung der Pflichtaufgabe weiterführende Aufgaben,			
	V1E41*	um den Schülerinnen und Schülern attraktive Lernanreize durch eine Differenzierung zu bieten.	0 1 2	1 oder 2 3 4
	V1E42	um den Schülerinnen und Schülern das Erreichen erweiterter Lernziele zu ermöglichen.	0 1 2	1 oder 2 3 4
	V1E43	um den Schülerinnen und Schülern die Möglichkeit zur Mitbestimmung zu geben.	0 1 2	3 oder 4 2 1
	V1E44*	um die Schülerinnen und Schüler bei zusätzlich erbrachter Leistung besonders loben zu können.	0 1 2	3 oder 4 2 1
Vignette 2	Angenommen die Lehrperson würde einen Test mit differenzierten Schwierigkeitsgraden erstellen , dann geschieht dies...			
	V2E11*	um die Zielerreichung aller Schülerinnen und Schüler in der Unterrichtsreihe vergleichend überprüfen zu können.	0 1 2	3 oder 4 2 1
	V2E12	um den Schülerinnen und Schülern eine Rückmeldung zu ihren individuellen Lernfortschritten zu geben.	0 1 2	1 oder 2 3 4
	V2E13	um die Selbstwirksamkeit der leistungsschwächeren Schülerinnen und Schüler bestärken zu können.	0 1 2	1 oder 2 3 4
	Angenommen die Lehrperson würde einen differenzierten Bewertungsmaßstab verwenden , dann geschieht dies...			
	V2E21	um eine soziale Bezugsnormorientierung zu ermöglichen.	0 1 2	3 oder 4 2 1
	V2E22*	um eine Vergleichbarkeit zwischen allen Schülerinnen und Schülern zu erlauben.	0 1 2	3 oder 4 1 2
	V2E23	um Rückschlüsse auf das allgemeine Leistungsvermögen der Klasse ziehen zu können.	0 1 2	3 oder 4 1 2
	V2E24*	um die Leistungsbereitschaft aller Schülerinnen und Schüler zu erhöhen.	0 1 2	1 oder 2 4 3
	V2E25	um den Schülerinnen und Schülern den Druck eines Leistungsvergleichs zu nehmen.	0 1 2	1 oder 2 3 4
Vignette 4	Die Lehrperson befragt die Schülerinnen und Schüler nach ihrem Vorwissen,			
	V4E11*	um den Schülerinnen und Schüler die Möglichkeit zu geben, ihr situationsbezogenes Interesse zu äußern.	0 1 2	1 oder 2 4 3
	V4E12	um den Einstieg entspannt und stressfrei zu gestalten.	0 1 2	3 oder 4 2 1
	V4E13*	um an das aktivierte Wissen der Schülerinnen und Schüler anknüpfen zu können.	0 1 2	1 oder 2 3 4

Vig	Code	Item	Pkt	Antwort	
Vignette 4	V4E14	um Diskussionen anzuregen.	0 1 2	3 oder 4 1 2	
	Die Lehrperson bespricht die Relevanz des Themas,				
	V4E21	um den Schülerinnen und Schülern die Nützlichkeit des Wissens für ihren Alltag aufzuzeigen.	0 1 2	1 oder 2 3 4	
	V4E22*	um das Vorwissen zu strukturieren.	0 1 2	3 oder 4 1 2	
	V4E23	um eine längere Arbeitsperiode einzuleiten.	0 1 2	3 oder 4 2 1	
	V4E24*	um das thematische Interesse zu wecken.	0 1 2	1 oder 2 4 3	
	V4E25	um die Erwartungen transparent zu gestalten.	0 1 2	3 oder 4 2 1	
Vignette 5	Die Lehrperson lässt die Schülerinnen und Schüler die Sozialform zur Bearbeitung der Aufgabe selbst wählen,				
	V5E11*	um den Schülerinnen und Schülern eine methodische Öffnung zu geben.	0 1 2	1 oder 2 4 3	
	V5E12*	um die sozialen Kompetenzen der Schülerinnen und Schüler zu fördern.	0 1 2	1 oder 2 4 3	
	V5E13*	um sich den Lernstoff entsprechend der Vorlieben aneignen zu können.	0 1 2	1 oder 2 4 3	
	V5E14	um die Schülerinnen und Schüler zu belohnen.	0 1 2	3 oder 4 2 1	
	Die Lehrperson stellt den Schülerinnen und Schülern zwei unterschiedliche Wahlpflichtaufgaben vor,				
	V5E21	um den Schülerinnen und Schülern die Möglichkeit zu geben, Aufgaben entsprechend der eigenen Stärken zu wählen.	0 1 2	1 oder 2 3 4	
	V5E22	um die Relevanz des Themas vielseitig zu verdeutlichen.	0 1 2	3 oder 4 1 2	
	V5E23	um den Schülerinnen und Schüler die Möglichkeit zu geben, interessenbezogene Aufgaben zu wählen.	0 1 2	1 oder 2 3 4	
	V5E24	um die Schülerinnen und Schüler aktiv beim Unterrichtsablauf mitwirken zu lassen.	0 1 2	3 oder 4 1 2	
Vignette 6	Die Lehrperson benennt das Thema der Stunde,				
	V6E11	um den Schülerinnen und Schülern die Möglichkeit zu geben, das Thema der Stunde in den Gesamtkontext einzuordnen.	0 1 2	1 oder 2 3 4	
	V6E12	um die Relevanz zu verdeutlichen.	0 1 2	3 oder 4 1 2	
	V6E13*	um den Aufbau einer gut strukturierten Wissensbasis anzubahnen.	0 1 2	1 oder 2 4 3	
	Angenommen die Lehrperson würde die Lernziele der Stunde beschreiben, dann geschieht dies...				
	V6E21*	um zur Strukturierung der Unterrichtsstunde beizutragen.	0 1 2	1 oder 2 4 3	
	V6E22	um die Schülerinnen und Schüler nicht zu überfordern.	0 1 2	3 oder 4 1 2	

Vig	Code	Item	Pkt	Antwort		
Vignette 6	V6E23	um einen persönlichen Bezug herzustellen.	0 1 2	3 oder 4 2 1		
	V6E24	um die Schülerinnen und Schüler die Möglichkeit zu geben, die Lernzielerwartungen nachzuvollziehen.	0 1 2	1 oder 2 3 4		
	Die Lehrperson beschreibt das methodische Vorgehen der Gruppenarbeit,					
	V6E31*	um eine klare Ordnung zu schaffen.	0 1 2	1 oder 2 4 3		
	V6E32*	um sich in die Methode hineindenken zu können.	0 1 2	1 oder 2 4 3		
	V6E33	um der Unlust zu Gruppenarbeiten entgegenwirken zu können.	0 1 2	3 oder 4 2 1		
	V6E34	um die Schülerinnen und Schüler bei der Unterrichtsgestaltung einbeziehen zu können.	0 1 2	3 oder 4 1 2		
	V6E35	um den Schülerinnen und Schülern kreative Freiräume zu geben.	0 1 2	3 oder 4 1 2		
	Die Lehrperson beschreibt die Leistungserwartung der Gruppenarbeit,					
	V6E41*	um einen Orientierungsrahmen zu geben.	0 1 2	1 oder 2 3 4		
	V6E42	um auf die individuellen Leistungen der Schülerinnen und Schüler eingehen zu können.	0 1 2	3 oder 4 1 2		
	V6E43	um eine transparente Beurteilung zu schaffen.	0 1 2	1 oder 2 3 4		
	V6E44*	um den Aufbau verlässlicher Arbeitsbeziehungen in der Gruppe zu fördern.	0 1 2	3 oder 4 2 1		
	Vignette 7	Die Lehrperson verdeutlicht, dass Fehler Teil des Lernprozesses sind,				
		V7E11	um Fehler nicht in Verbindung mit negativen Konsequenzen zu bringen.	0 1 2	1 oder 2 3 4	
V7E12*		um die Risikobereitschaft der Schülerinnen und Schüler zu erhöhen, sich auch an schwierige Aufgaben zu trauen.	0 1 2	1 oder 2 4 3		
V7E13		um den Schülerinnen und Schülern Aufgaben stellen zu können, bei denen es mehrere Lösungen gibt.	0 1 2	3 oder 4 2 1		
Angenommen die Lehrperson würde auf die Fehler der Schülerinnen und Schüler eingehen , dann geschieht dies...						
V7E21		um den Schülerinnen und Schülern die Möglichkeit zu geben, den Stoff wirklich zu verstehen.	0 1 2	1 oder 2 3 4		
V7E22		um die Vermittlung des Unterrichtsinhalts zu überprüfen.	0 1 2	3 oder 4 1 2		
V7E23*		um das Fehlerklima zu reflektieren.	0 1 2	3 oder 4 1 2		
V7E24	um anhand der Fehler aufzuzeigen, wie diese überwunden werden können.	0 1 2	1 oder 2 3 4			

Dimension Vorhersagen				
Itemstamm: Beurteilen Sie welche AUSWIRKUNGEN ...				
Vig	Code	Item	Pkt	Antwort
Vignette 1	die Anpassung der Arbeitszeit für den weiteren Lehr-Lernprozess hat.			
	V1V11	Deshalb erreichen die Schülerinnen und Schüler am Ende der Arbeitszeit die Erfüllung der gestellten Pflichtaufgabe.	0 1 2	1 oder 2 4 3
	V1V12	Deshalb langweilen sich die schnelleren Schülerinnen und Schüler am Ende der Unterrichtsstunde.	0 1 2	3 oder 4 1 2
	V1V13	Deshalb beeilen sich die Schülerinnen und Schüler beim nächsten Mal nicht mehr.	0 1 2	3 oder 4 1 2
	zusätzlich vorbereitete Hilfestellungen für den weiteren Lehr-Lernprozess haben.			
	V1V21	Deshalb erreichen die Schülerinnen und Schüler einen Lernfortschritt.	0 1 2	1 oder 2 4 3
	V1V22*	Deshalb werden Leistungsunterschiede für alle sichtbar gemacht.	0 1 2	3 oder 4 1 2
	V1V23	Deshalb denken die Schülerinnen und Schüler nicht selber nach.	0 1 2	3 oder 4 2 1
	V1V24	Deshalb nutzen die Schülerinnen und Schüler das Hilfsangebot voreilig.	0 1 2	3 oder 4 2 1
	V1V25	Deshalb fühlen sich einige Schülerinnen und Schüler benachteiligt.	0 1 2	3 oder 4 2 1
	die Staffelung der Aufgaben nach unterschiedlichem Schwierigkeitsgrad für den weiteren Lehr-Lernprozess hat.			
	V1V31	Deshalb wird die Selbstwirksamkeit der leistungsschwächeren Schülerinnen und Schüler erhöht.	0 1 2	1 oder 2 3 4
	V1V32	Deshalb gibt es eine negative Beeinträchtigung durch eine Über- oder Unterforderung.	0 1 2	3 oder 4 2 1
	V1V33*	Deshalb überschätzen sich die Schülerinnen und Schüler tendenziell.	0 1 2	3 oder 4 1 2
	V1V34	Deshalb sind die Schülerinnen und Schüler verunsichert.	0 1 2	3 oder 4 2 1
	die Verwendung weiterführender Aufgaben bei schnelleren Schülerinnen und Schülern für den weiteren Lehr-Lernprozess hat.			
	V1V41	Deshalb lösen auch die langsameren Schülerinnen und Schüler ihre Pflichtaufgaben schneller.	0 1 2	3 oder 4 1 2
	V1V42	Deshalb haben die langsameren Schülerinnen und Schüler weniger Zeitdruck bei der Bearbeitung der Pflichtaufgaben.	0 1 2	1 oder 2 4 3
	V1V43	Deshalb entwickeln die schnelleren Schülerinnen und Schüler einen besonderen Ehrgeiz sich zu beeilen.	0 1 2	3 oder 4 1 2
	Vignette 2	die Erstellung eines leistungsdifferenzierten Tests für den weiteren Lehr-Lernprozess hat.		
V2V11*		Deshalb schneiden alle Schülerinnen und Schüler im Test gleich ab.	0 1 2	3 oder 4 2 1
V2V12		Deshalb empfinden die meisten Schülerinnen und Schüler den Tests als ungerecht.	0 1 2	3 oder 4 2 1

Vig	Code	Item	Pkt	Antwort
Vignette 2	V2V13	Deshalb wird das Unterrichtsklima in der Klasse positiv beeinflusst.	0 1 2	1 oder 2 4 3
	V2V14	Deshalb wird die Selbstwirksamkeit der leistungsstärkeren Schülerinnen und Schüler verringert.	0 1 2	3 oder 4 2 1
	die Verwendung eines differenzierten Bewertungsmaßstabs für den weiteren Lehr-Lernprozess hat.			
	V2V21	Deshalb fühlen sich die leistungsschwächeren Schülerinnen und Schüler benachteiligt.	0 1 2	3 oder 4 2 1
	V2V22	Deshalb vergleichen sich die Schülerinnen und Schüler eher mit den Anderen aus der Klasse.	0 1 2	3 oder 4 1 2
	V2V23	Deshalb geben sich die Schülerinnen und Schüler bei der Lösung der Aufgaben weniger Mühe.	0 1 2	3 oder 4 2 1
Vignette 4	der Bezug zum Vorwissen für den weiteren Lehr-Lernprozess hat.			
	V4V11	Deshalb entwickeln die Schülerinnen und Schüler kritische Fragen.	0 1 2	3 oder 4 2 1
	V4V12	Deshalb erreichen die Schülerinnen und Schüler ein höheres kognitives Funktionsniveau.	0 1 2	1 oder 2 4 3
	V4V13*	Deshalb fokussieren die Schülerinnen und Schüler die Aufmerksamkeit auf die Thematik.	0 1 2	1 oder 2 3 4
	V4V14*	Deshalb wird das Anknüpfen an den neuen Lernstoff erleichtert.	0 1 2	1 oder 2 3 4
	die Besprechung der Relevanz des Themas für den weiteren Lehr-Lernprozess hat.			
	V4V21	Deshalb werden die Schülerinnen und Schüler animiert, selbstständig mehr zum Thema herauszuarbeiten.	0 1 2	1 oder 2 4 3
	V4V22*	Deshalb sind die Schülerinnen und Schüler motiviert, sich mit dem Thema zu beschäftigen.	0 1 2	1 oder 2 3 4
V4V23	Deshalb gelingt den Schülerinnen und Schülern der Transfer des neuen Wissens auf andere Situationen.	0 1 2	3 oder 4 1 2	
Vignette 5	die freie Wahl der Sozialform für den weiteren Lehr-Lernprozess hat.			
	V5V11*	Deshalb wird der Klassenzusammenhalt gestärkt.	0 1 2	3 oder 4 1 2
	V5V12	Deshalb arbeiten nahezu alle Schülerinnen und Schüler in Partnerarbeit.	0 1 2	3 oder 4 2 1
	V5V13	Deshalb verarbeiten die Schülerinnen und Schüler den neuen Lernstoff gut.	0 1 2	3 oder 4 1 2
	die Auswahl zwischen zwei Wahlpflichtaufgaben für den weiteren Lehr-Lernprozess hat.			
	V5V21	Deshalb wird die Lernbereitschaft der Schülerinnen und Schüler erhöht.	0 1 2	1 oder 2 3 4
	V5V22*	Deshalb haben die Schülerinnen und Schüler keine Angst davor, Fehler zu machen.	0 1 2	3 oder 4 1 2
	V5V23	Deshalb beginnen die Schülerinnen und Schüler die Bearbeitung der Aufgaben durch den Auswahlprozess deutlich später.	0 1 2	3 oder 4 2 1
V5V24	Deshalb setzt sich jede Schülerin und jeder Schüler mit einem Thema auseinander.	0 1 2	1 oder 2 3 4	

Vig	Code	Item	Pkt	Antwort
Vignette 6	die Benennung des Themas für den weiteren Lehr-Lernprozess hat.			
	V6V11*	Deshalb greifen die Schülerinnen und Schüler auf ihr Vorwissen zurück.	0 1 2	1 oder 2 4 3
	V6V12	Deshalb arbeiten einige Schülerinnen und Schüler nicht mit.	0 1 2	3 oder 4 1 2
	V6V13	Deshalb stellen die Schülerinnen und Schüler einen thematischen Bezug her.	0 1 2	1 oder 2 3 4
	die Beschreibung der Lernziele für den weiteren Lehr-Lernprozess hat.			
	V6V21	Deshalb können die Schülerinnen und Schüler die Lernziele, die sie am Ende der Unterrichtseinheit erreicht haben sollen, genau beschreiben.	0 1 2	1 oder 2 3 4
	V6V22*	Deshalb arbeiten die Schülerinnen und Schüler zielgerichtet.	0 1 2	1 oder 2 4 3
	V6V23	Deshalb erfüllen die Schülerinnen und Schüler nur die Mindestanforderungen.	0 1 2	3 oder 4 2 1
	V6V24	Deshalb können die Schülerinnen und Schüler die Lernzielerreichung überprüfen.	0 1 2	1 oder 2 3 4
	die Beschreibung des methodischen Vorgehens der Gruppenarbeit für den weiteren Lehr-Lernprozess hat.			
	V6V31	Deshalb arbeiten die Schülerinnen und Schüler effektiv.	0 1 2	1 oder 2 4 3
	V6V32*	Deshalb weiß jede Schülerin und jeder Schüler wie Vorgegangen werden soll.	0 1 2	1 oder 2 4 3
	V6V33	Deshalb halten sich die Schülerinnen und Schüler an die planmäßige Vorgehensweise.	0 1 2	1 oder 2 4 3
	die Beschreibung der Leistungserwartung der Gruppenarbeit für den weiteren Lehr-Lernprozess hat.			
	V6V41*	Deshalb orientieren sich die Schülerinnen und Schüler an den Vorgaben der Lehrperson.	0 1 2	1 oder 2 4 3
	V6V42	Deshalb wird die Selbstwirksamkeit der leistungsschwächeren Schülerinnen und Schüler während der Gruppenarbeit gestärkt.	0 1 2	3 oder 4 2 1
	V6V43	Deshalb haben die Schülerinnen und Schüler Angst, eine schlechte Note zu erhalten.	0 1 2	3 oder 4 2 1
	Vignette 7	Fehler als Teil des Lernprozesses zu sehen für den weiteren Lehr-Lernprozess hat.		
V7V11		Deshalb werden Schülerinnen und Schüler, die sich nicht verbessern, gelobt.	0 1 2	3 oder 4 2 1
V7V12*		Deshalb üben die Schülerinnen und Schüler Fehler aufzudecken und zu korrigieren.	0 1 2	1 oder 2 4 3
V7V13		Deshalb trauen sich die Schülerinnen und Schüler, ihre Vermutungen zu äußern.	0 1 2	1 oder 2 3 4
das Eingehen auf Fehler für den weiteren Lehr-Lernprozess hat.				
V7V21		Deshalb nutzen die Schülerinnen und Schüler Fehler zum Lernen.	0 1 2	1 oder 2 4 3
V7V22		Deshalb treten die gleichen Fehler weniger auf.	0 1 2	1 oder 2 4 3
V7V23	Deshalb haben die Schülerinnen und Schüler Angst, sich zu äußern.	0 1 2	3 oder 4 2 1	

H. Itemkennwerte

Tabelle 48 Itemkennwerte des 3-dimensionalen, finalen Partial-Credit-Modells (r_{pbis} – Punktbiseriale Korrelation (Trennschärfe), $M\theta_v$ – mittlere Fähigkeit einer Person der Kategorie c , δ – Schwellenparameter, Q3 – lokale stochastische Abhängigkeit (basiert auf einer Matrix von 104 x 104 Einträgen, daher werden nur die deutlich kritischen Werte angegeben), $M\sigma_i$ – mittlere Itemschwierigkeit, SE – Standardfehler der Itemschwierigkeit)

Item	c	r_{pbis}	$M\theta_v$	Outfit	Infit	δ	Q3	$M\sigma_i$	SE
V1BS2	0	-0,391	-0,508					0,033	0,213
	1	0,083	-0,033	0,958	0,960	-3,920			
	2	0,317	0,313	0,958	0,990	-1,132			
V1BS3	0	-0,296	-0,403					0,053	0,214
	1	-0,028	-0,097	1,072	1,018	-3,931			
	2	0,356	0,371	0,909	0,970	-1,102			
V1BS5	0	-0,050	-0,242					-1,669	0,324
	1	-0,092	-0,164	1,468	1,072	-5,508			
	2	0,109	-0,029	1,095	1,084	-2,866			
V2BS2	0	-0,429	-0,524					-0,628	0,203
	1	-0,150	-0,388	1,036	0,991	-2,556			
	2	0,483	0,156	0,965	0,961	-2,075			
V2BS3	0	-0,346	-0,347					0,012	0,189
	1	-0,105	-0,166	1,068	1,041	-2,932			
	2	0,457	0,351	0,947	0,951	-1,352			
V2BS4	0	-0,433	-0,958					-1,019	0,244
	1	-0,307	-0,284	0,662	0,905	-4,881			
	2	0,549	0,304	0,855	0,873	-2,210			
V4BS3	0	-0,303	-0,292					-0,156	0,187
	1	0,071	0,101	1,271	1,180	-1,963			
	2	0,265	0,098	1,280	1,197	-1,610			
V4BS4	0	-0,078	-0,233					-1,154	0,246
	1	-0,121	-0,179	1,525	1,083	-4,641			
	2	0,160	0,016	1,115	1,100	-2,434			
V4BS6	0	-0,318	-0,163					1,064	0,254
	1	0,139	0,315	1,043	1,031	-0,787			
	2	0,273	0,396	1,272	1,044	-0,387			
V5BS3	0	-0,338	-0,237					0,971	0,236
	1	0,160	0,093	0,999	1,004	-2,266			
	2	0,308	0,520	0,832	0,977	-0,352			
V5BS5	0	-0,372	-0,284					0,789	0,223
	1	0,191	0,103	1,007	0,995	-2,568			
	2	0,292	0,431	0,855	0,992	-0,514			
V5BS6	0	-0,433	-0,583					-0,314	0,200
	1	-0,226	-0,230	0,892	0,952	-3,859			
	2	0,615	0,473	0,769	0,831	-1,581			
V5BS7	0	-0,359	-0,799					-1,139	0,246
	1	-0,337	-0,362	0,814	0,941	-4,631			
	2	0,527	0,225	0,857	0,883	-2,417			

V6BS2	0	-0,496	-0,839				-0,940	0,221
	1	-0,223	-0,340	0,788	0,877	-3,747		
	2	0,554	0,196	0,851	0,882	-2,319		
V6BS3	0	-0,405	-0,369				0,073	0,188
	1	0,044	-0,017	1,038	1,019	-2,591		
	2	0,384	0,273	0,980	1,020	-1,320		
V6BS6	0	-0,335	-0,328				0,241	0,196
	1	-0,009	-0,076	1,058	1,024	-3,006		
	2	0,396	0,384	0,896	0,961	-1,080		
V6BS7	0	-0,362	-0,258				0,816	0,222
	1	0,114	0,050	0,998	0,981	-2,273		
	2	0,378	0,556	0,797	0,970	-0,529		
V7BS2	0	-0,144	-0,344				-1,173	0,249
	1	-0,160	-0,208	1,373	1,096	-4,502		
	2	0,234	0,065	1,073	1,053	-2,481		
V7BS4	0	-0,255	-0,125				1,221	0,287
	1	0,146	0,702	1,253	1,054	-0,368		
	2	0,213	0,338	1,466	1,076	-0,245		
V7BS6	0	-0,281	-0,427				-0,399	0,205
	1	-0,131	-0,155	1,006	1,019	-4,034		
	2	0,370	0,256	0,975	0,967	-1,647		
V1E11	0	-0,041	-0,183				-1,962	0,404
	1	-0,261	-0,302	1,329	1,031	-5,228		
	2	0,265	-0,027	0,977	0,989	-3,281		
V1E13	0	-0,230	-0,352				-0,975	0,235
	1	-0,015	-0,085	1,002	0,994	-4,499		
	2	0,150	-0,022	1,039	1,038	-2,247		
V1E21	0	-0,183	-0,272				-0,820	0,222
	1	-0,063	-0,107	1,043	1,021	-4,341		
	2	0,180	-0,005	1,024	1,014	-2,093		
V1E22	0	-0,097	-0,153				-0,430	0,199
	1	-0,148	-0,146	1,108	1,051	-3,838		
	2	0,228	0,034	0,982	0,989	-1,725		
V1E23	0	-0,322	-0,204				0,054	0,180
	1	0,335	0,279	0,981	0,985	-1,998		
	2	0,109	-0,024	1,104	1,091	-1,386		
V1E32	0	-0,459	-0,274				0,818	0,246
	1	0,304	0,050	0,923	0,929	-3,072		
	2	0,281	0,291	0,852	0,980	-0,366		
V1E33	0	-0,211	-0,341				-1,032	0,241
	1	-0,028	-0,092	0,955	1,000	-4,544		
	2	0,148	-0,024	1,028	1,019	-2,306		
V1E34	0	-0,159	-0,214				-0,689	0,205
	1	0,019	-0,065	1,069	1,042	-3,684		
	2	0,103	-0,040	1,078	1,070	-2,046		
V1E42	0	-0,176	-0,426				-1,606	0,323
	1	-0,070	-0,121	0,985	1,004	-5,187		
	2	0,138	-0,041	1,031	1,031	-2,866		

V1E43	0	-0,260	-0,198				0,362	0,204
	1	-0,016	-0,083	0,976	0,986	-3,082		
	2	0,355	0,221	0,881	0,927	-0,925		
V1E44	0	-0,378	-0,258				0,561	0,224
	1	0,204	0,005	0,944	0,948	-3,250		
	2	0,257	0,196	0,901	0,989	-0,644		
V2E12	0	-0,139	-0,267				-1,204	0,255
	1	-0,224	-0,204	0,970	1,005	-4,606		
	2	0,287	0,012	0,984	0,978	-2,500		
V2E13	0	-0,239	-0,240				-0,291	0,192
	1	-0,137	-0,139	1,028	0,994	-3,629		
	2	0,351	0,106	0,974	0,984	-1,598		
V2E21	0	-0,087	-0,106				0,854	0,235
	1	-0,088	-0,122	1,048	1,043	-2,581		
	2	0,291	0,274	0,852	0,980	-0,435		
V2E23	0	-0,310	-0,192				0,163	0,181
	1	0,111	0,026	1,017	1,015	-2,072		
	2	0,245	0,059	1,053	1,054	-1,265		
V2E24	0	-0,239	-0,246				-0,469	0,191
	1	0,138	0,003	1,037	1,002	-3,351		
	2	0,078	-0,050	1,073	1,065	-1,840		
V2E25	0	-0,244	-0,281				-0,498	0,203
	1	-0,064	-0,103	1,005	0,992	-4,036		
	2	0,255	0,049	0,985	0,989	-1,768		
V4E12	0	-0,087	-0,083				2,081	0,499
	1	0,049	-0,011	1,027	1,002	-0,709		
	2	0,104	0,261	0,779	0,998	0,700		
V4E14	0	-0,123	-0,098				0,690	0,215
	1	0,076	0,065	1,107	1,057	-1,039		
	2	0,093	0,003	1,212	1,079	-0,769		
V4E21	0	-0,146	-0,400				-1,767	0,352
	1	-0,087	-0,132	0,988	1,000	-5,275		
	2	0,139	-0,040	1,028	1,018	-3,042		
V4E23	0	-0,372	-0,213				0,736	0,225
	1	0,264	0,059	0,952	0,956	-2,765		
	2	0,186	0,135	1,016	1,016	-0,540		
V4E25	0	-0,085	-0,095				1,444	0,320
	1	-0,013	-0,080	1,048	1,027	-2,232		
	2	0,249	0,425	0,724	0,973	0,205		
V5E14	0	-0,261	-0,269				-0,302	0,198
	1	-0,136	-0,129	0,948	0,967	-3,912		
	2	0,372	0,135	0,912	0,940	-1,556		
V5E21	0	-0,048	-0,134				-1,219	0,252
	1	-0,174	-0,221	1,209	1,057	-3,920		
	2	0,181	-0,032	1,056	1,039	-2,609		
V5E22	0	-0,251	-0,161				0,067	0,177
	1	0,120	0,240	1,011	1,014	-1,506		
	2	0,217	0,023	1,033	1,035	-1,400		

V5E23	0	-0,325	-0,919				0,490	-2,012	0,418
	1	-0,156	-0,241	0,646	0,941	-4,900			
	2	0,283	-0,030	0,936	0,971	-3,383			
V5E24	0	-0,188	-0,105					0,876	0,240
	1	0,074	0,119	1,204	1,018	-0,737			
	2	0,169	0,086	1,384	1,039	-0,589			
V6E11	0	-0,048	-0,137					-1,085	0,245
	1	-0,159	-0,151	1,450	1,036	-4,701			
	2	0,180	-0,010	1,040	1,040	-2,338			
V6E12	0	-0,258	-0,153					0,240	0,182
	1	0,041	-0,005	0,996	1,004	-1,479			
	2	0,247	0,058	1,035	1,036	-1,219			
V6E22	0	-0,315	-0,198					-0,010	0,176
	1	0,152	0,128	0,995	0,994	-1,863			
	2	0,235	0,030	1,026	1,029	-1,462			
V6E23	0	-0,295	-0,211					0,383	0,202
	1	0,021	-0,065	0,976	0,980	-3,077			
	2	0,356	0,227	0,883	0,953	-0,902			
V6E24	0	-0,317	-0,784					-1,777	0,353
	1	-0,157	-0,184	0,615	0,940	-5,246			
	2	0,271	-0,012	0,967	0,966	-3,060			
V6E33	0	-0,215	-0,204					0,079	0,200
	1	-0,090	-0,108	0,984	0,992	-3,628			
	2	0,347	0,182	0,879	0,945	-1,152			
V6E34	0	-0,326	-0,208					0,061	0,178
	1	0,209	0,104	0,970	0,960	-2,258			
	2	0,174	0,015	1,106	1,056	-1,361			
V6E35	0	-0,329	-0,168					0,395	0,189
	1	0,165	0,134	0,965	0,998	-1,530			
	2	0,251	0,085	1,001	1,036	-1,053			
V6E42	0	-0,379	-0,312					-0,463	0,186
	1	0,209	0,108	0,950	0,966	-2,788			
	2	0,184	-0,011	1,079	1,067	-1,885			
V6E43	0	-0,202	-0,411					-1,441	0,288
	1	-0,254	-0,234	0,985	0,984	-4,848			
	2	0,336	0,016	0,974	0,983	-2,736			
V7E11	0	-0,211	-0,494					-1,775	0,373
	1	-0,134	-0,268	1,322	0,969	-4,004			
	2	0,240	-0,041	1,029	0,994	-3,204			
V7E13	0	-0,367	-0,173					1,094	0,257
	1	0,259	0,094	0,934	0,955	-2,126			
	2	0,228	0,243	0,883	0,998	-0,232			
V7E21	0	-0,107	-0,416					-2,016	0,445
	1	-0,135	-0,141	0,878	0,997	-6,395			
	2	0,161	-0,024	1,017	1,015	-3,025			
V7E22	0	-0,318	-0,152					0,448	0,194
	1	0,026	-0,005	1,008	1,009	-1,136			
	2	0,317	0,127	1,019	1,009	-1,019			

V7E24	0	-0,086	-0,177					-0,954	0,233
	1	-0,049	-0,096	1,105	1,026	-4,600			
	2	0,098	-0,036	1,041	1,039	-2,200			
V1V11	0	-0,303	-0,493					-0,941	0,263
	1	0,094	-0,001	0,803	0,963	-5,181			
	2	0,059	-0,005	1,059	1,049	-2,007			
V1V12	0	-0,359	-0,218					-0,043	0,181
	1	0,342	0,218	0,987	0,977	-2,621			
	2	0,064	-0,003	1,113	1,095	-1,444			
V1V13	0	-0,388	-0,203					0,041	0,179
	1	0,094	0,062	0,985	0,982	-2,082			
	2	0,329	0,126	1,008	1,016	-1,395			
V1V21	0	-0,357	-0,442					-0,901	0,227
	1	0,148	0,053	0,852	0,958	-4,151			
	2	0,086	-0,005	1,065	1,060	-2,223			
V1V23	0	-0,509	-0,369				0,431	0,027	0,199
	1	0,109	0,010	0,883	0,918	-3,613			
	2	0,401	0,250	0,877	0,933	-1,220			
V1V24	0	-0,366	-0,140				0,431	1,471	0,321
	1	0,296	0,146	0,950	0,956	-2,192			
	2	0,198	0,373	0,839	0,987	0,229			
V1V25	0	-0,377	-0,465					-0,638	0,229
	1	-0,011	-0,039	0,809	0,942	-4,637			
	2	0,259	0,106	0,968	0,980	-1,788			
V1V31	0	-0,208	-0,262					-0,468	0,232
	1	-0,044	-0,048	0,964	0,990	-4,679			
	2	0,199	0,096	0,997	1,001	-1,545			
V1V32	0	-0,350	-0,384					-0,645	0,213
	1	-0,032	-0,050	0,889	0,934	-4,152			
	2	0,275	0,088	0,973	0,976	-1,921			
V1V34	0	-0,340	-0,323					-0,145	0,215
	1	0,033	-0,024	0,893	0,954	-4,198			
	2	0,275	0,168	0,932	0,961	-1,278			
V1V41	0	-0,293	-0,189					-0,182	0,180
	1	0,177	0,136	1,010	1,014	-2,382			
	2	0,157	0,030	1,077	1,063	-1,612			
V1V42	0	-0,243	-0,222					-0,320	0,197
	1	0,087	0,006	0,994	0,984	-3,735			
	2	0,118	0,027	1,075	1,047	-1,613			
V1V43	0	-0,098	-0,060					0,570	0,203
	1	-0,026	-0,083	1,155	1,092	-1,144			
	2	0,117	0,049	1,170	1,078	-0,889			
V2V12	0	-0,375	-0,128				0,460	1,353	0,290
	1	0,226	0,131	0,934	0,947	-1,927			
	2	0,339	0,569	0,620	0,930	0,037			
V2V13	0	-0,324	-0,128				0,460	0,377	0,187
	1	0,114	0,157	1,010	0,989	-1,389			
	2	0,281	0,144	1,048	1,017	-1,079			

V2V14	0	-0,398	-0,279				0,422	0,209	0,205
	1	0,155	0,031	0,932	0,951	-3,583			
	2	0,267	0,197	0,968	0,981	-1,001			
V2V21	0	-0,335	-0,312					-0,304	0,203
	1	0,023	-0,020	0,919	0,967	-4,020			
	2	0,256	0,123	0,984	0,978	-1,533			
V2V22	0	-0,331	-0,185					-0,089	0,177
	1	0,103	0,101	0,987	0,997	-2,020			
	2	0,267	0,084	1,030	1,024	-1,536			
V2V23	0	-0,334	-0,242				0,422	0,177	0,204
	1	-0,011	-0,034	0,947	0,974	-3,609			
	2	0,396	0,298	0,910	0,955	-1,035			
V4V11	0	-0,036	-0,038					1,902	0,420
	1	0,045	0,020	1,081	1,036	-0,921			
	2	-0,012	-0,065	1,268	1,019	0,525			
V4V12	0	-0,087	-0,124					-0,759	0,215
	1	-0,171	-0,126	1,104	1,048	-4,084			
	2	0,222	0,055	1,006	1,006	-2,068			
V4V21	0	0,023	-0,008					-0,716	0,212
	1	-0,128	-0,101	1,284	1,073	-4,048			
	2	0,107	0,011	1,052	1,046	-2,024			
V4V23	0	-0,149	-0,060					0,884	0,237
	1	0,050	0,105	1,138	1,028	-0,727			
	2	0,136	0,099	1,212	1,043	-0,581			
V5V12	0	-0,179	-0,070					1,835	0,395
	1	0,175	0,108	0,987	0,996	-1,652			
	2	0,030	0,050	1,114	1,017	0,556			
V5V13	0	-0,059	-0,044					0,623	0,207
	1	0,027	0,099	1,255	1,075	-0,888			
	2	0,054	0,010	1,276	1,074	-0,847			
V5V21	0	-0,234	-0,443					-1,247	0,281
	1	-0,045	-0,052	0,890	0,970	-5,223			
	2	0,147	0,026	1,021	1,020	-2,404			
V5V23	0	-0,345	-0,180					0,617	0,211
	1	0,196	0,067	0,957	0,963	-2,858			
	2	0,222	0,196	0,973	1,003	-0,665			
V5V24	0	-0,206	-0,219					-0,379	0,207
	1	0,004	-0,029	1,003	1,003	-4,186			
	2	0,159	0,063	1,022	1,031	-1,585			
V6V12	0	-0,454	-0,456					-0,699	0,205
	1	0,302	0,163	0,809	0,912	-3,703			
	2	0,057	-0,010	1,097	1,078	-2,055			
V6V13	0	-0,104	-0,160					-0,798	0,235
	1	-0,128	-0,081	1,149	1,007	-4,728			
	2	0,191	0,064	0,996	1,000	-1,970			
V6V21	0	-0,231	-0,290					-0,771	0,221
	1	-0,054	-0,057	0,999	0,978	-4,357			
	2	0,201	0,055	1,008	1,009	-2,030			

V6V23	0	-0,397	-0,236				0,535	0,216
	1	0,241	0,068	0,939	0,949	-3,270		
	2	0,216	0,197	0,940	1,001	-0,672		
V6V24	0	-0,203	-0,458				-1,467	0,315
	1	-0,207	-0,135	0,809	0,982	-5,478		
	2	0,280	0,068	0,967	0,966	-2,614		
V6V31	0	-0,341	-0,334				-0,688	0,202
	1	-0,081	-0,090	0,969	0,967	-3,456		
	2	0,331	0,080	0,961	0,970	-2,072		
V6V33	0	-0,265	-0,329				-0,979	0,228
	1	-0,043	-0,065	0,914	0,978	-3,718		
	2	0,219	0,029	0,999	1,005	-2,365		
V6V42	0	0,186	0,036				1,310	0,296
	1	-0,165	-0,114	1,128	1,111	-2,616		
	2	-0,058	-0,141	1,438	1,036	0,137		
V6V43	0	-0,355	-0,266				0,014	0,196
	1	-0,047	-0,051	0,924	0,958	-3,599		
	2	0,419	0,263	0,891	0,938	-1,239		
V7V11	0	-0,368	-0,477				-0,760	0,230
	1	-0,107	-0,075	0,834	0,958	-4,653		
	2	0,334	0,132	0,952	0,957	-1,942		
V7V13	0	-0,049	-0,164				-1,962	0,418
	1	-0,258	-0,321	1,058	1,011	-4,844		
	2	0,260	0,011	0,940	0,971	-3,333		
V7V21	0	-0,170	-0,216				-0,688	0,218
	1	-0,059	-0,057	1,029	1,006	-4,372		
	2	0,171	0,049	1,033	1,029	-1,925		
V7V22	0	-0,202	-0,211				-0,444	0,202
	1	0,055	-0,006	1,005	1,002	-4,020		
	2	0,101	0,022	1,040	1,032	-1,706		
V7V23	0	-0,135	-0,123				-0,146	0,191
	1	-0,184	-0,118	1,023	1,021	-3,541		
	2	0,321	0,158	0,949	0,964	-1,443		

I. Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich diese Dissertation selbstständig verfasst habe und keine anderen als die darin angegebenen Quellen und Hilfsmittel verwendet habe. Ich habe keine entgeltliche Hilfe von Vermittlungs- oder Beratungsdiensten in Anspruch genommen. Alle wörtlich oder inhaltlich übernommenen Stellen habe ich als solche gekennzeichnet.

Ich versichere außerdem, dass ich die Arbeit nur in diesem Promotionsverfahren eingereicht habe und weder in gleicher noch in ähnlicher Form als Prüfungsarbeit im In- oder Ausland abgegeben habe.

Mir ist bewusst, dass ein Verstoß gegen die vorgenannten Punkte den Entzug des Dokortitels bedeutet und gegebenenfalls auch weitere rechtliche Konsequenzen haben kann.

Hinweis zur Nutzung von KI-gestützten Assistenten

Bei der Erstellung dieser Dissertation wurden punktuell generative KI-Werkzeuge (ChatGPT, Perplexity, DeepL) eingesetzt, dies ausschließlich als Formulierungshilfe einzelner Sätze und Wörter, vereinzelt Literaturrecherche sowie für Textübersetzungen. Die inhaltliche Verantwortung, die Struktur der Arbeit, sämtliche Analysen und Schlussfolgerungen liegen vollständig bei der Verfasserin. Die Nutzung erfolgte stets unter Wahrung wissenschaftlicher Integrität und dienten ausschließlich der Unterstützung im Schreibprozess.

Ort und Datum

Unterschrift