

## Visualizing quantum states: A pilot study on problem solving in quantum information science education

Jonas Bley<sup>1</sup>, Eva Rexigel<sup>1</sup>, Alda Arias<sup>1</sup>, Lars Krupp<sup>2,3</sup>, Steffen Steinert<sup>4,5</sup>,  
Nikolas Longen<sup>2</sup>, Paul Lukowicz<sup>2,3</sup>, Stefan Küchemann<sup>4</sup>, Jochen Kuhn<sup>4</sup>,  
Maximilian Kiefer-Emmanouilidis<sup>1,2,3</sup> and Artur Widera<sup>1</sup>

<sup>1</sup>Department of Physics and Research Center OPTIMAS, RPTU Kaiserslautern-Landau,  
Kaiserslautern, 67663, Germany

<sup>2</sup>RPTU Kaiserslautern-Landau, Department of Computer Science and Research Initiative QC-AI  
Kaiserslautern, 67663, Germany

<sup>3</sup>Embedded Intelligence, German Research Center for Artificial Intelligence,  
Kaiserslautern, 67663, Germany

<sup>4</sup>Faculty of Physics, Chair of Physics Education, Ludwig-Maximilians-Universität München  
(LMU Munich), Munich, 80539, Germany

<sup>5</sup>Department of Electrical and Computer Engineering, RPTU Kaiserslautern-Landau,  
Kaiserslautern, 67663, Germany

 (Received 14 March 2025; accepted 14 November 2025; published 19 December 2025)

In the rapidly evolving interdisciplinary field of quantum information science and technology, a major obstacle is the need to understand advanced mathematics to solve complex problems. Current findings in educational research suggest that incorporating visualizations into problem-solving settings can have beneficial effects on students' performance and cognitive load compared to relying solely on symbolic problem-solving content. Visualizations like the (dimensional) circle notation enable us to represent not only single-qubit but also more complex multiqubit states, entanglement, and quantum algorithms. In this pilot study, we aim to take an initial step toward identifying the contexts in which students benefit from the presentation of visualizations of single- and multiqubit systems in addition to mathematical formalism. For this purpose, we propose a set of test items and a comprehensive methodology to assess students' performance and cognitive load when solving problems. This is a pilot investigation with a large breadth of questions intended to generate hypotheses and guide larger-scale, more focused studies in the future. Specifically, we compare two approaches: using the mathematical-symbolic Dirac notation alone and using it in combination with the (dimensional) circle notation. In surveys in one-, two- and three-qubit systems, we gather qualitative data from five, five and two think-aloud interviews, identifying problems that students encounter and their problem-solving strategies. In addition, we analyze quantitative data (performance and cognitive load) from 23, 27, and 17 participants in surveys on one-, two-, and three-qubit systems recruited mainly from our quantum computing lectures. We find that most of the test items are appropriate for a heterogeneous target group, as they can differentiate between participants in terms of performance and time taken. In general, the A-B crossover structure of the study is suitable for investigating the benefits of visualization for problem solving with regard to length and feasibility. Future studies should, however, be narrower in scope, given the observable dependence on student characteristics and context, with particular interest lying in the further investigation of the Hadamard gate, the CNOT gate, and entanglement in multiqubit systems.

DOI: [10.1103/hdpv-frft](https://doi.org/10.1103/hdpv-frft)

### I. INTRODUCTION

The interdisciplinary nature of the field of quantum information science and technology (QIST) necessitates concerted efforts to educate a workforce with diverse academic backgrounds [1]. This implies that the quantum education community needs to find ways to introduce the field of QIST to a broad audience. As a foundational concept, qubit systems are inherently interdisciplinary and span all of QIST. The “qubit-approach” can therefore be

---

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

used to abstractly introduce a new, heterogeneous audience to the field. As a result, the question of how best to make this approach more accessible as possible is raised. Fortunately, existing didactical theory provides a foundation from which to draw ideas.

Visualization has been shown to be a useful tool in science, technology, engineering, and mathematics (STEM) education [2–7]. In QIST education in particular, visualization has been proposed to be useful for learning [8,9]. The field of QIST is based on the utilization of multiqubit systems, but the effectiveness of presenting visualizations of qubit states in problem solving, in terms of performance [10] and cognitive load [11], is currently unknown. Understanding the contexts and conditions under which learners benefit from such visualizations would enable us to provide QIST educators with empirically grounded recommendations for how to most effectively integrate visualizations into teaching. This is an important step toward solving the challenge of high heterogeneity of target groups in QIST education. It remains unclear whether, and under which conditions, learning and problem solving are supported by providing visualizations alongside the mathematical-symbolic representation [e.g., the Dirac notation (DN)] in QIST education. One theoretical construct that may offer explanatory insight is that of representational competence [12]. We identify three overarching questions in the context of visualizations in QIST education: In what contexts does incorporating visualizations support students' performance in task solving on quantum operations and entanglement? How does the inclusion of visualizations affect cognitive load? How are performance and cognitive load affected by representational competence?

In order to answer these questions, appropriate and validated test instruments on task solving in QIST are needed. However, validated test instruments in the context of QIST education, and in particular on quantum operations and entanglement, are scarce in single-qubit systems and do, to our knowledge, not exist for multiqubit systems. As an essential first step in the investigation of these questions, we conducted a pilot study validating relevant test instruments to investigate the overarching questions. In this way, we initiate an investigation into whether visualizations provide a benefit for solving problems in QIST when provided in addition to the DN and under what circumstances such benefits occur. The instruments, survey design, and preliminary results presented here lay the groundwork for future, more targeted studies that can build on this foundation.

The paper is structured as follows. First, we introduce the research background in quantum education and the visualizations used in Sec. II. We follow with a description of the study design in Sec. III and describe the creation of the test items in Sec. IV. The preliminary results are summarized, and the test instruments are discussed with respect to their suitability to the target group and indications of effects

worth exploring further in Sec. VI. We conclude and provide an outlook for future research in Sec. VIII.

## II. RESEARCH BACKGROUND

The research background of this study includes the functions that multiple external representations (MERs) serve (Sec. II A), an overview of educational research on representations in quantum education and visualizations of multiqubit quantum states and entanglement. This encompasses the two representations investigated in the present study, Circle notation (CN) and dimensional circle notation (DCN) (Sec. II B), as well as the associated measures, including cognitive load, multiple-choice test design, and time taken (Sec. II C).

### A. Learning and problem solving with visualizations in QIST

Based on educational research, it is well known that learning and problem solving in STEM can be supported by focusing not only on text-based and mathematical-symbolic representations (e.g., written text and formulas) but also on visualizations (e.g., pictures and diagrams) [13,14]. From a theoretical perspective, the benefit of learning with text accompanied by visualization can be explained by more efficient information processing and the dual-channel structure of the working memory [15,16]. When learning with Dirac notation alone, the processing load of information is solely on the verbal channel of working memory. On the other hand, when learning with text and visualization, both the verbal and visual channels of working memory are activated and integrated in the construction of coherent mental schemata [15,16]. This distributed information processing across two channels reduces the load on each single channel and, thus, the risk of cognitive overload.

Within her Design, Functions, Tasks (DeFT) framework, Ainsworth [17] has formulated three main functions of MERs to support learning. First, MERs can *complement each other* by containing different information or supporting different processes. Second, they can *constrain each other*, e.g., by familiarity of the learners with a subset of the MERs or inherent properties. Third, incorporating MERs can *construct a deeper understanding* as learners are confronted with the abstraction of the underlying knowledge structures, the extension of knowledge to an unknown representation, or the improvement of the understanding of the relations between different representations.

However, incorporating additional representations can also impose additional cognitive load on students, as they must understand not only how each representation conveys information but also how to translate between them. Therefore, the learning effectiveness of MERs does not only rely on the learning material but also on the characteristics of the learners [17,18]. To learn and solve problems

with MERs efficiently, students must possess representational competence [5,12,19]. Rau divides representational competence with multiple visual representations into three domains [5]:

Conceptual competencies: *visual understanding*, the ability to connect the representation to concepts and the understanding of the main features, and *connective understanding*, the ability to connect and compare the representation with other representations.

Perceptual competencies: *visual fluency*, the efficiency of connecting the representation to concepts and the ability to work effectively within the representation, and *connective fluency*, the efficiency and flexibility of connecting multiple representations to each other and switching between them.

Meta-representational skills: the ability to choose suitable representations based on task demands, context, and one's own ability level and personal goals.

### B. Representations in QIST education

Previous research has focused on learning quantum physics with representations such as the Dirac and vector and matrix notation of single-qubit states and operations, considering metarepresentational competencies like assessing the suitability of these representations [20]. The authors found that, rather than focusing solely on students' *learning difficulties*, examining how representations are used to solve *tasks* can yield valuable insights into more effective teaching approaches. Extending this line of research to include visualization techniques represents a natural next step.

Because of the field's complexity and counterintuitive nature, visualizations are widely used in teaching and learning quantum physics. Several have been investigated for their educational efficacy and associated difficulties. Common visualizations of qubit states, such as the unit circle, the Bloch sphere, coin analogies, and quantum circuits, are summarized in Ref. [21]. Different visualizations of a beam splitter were examined in Ref. [22], leading to an interactive visualization tool of the Mach-Zehnder interferometer [23], an optical implementation of a single-qubit system. Student difficulties with the Mach-Zehnder interferometer when learning with an interactive visualization were also analyzed in Ref. [24]. One of the most prominent visualizations of single-qubit systems is the Bloch sphere, which represents quantum operations as rotations and projections in three-dimensional space. Difficulties related to the Bloch sphere, the concept of a basis, and the measurement process were analyzed in Ref. [25]. Their tutorial improved student performance regarding measurement probabilities and outcomes, demonstrating the Bloch sphere's utility in identifying equivalent states (up to a global phase) and visualizing single-qubit unitary operations as rotations. In Ref. [26], a new learning sequence using the Mach-Zehnder interferometer was proposed, combining polarization

visualizations, an interactive GeoGebra animation, a graphical experiment setup, and abstract representations. Such studies highlight the diversity of representations even within a narrow context like the Mach-Zehnder interferometer.

### 1. Representations of multiqubit system

QIST relies heavily on linear algebra in complex Hilbert spaces [27,28], where quantum states can be represented as vectors, and quantum operations correspond to unitary operations or projections. As systems expand to multiple qubits, these relationships become increasingly abstract and difficult to grasp. To address this challenge, various visualization techniques have been developed. Advanced visualizations include geometric depictions of quantum states, entanglement [27,29–31], and multiqubit states [32–35]. The diagrammatic ZX-calculus [36–39] describes operations and algorithms in a “LEGO-like” manner [40], simplifying standard quantum circuit notation. Color code visualizations [41–44] depict quantum error correction algorithms on hypercubes or hypercubelike lattices. The DROPS representation visualizes quantum states and entanglement via generalized Wigner functions [45–47] and has evolved into the BEADS representation [48], a more intuitive extension of the Bloch sphere to multiqubit systems.

While these visualizations are basis-independent, our work focuses on the computational basis, which is commonly used for introductory purposes [49]. Dynamic visualizations of quantum states and circuits, such as the “quantum puzzle game” based on state diagrams [50–52], exemplify this trend. The circle notation (CN) is a visual representation of the complex amplitudes of states in the computational basis using circles with gauges [53]. While in CN, amplitudes are aligned in a row as you would when writing states in the Dirac notation, qubits can also be assigned axes in space [54,55]. This dimensional approach has the possible benefit of making quantum operations more intuitive and can be combined with CN [then called dimensional circle notation (DCN)]. It can be used to visualize the entanglement properties of qubit systems [56].

Currently, there is little empirical evidence on the effectiveness of different visualizations of multiqubit systems for learning or problem solving. Because learning and problem solving share cognitive processes [10], differing only in the decision-making stage [57], and in a similar line of thought to Ref. [20], insights from problem-solving studies may improve instruction by identifying when and how to employ visualization techniques. Enhanced problem-solving performance could, in turn, lead to deeper conceptual understanding of underlying concepts like superposition and entanglement, which are known to show intrinsic susceptibility to misunderstandings [24,58].

Referring to the DeFT framework, using visualizations to support problem-solving leverages their complementary function, adding visual access to qubit characteristics and

gate operations. When students are presented with both visual and symbolic representations, they gain multiple new routes to understanding QIST concepts. In addition to this, the use of visualization facilitates the understanding of the corresponding symbolic descriptions. When using both visual and symbolic representations, students can be supported in extending existing knowledge structures to mathematical descriptions. For example, in the context of entanglement, dimensional visualizations like DCN provide new representations of the separability of qubit systems or even subsystems [56] (see also Sec. IV D). While CN is merely a representation of complex numbers, DCN adds an additional layer of information by spatially relating the basis states. When conducting a study comparing CN and DCN, one may be able to distinguish the benefits of both of these functions separately. If it is found out that benefits lie merely in the introduction of dimensionality, one could conclude that putting complex numbers on the edges of a cube is sufficient, instead of also using the circles.

However, visualizations also have limitations. In one-qubit systems, the Bloch sphere and the generalization to multiqubit systems in the form of the QuBeads representation [48] might make operations and entanglement more intuitive than CN by relating them to rotations in space and could therefore be used to complement mathematical notation more effectively. Second, these representations take up a lot of space and can display information only approximately, making mathematical symbolism and the circuit notation that can give an overview of a complete quantum algorithm, the more concise choice to constrain interpretation. Especially in systems beyond three qubits, and especially DCN takes up a large amount of space due to the additional axes that are presented and the need for spatially separating the cubes. CN does not have this drawback to this extent, e.g., one can find eight-qubit systems being visualized on two thirds of a page in a book in Ref. [53]. Furthermore, visualizations depict numeric relationships but cannot replace algebraic proofs, which remain essential for formal understanding. Thus, we advocate for their use in conjunction with, rather than as replacements for, symbolic representations, in order to fulfill the functions of the DeFT framework.

## C. Relevant outcome measures and test construction

### 1. Cognitive load

It is essential to consider the cognitive load imposed on learners when designing tasks and learning materials, as working memory capacity limits both learning and problem solving [11]. According to cognitive load theory, cognitive load comprises three components: extraneous [extraneous cognitive load (ECL)], intrinsic [intrinsic cognitive load (ICL)], and germane [germane cognitive load (GCL)] load [59]. ECL arises from presentation design; reducing it frees resources for goal-relevant processing. ICL reflects inherent task complexity and can be reduced only through

simplification or scaffolding. GCL refers to the cognitive effort invested in schema construction. Effective instructional design should therefore minimize ECL and maximize GCL.

When incorporating a visualization in addition to the DN, learner characteristics can determine the imposed cognitive load and whether the visualization helps or not. For novices, visualizations may reduce ICL by providing concrete anchors, whereas for experts, redundant visuals may increase ECL. This phenomenon is known as the *expertise reversal effect* [60]. Conversely, learners with strong representational understanding—e.g., of gate operations or separability in DCN—may derive greater benefit. Such differences exemplify aptitude–treatment interactions [61].

### 2. Multiple-choice test item design

Multiple-choice items were chosen for this pilot study due to their practical advantages over open-ended questions [62]. They are efficient, easily scored, and suitable for analyzing response time without confounding factors such as writing speed. When validated, they offer a reliable assessment while leveraging the testing effect, which enhances retention [63].

Designing effective multiple-choice items requires attention to distractors—plausible incorrect options that reflect common misconceptions [63,64]. In this study, such misconceptions include confusion about qubit order, misinterpretation of measurement probability (as magnitude rather than squared magnitude) [65], conflation of basis states with qubits [66], and confusing separability with entanglement [67]. If distractors do not follow common misconceptions, they should share characteristics with the correct answer or the other distractors in order to be competitive [68]. Distractors should be designed in a way that leaves no hint toward the correct answer without reading the question.

### 3. Time taken

Response time provides additional insights into task difficulty. According to the distance–difficulty hypothesis [69,70], the logarithm of the response time is inversely proportional to the difference between a student’s ability and an item’s difficulty: students spend more time on tasks close to their ability level. For example, in Ref. [71], it was confirmed that in complex tasks, students who took longer also performed better. Similarly, it was found that children with higher levels of ability take longer to incorrectly solve the questions [72]. When examining the effect of visualizations, longer response times for items with visualizations may indicate that the visualization engages additional problem-solving strategies and aligns task difficulty with the student’s ability.

When only considering the time taken for correctly solved questions, most of the results where the question was above the skill level of the participant are eliminated (except for those questions that were correctly solved by

random chance). Therefore, analyzing the time taken exclusively for correctly solved questions focuses on students whose ability matches or exceeds the question's difficulty. If a question becomes trivial for a student when presented with a visualization, they will take less time to correctly solve questions. Consequently, when requiring that both the question without and with visualization be solved correctly, the average time taken may be lower if the visualization indeed simplifies the question.

The time taken for task solving is a multidimensional construct. Clear results are expected to be very difficult to obtain, as there are many factors that play a role. Nevertheless, according to the distance-difficulty hypothesis [69,70], time could serve as an indicator of how well a task's difficulty aligns with student ability. For comparisons between questions to be meaningful, the questions should be of similar complexity. Furthermore, a detailed analysis of the relationship between question difficulty and the time taken for correct answers could offer additional insights into the benefits of providing visualization—for instance, if a difficult question requires more time with visualization than without, it might indicate that the visualization supports the use of deeper or alternative problem-solving strategies.

### III. RESEARCH QUESTIONS

Based on current educational research, we expect that the use of visualizations such as circle notation (CN) and dimensional circle notation (DCN) can leverage the known advantages of learning with Multiple External Representations (MERs) in Quantum Information Science and Technology (QIST) by constructing deeper understanding of entanglement and by complementing the mathematical-symbolic DN by offering new strategies to predict the outcome of quantum gate operations. In this study, we take the first step in the design and validation of test items that can be used to verify this expectation by comparing the task-solving abilities on items presented in DN with those in which DN is accompanied by CN or DCN. Such tests should take into account the representational competence of the participants. This competence is divided into conceptual competencies related to visual and connectional understanding and the perceived cognitive load during tasks without and with visualization [5]. Here, we refer to representational understanding instead of visual understanding, including competence with complex numbers in mathematical-symbolic notation. It is not clear whether this is an adequate generalization—namely, whether Rau's construct of representational competence, adapted to incorporate representations beyond those that are purely graphical, still predicts performance increases. In this work, we focus on conceptual competencies rather than perceptual or metarepresentational competencies for two reasons: we discard perceptual competencies for practical reasons, as it is unclear how to measure the fluency or

efficiency described by Rau independently of the accuracy of solving the questions. Efficiency is conceived as performance over effort, and there are multiple possible ways to define both [73], while accuracy should also be part of this metric. As the representations used in this work do not differ in level of abstraction or in the information presented, we see metarepresentational competencies as less important for solving the given problems. In the framework used, we remain with competencies involving connecting representations and understanding their functioning.

Another point that is unclear is whether this framework holds in task solving as well as in learning. However, there is a strong argument that task solving and learning share many similarities, except for differences in the decision-making phase [57].

Because the supposed benefits are likely not only learner dependent but also context dependent, we explore the effects of presenting a visualization in different contexts, namely measurements, various unitary operations, and entanglement and separability. The aim of this pilot study is to create test items and a suitable study design to answer the overarching questions described in Sec. I of the conditions for benefits of visualization for problem solving in QIST. To this end, we formulate the following research questions.

**RQ1:** Are the proposed test items suitable, in terms of difficulty, for a heterogeneous target group?

**RQ2:** Which test items should be omitted or redesigned?

**RQ3:** Is the study design suitable for answering the overarching questions?

In addition, we aim to identify indications of effects related to the overarching questions that warrant further exploration in future studies.

### IV. METHODS

In order to gain insights into the effects and relevant conditions of including visualizations like CN and DCN in students' problem solving, we decided to compare the performance and cognitive load of participants with and without visualization within participants (intrapersonally) and relate them to representational competence. As is typical for a pilot study, our aim is to learn about the suitability of the test items that are presented to the participants and the study design. All test items were reviewed by experts using the design principles described in Sec. II C. They are provided in Supplemental Material [74], with the correct answers highlighted. The order of the answers was randomly assigned to each participant. In Sec. IV A, we define the overall structure of the pilot study. Some students volunteered for think-aloud interviews, the methodology of which we describe in Sec. IV B. We then describe the group of participants in Sec. IV C. In Sec. IV D, we describe the introductory slides used. In Secs. IV E and IV F, we outline the design of the questions on representational and connectional understanding.

In Sec. IV G, we describe the test items of the main parts of the survey. We finish with the cognitive load test items in Sec. IV H.

### A. Structure of the study

The main study consisted of three online multiple-choice tests on one-, two-, and three-qubit systems, respectively. Each of the tests took about 45 min.

In each survey, except for the first, the participants were randomly assigned to two groups, where one group was presented with CN and the other group was presented with DCN (in one-qubit systems, CN and DCN are identical). Both CN and DCN were included in the study for the following reason: by comparing CN and DCN in a study such as this, one may be able to investigate the effects of visualizing complex numbers separately from those of introducing dimensionality.

Each survey was separated into two parts, A and B, where questions using the same contexts (gate operations, measurements, entanglement, and separability) were asked in both parts, while all questions were identical within one part (with or without visualization). The participants were divided into two further groups, independent of the CN and DCN group: one group was presented with visualization in part A and without visualization in part B (group vis-math), while the order was reversed for the other group (group math-vis). The same questions were asked to all students in part A, but one group received them with visualization and one without; in part B, the roles were reversed. This resulted in two independent groups in the one-qubit survey and four independent groups in the two- and three-qubit surveys. After each part, we measured the perceived cognitive load. Before the surveys, participants viewed introductory slides explaining the respective visualization, operations, and entanglement concepts and answered questions assessing representational competence. At the end, we collected demographic information (age, level of education, and field of study). The overall structure of the surveys is shown in Fig. 1.

### B. Qualitative part of the study and participants

To discover which problem-solving strategies students use and which problems they encounter with or without

visualization, we conducted think-aloud interviews with students from our quantum computing lecture [75]. This allowed us to identify reasons for differences in performance and cognitive load when visualization was provided. The performances of the interviewed students were included in the main survey results, but not their response times, as they required additional time to explain their thought processes.

Some students were recruited specifically for these think-aloud interviews. They were instructed to think about the question, select an answer, and then explain their reasoning. Participants received a compensation of 15€ for the one- and two-qubit surveys and 10€ for the three-qubit survey.

Five participants volunteered for think-aloud interviews in the one- and two-qubit surveys, two of whom also participated in the three-qubit survey. Of these participants, three were aged 22–25 and two were 26–30. Two identified as female and three as male. Three participants had an information and computer science background and two came from mathematics; one held a master’s degree and four a bachelor’s degree. All five had been introduced to both DN and CN before the survey. This group can be regarded as representative of the broader participant cohort in terms of academic background. The participants in the qualitative study were randomly assigned the conditions shown in Table I.

### C. Participants of the quantitative part

Participants in the main study received 10€ compensation per test. Most participants were drawn from our quantum computing lecture. In the surveys on one-, two-, and three-qubit systems, there were 18, 22, and 15 participants, respectively (excluding the participants of the qualitative part). At the end of the first survey, demographic information was collected. Table II shows the number of students assigned to the different conditions, including both general study participants and interview participants.

Of the 18 participants in the first survey, 8 reported that they had not been introduced to DN before and 9 reported prior exposure to CN. Among the participants in the one-, two-, and three-qubit studies, 1(1)(1) was aged 18–21, 12(11)(9) were 22–25, and 3(2)(3) were 26–30. Furthermore, 16(14)(13) were male and 1(1)(1) female.

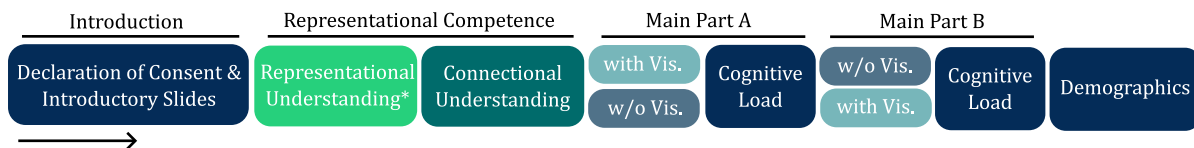


FIG. 1. Overall structure of the study. In the two- and three-qubit surveys, the participants were first grouped into circle notation or dimensional circle notation randomly. During all surveys, the participant groups were split during the main parts of the studies into group vis-math or group math-vis, depending on whether they were presented with or without visualization first. The questions within one of the parts A or B only differed in whether they were with visualization or not. This results in two groups in the one-qubit survey and four groups in the two- and three-qubit surveys. \*Representational understanding was only tested in the one-qubit survey.

TABLE I. Assignment of conditions in the qualitative analysis to groups assigned questions with visualization first or second and assigned to circle notation or dimensional circle notation in the case of the two- and three-qubit surveys.

Student	One qubit	Two qubits	Three qubits
1	Visualization first	CN, visualization first	
2	Visualization first	CN, visualization first	DCN, no visualization first
3	Without visualization first	DCN, without visualization first	
4	Without visualization first	DCN, visualization first	DCN, visualization first
5	Without visualization first	DCN, without visualization first	

Regarding educational level, 4(3)(4) participants had completed high school, 9(7)(5) held a bachelor's degree, 3(3)(3) held a master's degree, and 1(1)(1) held a Ph.D. In terms of academic field, there was 1(1)(1) in economics, 8(8)(7) in information/computer science, 3(3)(2) in computer science and engineering, 4(4)(4) in mathematics, and 1(1)(1) in physics. Of these 18 participants, 13 completed all three surveys, two completed only the first, two completed the first and second, and one completed the first and third. We do not have personal information for those who participated only in the second and/or third survey, but we assume that they were familiar with DN and CN, as they were also recruited from our quantum computing lecture. Most participants were new to quantum science, as it is not typically included in mathematics or computer science curricula, and the quantum computing lecture was introductory in nature. Except for the students in physics and the Ph.D. holders, we assume that all participants had taken no more than one or two QIST-related courses.

#### D. Introductory slides

At the beginning of each survey, participants were presented with introductory slides that explained all key ideas necessary to solve the study questions, using both mathematical symbolism and the corresponding visualization, CN or DCN, depending on their assigned condition.

TABLE II. Assignment numbers of students to the two or four conditions of the different surveys, to groups assigned questions without visualization first or second, and circle notation or dimensional circle notation as visualization in the case of the two- and three-qubit surveys. Comparisons are made both between groups of the same column and within participants. In brackets are the numbers of the qualitative part of the survey. These are also included in the quantitative part.

Number of qubits	One		Two		Three	
	CN	CN	DCN	CN	DCN	
Without visualization first	12(3)	6(1)	8(1)	5	3(1)	
Visualization first	11(2)	6(2)	7(1)	5	4(1)	
Sum	23(5)	12(3)	15(2)	10	7(2)	

Participants read through the slides at their own pace. The introductory slides covered the following topics:

- Visualization of qubit states.
- Action of single-qubit gates in single-qubit systems.
- Action of gates in multiqubit systems (only shown in the second and third surveys).
- Measurement probabilities and the resulting state after measurement.
- Separability and entanglement (only shown in the second and third surveys).

All introductory slides of the one-, two-, and three-qubit survey are shown in Supplemental Material [74]. They were designed to introduce the visualization as a complement to the DN and to demonstrate the strategies specific to each representation. The introductory slides are supposed to reduce the influence of previous knowledge on the test results and are designed to thoroughly explain ideas, making them accessible even to participants with only a basic mathematical background. This approach was chosen because the main purpose of the questions was to determine whether students benefit from visualization, with prior knowledge being uncontrolled for due to the lack of validated test instruments for this construct. To avoid hinting at some students toward important problem-solving strategies more than others, participants were not given the opportunity to ask questions for this or any other part of the survey. The participants were instructed to read the introductory slides, answer the survey in good faith, and not use any additional tools. The participants took on average  $55 \pm 33$ ,  $127 \pm 68$ , and  $141 \pm 95$  s on slides of the one-, two-, and three-qubit surveys, and at least 12, 49, and 56 s per slide. In general, the participants engaged with the introductory part of the study. The relevant concepts, aligned with the introductory slides—with which, in principle, all survey questions could be answered correctly without prior quantum knowledge beyond basic mathematics—are described in the following.

#### 1. Visualization of qubit states

The CN uses filled circles with pointers to visualize the amplitudes  $c_i = r_i e^{i\phi_i}$  of a quantum state  $|\psi\rangle = \sum_i c_i |i\rangle$ . We refer to qubit states using the computational basis, which means that  $i$  is written in binary form, with the

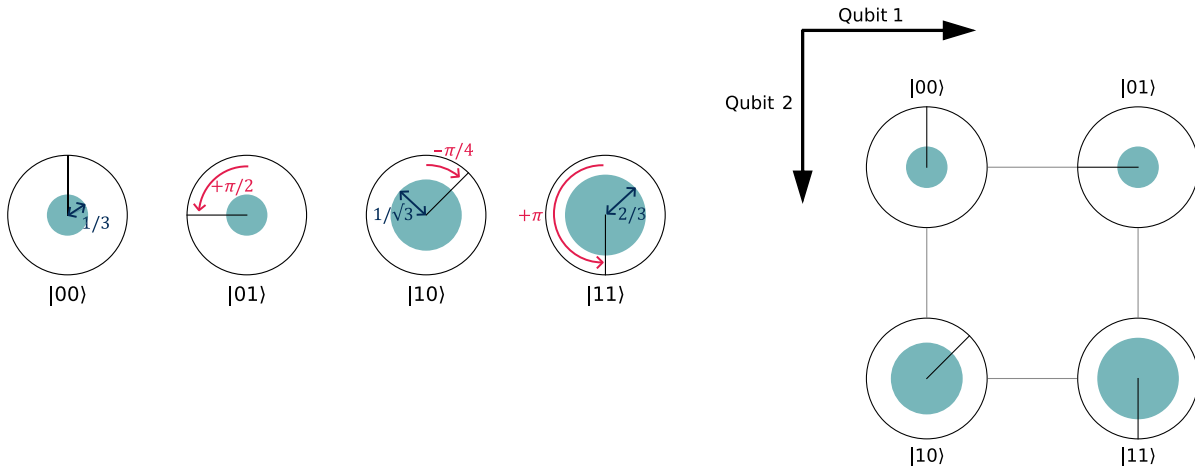


FIG. 2. The state  $|\psi\rangle = \frac{1}{3}|00\rangle + \frac{1}{3}e^{i\pi/2}|01\rangle + \frac{1}{\sqrt{3}}e^{-i\pi/4}|10\rangle - \frac{2}{3}|11\rangle$  in circle notation (left) [53] and in dimensional circle notation (right). The qubits are ordered in from right to left,  $|2, 1\rangle$ .

rightmost digit being the least significant and also the number corresponding to the first qubit. The magnitude  $r_i$  is represented by the radius of the filling circle and the gauge at an angle  $\varphi_i$ , measured counterclockwise from the 12 o'clock axis. In dimensional notations, each qubit is assigned an axis in space [54,56]. Figure 2 shows the same two-qubit state in CN and in DCN. Participants were shown a version of this image that corresponded to their assigned visualization type and the number of qubits in the respective survey.

**2. Action of single-qubit states in single-qubit systems**

The relevant gates to manipulate individual qubits shown in the study are the X, the Z, and the H gates. The X gate flips the qubit and swaps the two coefficients of the  $|0\rangle$  and  $|1\rangle$  states. The Z gate flips the relative phase by  $\pi$ , i.e., changes the sign in front of the coefficient of the  $|1\rangle$ -state. The H (Hadamard) gate creates equal superpositions from the  $|0\rangle$  and  $|1\rangle$  basis states but introduces a  $\pi$  phase flip when acting on the  $|1\rangle$  basis state. The X, Z, and H gates all act inversely in the same way. How these gates function is visualized in Fig. 3.

**3. Action of gates in multiqubit systems**

Single-qubit gates act along the axis of one qubit as if they were operating on the individual pairs of basis states along that axis. The  $X_y$  gate swaps the coefficients of the basis states of a given state, such that the basis states are exchanged with respect to qubit  $y$ . An example slide used in the study, where this concept is visualized in DCN, is shown in Fig. 4.

**4. Measurement probabilities and the resulting state after measurement**

Because a quantum state  $|\psi\rangle = \sum_i c_i|i\rangle$  is normalized, i.e.,  $\sum_i |c_i|^2 = 1$ , the sum of the areas of the inner circles

equals 1. Each area can be interpreted as the measurement probability of the corresponding basis state. After measurement, the state of the system collapses into a renormalized state corresponding to the measurement outcome. An example introductory slide for participants of the two-qubit DN survey is shown in Fig. 5

**5. Separability and entanglement**

In the computational basis, separability of pure states can be seen as symmetry—apart from a complex factor—with respect to each qubit of the system [56]. To determine whether a system is separable, one must examine the ratios between coefficients along the axis of one qubit. If all ratios are equal to a constant  $r$ , the state is separable with respect to that qubit. In two-qubit systems, separability can be observed through symmetry along the axis of one qubit,

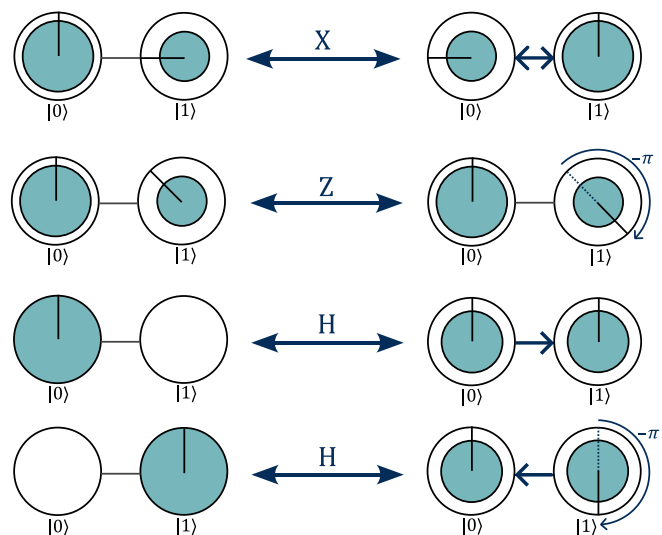


FIG. 3. The X, Z, and H gates in a single-qubit system in circle notation.

**Please read the following info on single-qubit gates in two-qubit systems.**

In two-qubit systems, single-qubit gates change all instances of that qubit as part of the system.

For example, the  $X_y$ -gate swaps coefficients of basis states of a given two-qubit state such that the basis states are swapped in terms of qubit  $y$ :

$$1/2 |00\rangle + 1/\sqrt{2}e^{-i\pi/4} |01\rangle - 1/\sqrt{12} |10\rangle + 1/\sqrt{6}e^{3i\pi/4} |11\rangle \rightarrow_{X_1} 1/\sqrt{2}e^{-i\pi/4} |00\rangle + 1/2 |01\rangle + 1/\sqrt{6}e^{3i\pi/4} |10\rangle - 1/\sqrt{12} |11\rangle$$

$$1/2 |00\rangle + 1/\sqrt{2}e^{-i\pi/4} |01\rangle - 1/\sqrt{12} |10\rangle + 1/\sqrt{6}e^{3i\pi/4} |11\rangle \rightarrow_{X_2} -1/\sqrt{12} |00\rangle + 1/\sqrt{6}e^{3i\pi/4} |01\rangle + 1/2 |10\rangle + 1/\sqrt{2}e^{-i\pi/4} |11\rangle$$

In dimensional circle notation, the gates are applied along the axis of the targeted qubit on every pair of states like a single-qubit gate.

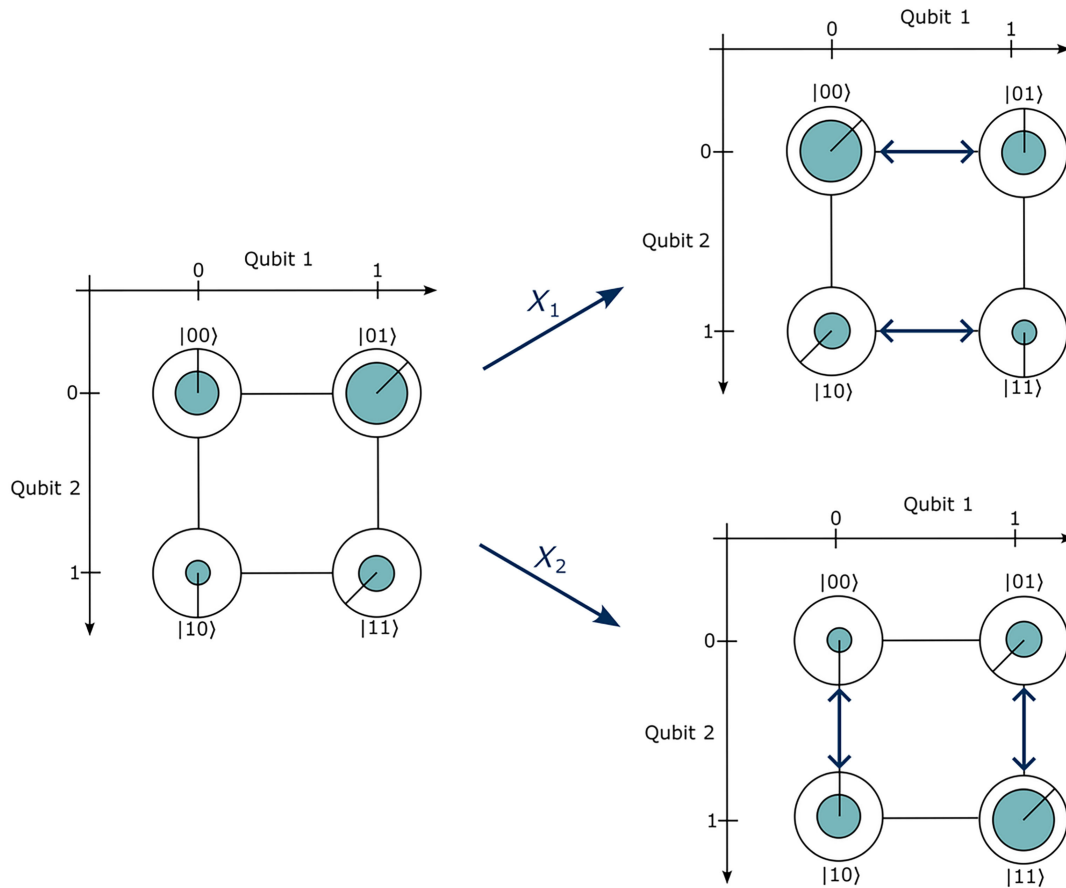


FIG. 4. Introductory example slide for the action of single-qubit gates in two-qubit systems visualized in dimensional circle notation.

apart from the ratio  $r$ . If this symmetry is absent, the system is entangled. In three-qubit systems, one can identify separability by symmetry with respect to the plane perpendicular to a qubit's axis: a qubit is separable from the rest of the system if such a symmetry exists. A system is called partially separable if only a subset of qubits is separable from the rest. This is illustrated in Fig. 6. Alternatively, entanglement can be explained as a dependency of the resulting postmeasurement state on the measurement outcome of the qubit in question.

**E. Representational understanding**

We use the term “representational understanding” instead of “visual understanding” as described by Rau [5], because we are also interested in representational

competence with the traditional mathematical-symbolic notation. In the one-qubit survey, participants were asked one question on the multiplication of complex numbers and one on the Euler formula (for example, translating between  $e^{i\pi/4}$  and  $1/\sqrt{2} + i/\sqrt{2}$ ), and similar questions using only CN. For this, four different questions were designed, two of which each participant received in CN and two of which each participant received in mathematical symbolism, assigned randomly. An example is shown in Fig. 7. These questions are limited in their applicability to basic arithmetic in DN and CN and do not capture spatial reasoning ability that is supposedly necessary for a benefit from DCN in multiqubit systems. There is a lack of alternatives to assess multiqubit representational understanding, which we did not address in this study due to time

**Please read the following info on measurements in two-qubit systems.**

When measuring e.g. qubit 1 in a two-qubit state  $|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle$ , the probability of measuring the value  $x$  is  $|\alpha_{0x}|^2 + |\alpha_{1x}|^2$ .

The resulting state of the system will be  $|\psi\rangle = \frac{\alpha_{0x}|0x\rangle + \alpha_{1x}|1x\rangle}{\sqrt{|\alpha_{0x}|^2 + |\alpha_{1x}|^2}}$ , where only the basis states where qubit 1 has the value  $x$  remain and the state is renormalized.

For example, consider the state  $|\psi\rangle = 1/2|00\rangle + 1/\sqrt{2}e^{-i\pi/4}|01\rangle + 1/\sqrt{6}|10\rangle + 1/\sqrt{12}e^{3i\pi/4}|11\rangle$ .

When measuring qubit 2, the probability  $P_2(0)$  of measuring 0 is  $(1/2)^2 + (1/\sqrt{2})^2 = 3/4$  while the probability  $P_2(1)$  of measuring 1 is  $1/4$ .

The state of the remaining system will be  $|\psi_0\rangle = \frac{1/2|00\rangle + 1/\sqrt{2}e^{-i\pi/4}|01\rangle}{\sqrt{3/4}} = \frac{1}{\sqrt{3}}|00\rangle + \frac{\sqrt{2}}{\sqrt{3}}e^{-i\pi/4}|01\rangle$  or  $|\psi_1\rangle = \frac{1/\sqrt{6}|10\rangle + 1/\sqrt{12}e^{3i\pi/4}|11\rangle}{1/\sqrt{4}} = \frac{\sqrt{2}}{\sqrt{3}}|10\rangle + \frac{1}{\sqrt{3}}e^{3i\pi/4}|11\rangle$ , respectively.

This can be visualized in circle notation:

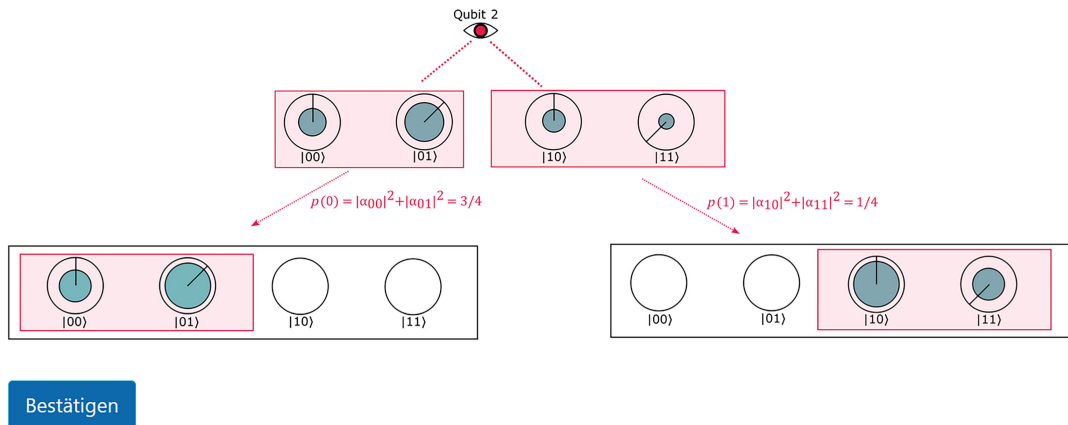


FIG. 5. Introductory example slide for measurements in two-qubit systems visualized in Dirac notation.

constraints on the side of the participants, as we did not want to exceed the 60 min threshold. Because many of the same students also participated in the other surveys and these questions are the most relevant for one-qubit systems, the representational competence questions of the one-qubit survey were not repeated in the other surveys. All representational understanding questions are provided in the Supplementary Material [74]. The questions were designed to include possible misconceptions about CN: for example, students might think that the gauge pointing to the right indicates a phase of 0, and that adding a positive phase means that the gauge turns clockwise.

Participants were asked to rate their confidence on each question using a Likert scale from 0 (random guess) to 4 (very sure). The responses were evaluated using a weighted scoring method as proposed in, e.g., Ref. [76] as having the benefit of reducing statistical error, but may be influenced by participant characteristics, such as self-esteem. The extent to which participants use the visualization is a deciding factor in whether the visualization provides a

benefit. We therefore expect the weighted measure to serve as a reliable moderator variable.

**F. Connectional understanding**

We designed test items to measure participants' connectional understanding between DN and the visualization. After the introductory slides, participants were asked four questions (six questions in the two-qubit survey) to translate between their assigned visualization and the DN. Two (three in the two-qubit survey) of these questions required translating from DN to visualization, and two (three) required translating back. One question addressed the translation of phase, one of magnitude, and in the two-qubit survey, and in the two-qubit survey, we added one question addressing both. An example task from the three-qubit survey is shown in Fig. 8. All connectional understanding questions of the one-, two-, and three-qubit surveys are included in Supplemental Material [74].

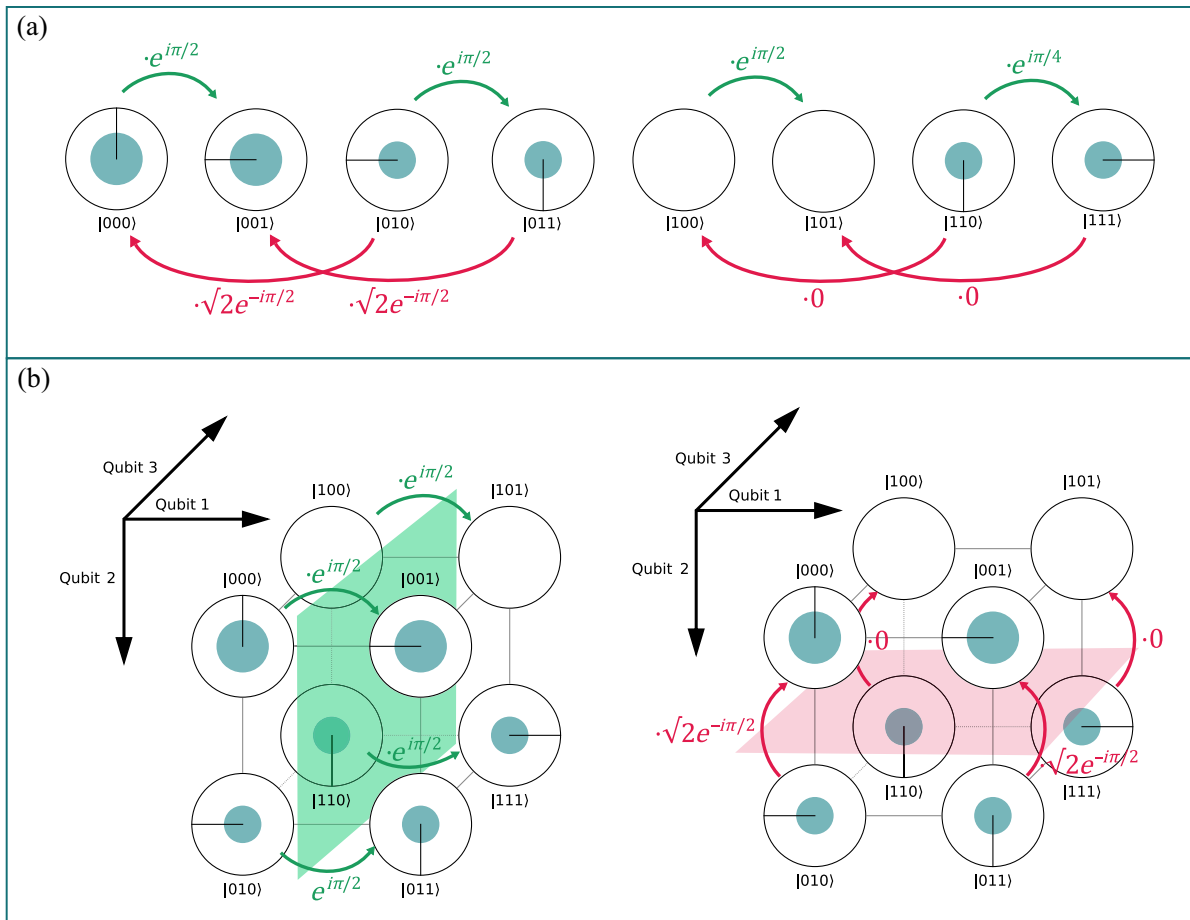


FIG. 6. Partial separability of the state  $|\psi\rangle_{3,2,1} = \frac{1}{2}|000\rangle + \frac{1}{2}e^{i\pi/2}|001\rangle + \frac{1}{2\sqrt{2}}e^{i\pi/2}|010\rangle - \frac{1}{2\sqrt{2}}|011\rangle - \frac{1}{2\sqrt{2}}|110\rangle + \frac{1}{2\sqrt{2}}e^{-i\pi/2}|111\rangle$  [56] in circle notation and dimensional circle notation. (a) Separability in standard circle notation in terms of qubit 1 can be seen by the fact that there exists a ratio  $r_1 = e^{i\pi/4}$  (in green) such that for any coefficients  $c_{xy0}$  and  $c_{xy1}$ ,  $c_{xy1} = r_1 e^{i\pi/4} c_{xy0}$  is the same everywhere. However, as is shown in red, this is not the case for qubit 2. (b) As depicted on the left-hand side, the state is separable in terms of qubit 1, because the state is symmetric apart from a complex factor in terms of qubit 1. This is apparent by the green symmetry plane. For the other two qubits, there exists no such ratio, i.e., no such symmetry planes, as shown here for qubit 2 with the red plane.

The connective understanding questions were designed to test participants’ ability to translate either phases or magnitudes between representations. The distractors were constructed so that no answer stood out regarding any specific feature—or that all did—ensuring that participants could not infer hints toward the correct answer without engaging with the question itself. In addition, the questions were intentionally diverse, illustrating different combinations of amplitudes.

### G. Quantum operations and entanglement

We tested the visualizations in one-, two-, and three-qubit systems. The X and Hadamard gates were tested for all system sizes, and the Z gate was included in the one- and three-qubit systems. In addition, questions about measurement probabilities and the resulting state after measurement were asked in all surveys. The phase gate was tested only in the one-qubit survey, while in the two- and three-qubit

surveys, the CNOT gate, (partial) separability, and (partial) entanglement were also examined. Figure 9 shows an example task on the effect of a Hadamard gate in a two-qubit system. A short explanation of how a given operation acts—or how entanglement or separability is defined—was provided to reduce the influence of prior knowledge, following the same reasoning as for the introductory slides. This also reduces the possibility of a false interpretation of the questions, in line with recommendations from [68].

#### 1. Questions of the one-qubit survey

In the one-qubit survey, we first asked about the equivalence of states that differ only by a global phase to test whether these phase differences are easier to recognize in the CN representation. The distractors were designed to include confusion between shifting only one of the two phases and swapping only magnitudes. The X-gate questions were designed to include the common confusion

What is the result of the following multiplication in circle notation?

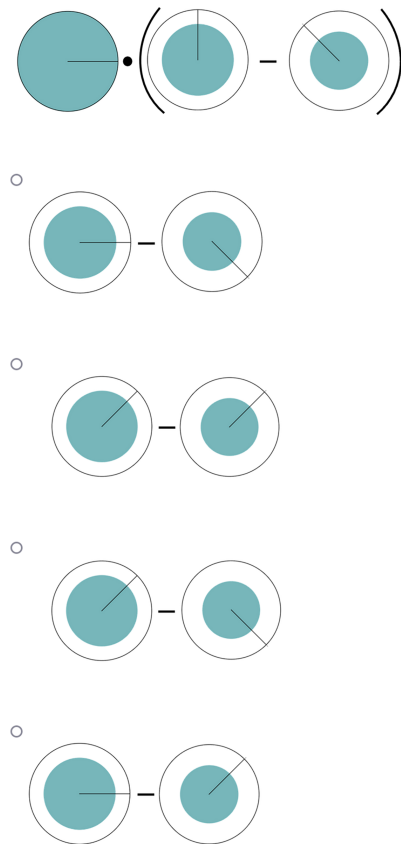


FIG. 7. Example question for representational understanding in circle notation for the calculation  $e^{-i\pi/2} \cdot (\frac{\sqrt{2}}{\sqrt{3}} - \frac{1}{\sqrt{3}} e^{i\pi/4}) = \frac{\sqrt{2}}{\sqrt{3}} e^{-i\pi/2} - \frac{1}{\sqrt{3}} e^{-i\pi/4}$ . The correct answer is the bottom most. The distractors include variations in the phase of the complex numbers depicted by the circles.

about which part of the amplitude to swap; that is, one distractor was chosen such that only the magnitude is swapped, and another such that only the phase is swapped. The final distractor was designed to equalize the magnitudes. For this question, participants needed to recognize the equivalence of two states differing only by a global phase  $(\frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2} e^{i\pi/2}|1\rangle \equiv \frac{1}{2} e^{-i\pi/2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle)$ .

The H-gate questions in the one-qubit survey were designed to be more challenging, as they required participants to apply a combination of superposition creation and destruction. They were meant to involve competence in distributive multiplication of complex numbers; however, discussions with participants revealed that these particular questions were too complex. We therefore decided to ask only about situations where amplitudes were either created or destroyed (but not both) in the two- and three-qubit surveys. The correct answer to the part A question also differed by a global phase  $(\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}} e^{i\pi/2}|1\rangle \equiv \frac{1}{\sqrt{2}} e^{-i\pi/4}|0\rangle + \frac{1}{\sqrt{2}} e^{i\pi/4}|1\rangle)$ . These global phase differences

might represent a disadvantage of CN compared to the Bloch sphere, where global phase information is omitted by construction. The Bloch sphere also provides a unique strategy to solve such questions by imagining a rotation around the diagonal axis between the  $x$  and the  $z$  axis on the sphere.

The Z-gate question included distractors involving confusion of a  $\pi$ -phase shift with a  $\pi/2$ -phase shift, the misconception that Z gates swap magnitudes, and a combination of both. For the phase-gate questions, we added distractors showing the phase shift in the wrong direction or doubled in either direction.

The measurement questions were designed to test whether the depiction of measurement probabilities by circle area aided participants' reasoning. Distractors included interpreting the real part without squaring it, assuming no statement about measurement was possible, and assuming the measurement result was always 1. The second measurement question, on the resulting state after measurement, included distractors with non-normalized or not properly normalized states and unchanged states.

## 2. Quantum operations in two- and three-qubit systems

The two- and three-qubit surveys followed the same structure as the one-qubit survey, except that the phase gate and global-phase equivalence questions were omitted in favor of entanglement and separability. Distractors were included where the gate was applied to the wrong qubit or to both qubits (for CNOT gate, we included a SWAP gate as a distractor). Additional distractors were designed to share similar visual or numerical features, ensuring that no option stood out.

The guess-the-gates questions in the two-qubit survey asked which combination of gates had been applied and included all combinations of  $H_i X_j$  or  $H_i Z_j$  gates, where  $i \neq j$ . The corresponding questions in the three-qubit survey included the  $H_i \text{CNOT}_{ij}$  or  $H_j \text{CNOT}_{ji}$  operations in part A and measurement operations in part B.

The measurement probability questions were designed similarly to those in the one-qubit survey but referred to percentage differences in the two-qubit survey and multiples in the three-qubit survey. We expected that the visualized areas might provide hints for students less comfortable with fractional values. Distractors depicted normalized states where the measured qubit always appeared in state 0 or 1. The measurement state questions were designed to include fewer misconceptions. One question asked participants to identify the correct phases after measurement; the correct answer was  $(\varphi_1, \varphi_1 + \pi/2)$ , while the distractors included  $(\varphi_1 + \pi, \varphi_1 - \pi/2)$ ,  $(\varphi_1 + \pi, \varphi_1 + \pi/2)$ , and  $(\varphi_1, \varphi_1 - \pi/2)$ . In amplitude-based measurement state questions, the amplitudes were swapped around.

What is the following state in circle notation?

$$\frac{1}{\sqrt{6}} |000\rangle + \frac{1}{\sqrt{6}} e^{-i\pi/2} |010\rangle + \frac{1}{\sqrt{6}} e^{3i\pi/4} |011\rangle + \frac{1}{\sqrt{6}} |100\rangle + \frac{1}{\sqrt{6}} e^{i\pi/4} |101\rangle + \frac{1}{\sqrt{6}} e^{i\pi/3} |110\rangle$$

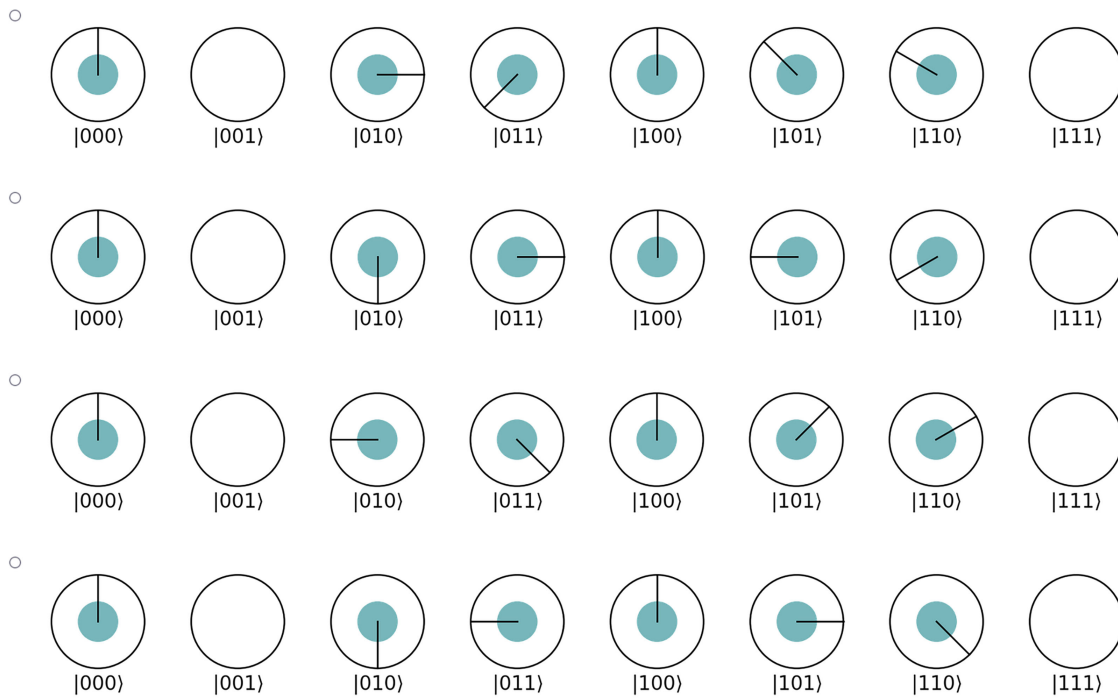


FIG. 8. Example question for translating from Dirac notation to circle notation, where a focus lies in the understanding of the phase in circle notation. The correct answer is the top most. The distractors are chosen to share features with the correct answer, with no answer standing out regarding a single feature.

### 3. Entanglement and separability

For each of the two- and three-qubit surveys, we designed four questions on entanglement. In the two-qubit survey, two questions asked participants to identify separable states and two to identify entangled ones. One question focused on magnitude entanglement (with equal phases), and the other on phase entanglement (with equal amplitudes). In the magnitude entanglement questions, amplitudes were permuted so that only one state was entangled or separable. In the phase separability questions, distractors included different ratio configurations (see Fig. 6):  $r_1 = e^{i\varphi}$ ,  $r_2 = e^{-i\varphi}$ ;  $r_1 = e^{i\varphi}$ ,  $r_2 = e^{2i\varphi}$ ; and  $r_1 = e^{i\varphi}$ ,  $r_2 = e^{-2i\varphi}$ . Because there exist far more entangled states than separable ones [77], the task “find the separable state” allowed for more distractor variations than “find the entangled state.” In the latter case, distractors were restricted to equal ratios, with the correct state chosen as  $r_1 = e^{i\varphi}$ ,  $r_2 = e^{-i\varphi}$ , and the others varying in phase direction and magnitude.

In the three-qubit survey, participants identified fully entangled or fully separable states, with partially entangled states used as distractors. To design the entanglement questions, we started from a fully entangled state (in phase or amplitude) and created distractors symmetric with regard to one qubit, making

that qubit separable from the system. To design questions asking to find fully separable states, we started with a fully symmetric state as the correct answer and generated distractors by applying  $\text{CNOT}_{xy}$  and  $\text{CNOT}_{xz}$  for  $y \neq z$  or by mixing amplitudes from multiple states. The amplitude questions only had four amplitudes that were non-zero, in the phase questions, all amplitudes were non-zero. All magnitudes were kept equal to reduce task complexity. Ideally, the number of questions would be doubled to include all combinations of entanglement/separability and phase/amplitude with and without visualization. In this pilot study, however, each participant was shown all four types only in one condition (with or without visualization), not both.

### H. Cognitive load

Perceived cognitive load was measured after each part of the survey (A and B). The following cognitive load test items were adopted from Ref. [78] to measure intrinsic cognitive load (ICL), germane cognitive load (GCL), and extraneous cognitive load (ECL):

ICL question 1: For the tasks of survey part A/B, many things needed to be kept in mind simultaneously.

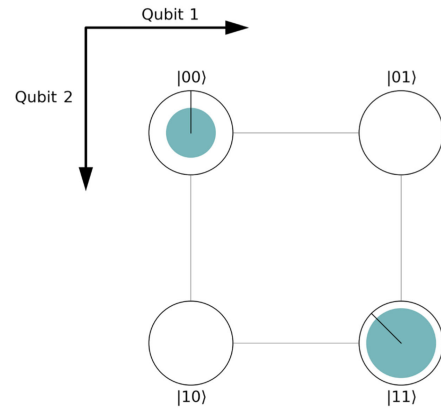
ICL question 2: The tasks of survey part A/B were very complex.

The  $H_x$ -gate (Hadamard-gate) creates an equally weighted superposition from basis states of qubit # $x$ .

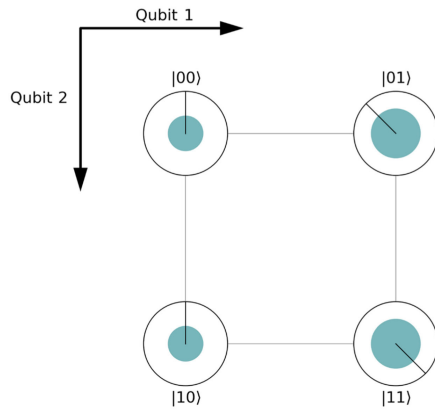
What is the outcome when an  $H_2$ -gate is applied to the following state?

(basis states are written as |qubit 2qubit 1>)

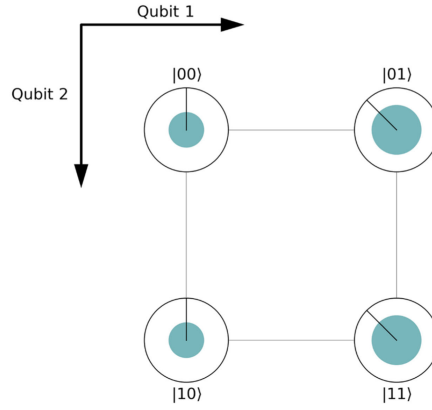
$$\frac{1}{\sqrt{3}}|00\rangle + \frac{\sqrt{2}}{\sqrt{3}}e^{i\pi/4}|11\rangle$$



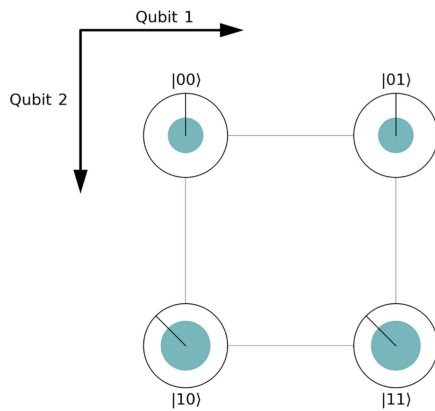
$$\circ \frac{1}{\sqrt{6}}|00\rangle + \frac{1}{\sqrt{3}}e^{i\pi/4}|01\rangle + \frac{1}{\sqrt{6}}|10\rangle - \frac{1}{\sqrt{3}}e^{i\pi/4}|11\rangle$$



$$\circ \frac{1}{\sqrt{6}}|00\rangle + \frac{1}{\sqrt{3}}e^{i\pi/4}|01\rangle + \frac{1}{\sqrt{6}}|10\rangle + \frac{1}{\sqrt{3}}e^{i\pi/4}|11\rangle$$



$$\circ \frac{1}{\sqrt{6}}|00\rangle + \frac{1}{\sqrt{6}}|01\rangle + \frac{1}{\sqrt{3}}e^{i\pi/4}|10\rangle + \frac{1}{\sqrt{3}}e^{i\pi/4}|11\rangle$$



$$\circ \frac{1}{\sqrt{6}}|00\rangle + \frac{1}{\sqrt{6}}|01\rangle + \frac{1}{\sqrt{3}}e^{i\pi/4}|10\rangle - \frac{1}{\sqrt{3}}e^{i\pi/4}|11\rangle$$

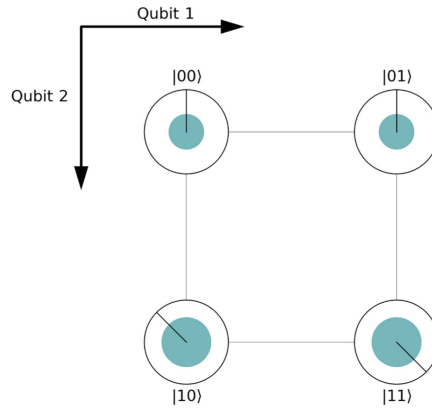


FIG. 9. Example task of predicting the outcome of a Hadamard gate. The correct answer is the one in the top left. The distractors are chosen such that the Hadamard gate is applied to the wrong qubit (bottom right), and the two other distractors sharing features with the first and the correct answer.

GCL question 1: For the tasks of survey part A/B, I had to highly engage myself.

GCL question 2: For the tasks of survey part A/B, I had to think intensively what things meant.

ECL question 1: During the tasks of survey part A/B, it was exhausting to find the important information.

ECL question 2: The design of the tasks of survey part A/B was very inconvenient for learning.

ECL question 3: During the tasks of survey part A/B, it was difficult to recognize and link the crucial information.

Responses were given on a Likert scale from 0 (completely wrong) to 6 (absolutely right).

We chose the naive version of the questionnaire from Ref. [78] to measure GCL, as we concluded that these items fit our use case of problem-solving tasks more appropriately. However, given the limited effects observed for the naive GCL items in Ref. [78], the interpretation of GCL should be viewed critically. The selected items target perceived mental effort—specifically, cognitive engagement and intensity of thinking—which are assumed to correlate with GCL [79].

### I. Data analysis

The small sample size of this pilot study does not allow for reliable statistical analysis. The sample is also heterogeneous, which is typical in QIST education, where participants' academic backgrounds vary widely. To address these challenges, we combine insights from the quantitative results with findings from the qualitative part of the study. This approach allows us to evaluate the suitability of the test items for a heterogeneous target group and to identify potential causes of performance and cognitive load differences between visualization conditions.

From the quantitative data, we infer the appropriateness of question types by examining answer correctness, time taken, and cognitive load. Depending on whether the average correctness was above 84%, between 51% and 84%, or below 51%, each question type in the main study was labeled as easy, moderate, or hard, respectively. We further distinguished between low and high variance, depending on the observed differences between and within groups with and without visualization. Larger differences indicate either stronger dependency on learner prerequisites, the presentation of visualization, or their interaction, and warrant further exploration. In this way, we identify promising directions for more targeted and robust future studies.

## V. INTERVIEW SUMMARY AND DISCUSSION

In order to gain a deeper understanding of the cognitive processes associated with answering the questions of this survey, we interviewed five students who had similar backgrounds and prior experiences to those of the main

study cohort. In doing so, we also assess the quality of the questions, looking especially for false reasoning that leads to correct answers. Here, we thematically summarize the takeaways from the individual participants' thought processes with regard to the following questions: Do the students find answer options that correspond to their misconceptions? How are the visualizations utilized? What problems exist when solving tasks with the visualizations, and do these problems correspond to problems in representational competence?

For an overview of the individual performances of the students in the tests, the scores of the five participants are shown in Table VI, and the perceived cognitive load is shown in Table VII.

### A. Representational competence

Problems with the representational competence questions are summarized in Table VIII. Although three of the five students achieved perfect scores in the symbolic and the visual representational competence, some problems still became apparent. Student no. 1 had no trouble with the symbolic form of complex numbers, but associated the left side of the circle with negative numbers. Student no. 2 had trouble with both the symbolic and the visual representation, twice reversing the direction of the angle of  $e^{-i\pi/4}$  and associating the gauge in the circle with the polar form in the complex plane, describing the line going to the right as the number 1. Student no. 3 interpreted complex numbers as representing a quantum state by separating the real part and the imaginary parts into a superposition. The student also had problems with properties of exponents and held the misconception that positive angles correspond to clockwise direction. Student no. 4 had no problems with the mathematical notation, but in the visual representation associated the right side of the circle with both the negative real part and the negative imaginary components of the complex number. Student no. 5 showed no difficulties.

### B. One-qubit survey

The problems and strategies used in the one-qubit survey are shown in Tables IX (Z and phase gate), X (global phase and X gate), and XI and XII (H gate and measurements) in the Appendix A. Here, we summarize the key takeaways. In the Z- and phase-gate questions, it became apparent that problems with the visualization included confusion about the gauge angle direction. Strategies that circle notation allows include rotating the line and solving exercises by approximation. Such visual strategies are also used in the global phase question, where student no. 1 had no strategy without the visualization. Student no. 2 solved the question with the visualization and used the symbolic notation to check their answer, corresponding to the *constrain-interpretation* function of the Design, Functions, Tasks (DeFT) framework. Most students were confused by the

X-gate question, leading them to disregard parts of the characteristics of the state in order to align the answers with their own ideas; only student no. 5 recognized that a global phase change should be added to arrive at the correct answer. The H-gate question was similarly difficult for the students, with some giving up and others trying to make sense of it by the explanation of the H gate in the question, without success. Only student no. 4 solved this question correctly—not by chance—and only with the visualization, using a mathematical strategy based on matrix multiplication. No visual strategies were used to solve the measurement probability questions. This choice might be due to difficulty in approximating areas (see Ref. [80]).

Some students missed the fact that magnitudes must be squared to obtain measurement probabilities. Student no. 1 approximated the answer by squaring the numbers. Another identified the square root depiction as a hint that measurement probabilities should be squared. The students did not experience problems with the measurement state questions. In summary, visual strategies only seemed relevant in the question involving the manipulation of phases. In these questions, visualization complemented the mathematics by providing alternative strategies, in accordance with the DeFT framework.

### C. Two-qubit survey

In the two-qubit survey, the student were divided into two more groups: Students no. 1 and 2 were assigned to circle notation (CN) and students no. 3, 4, and 5 to dimensional circle notation (DCN). The problems and strategies used in the one-qubit survey are shown in Tables XIII (X and H gate), XIV and XV (guess-the-gates and measurements), and XVI and XVII (separability and entanglement) in the Appendix A. In the X-gate questions, the students assigned to CN chose to swap the amplitudes of basis states in the standard way, while those assigned to DCN adopted visual strategies. Only student no. 3 experienced difficulties in the question with visualization, as they swapped states where qubit 2 is 1 (i.e., a  $CNOT_{21}$ ), instead of applying an  $X_2$  gate correctly. Here, the visual representation might have led to a reduction in the use of cognitive processes that are actually necessary to solve the question, as student no. 3 had answered the X-gate question without visualization correctly and with correct reasoning.

In the H-gate questions, student no. 1 did not know the concept of the H-gate collapsing superpositions, while students no. 3 and no. 4 chose the wrong qubit to perform the H gate on. The other students adopted successful visual strategies, arguing with CN or DCN, but were equally successful without visualization. The guess-the-gates question was more difficult but was still solved correctly by most of the students with and without visualization. In many cases, a random guess or intuition was the reason for this, and the students were, in general, cognitively overloaded by the question or lacked a strategy for solving it.

Only student no. 5 solved both guess-the-gates questions with a correct strategy, while student no. 4 solved the question correctly only without visualization by comparing special features of the coefficients. Student no. 2 solved the questions with visualization using a visual strategy, while relying on intuition without visualization.

In the measurement probability questions, student no. 1 had conceptual difficulties regarding what a measurement of only one qubit of a two-qubit system represents. Student no. 2 calculated the probabilities, ignoring the visualization, while students no. 3 and no. 5 tried solving by approximation but lacked knowledge of a suitable strategy. Student no. 4 had mathematical difficulties. Like in the one-qubit survey, all students ignored the visualization in the measurement probability questions. This, again, hints toward the visualization not providing much benefit to the participants when area comparisons would be required for a visual task-solving strategy.

In the separability and entanglement questions, students no. 1 and no. 3 lacked conceptual understanding. They identified entanglement as a characteristic of *what happens to one qubit also happens to another*, i.e., only definable by action and picked answers at random or by intuition. Student no. 2 only lacked coherent strategies in the questions without visualization. They correctly identified separable states by a notion of “following the trend,” whereas they thought the question of finding an entangled state was about finding a separable state. Students no. 4 and no. 5 adopted successful visual strategies, whereas student no. 4 imagined the circle notation in both questions where it was not presented, in order to solve the question. This strategy requires connectional understanding and is an example of connectional understanding providing benefits even when the representation in question is not present. Student no. 5 remembered the rule  $\alpha_{00}\alpha_{11} = \alpha_{01}\alpha_{10}$  for the coefficients of separable states to solve the question in symbolic notation.

In summary, we see a variety of different strategies in two-qubit systems, both symbolic and visual. These are not only learner dependent but also context dependent. For example, the students did not use visual strategies to solve measurement questions but preferred such strategies for the entanglement and separability questions when they knew how to adopt them. The results point toward the advantages of the visualization by offering strategies not available in the mathematical notation. At the end of the survey, student no. 1 stated: “At least with two qubits it’s way easier to do the operations, it’s just switching things around.” Student no. 5 stated: “I found the circle notation interesting, especially for entanglement and separability. I think it is easy to see whether they are separable or entangled and I did not know that it is possible to see it so easy and I also think it was sometimes hard for me to calculate things in my head.” Thus explaining the decrease in cognitive load with the visualization.

### D. Three-qubit survey

Only two students (no. 2 and no. 4) participated in the three-qubit survey. We note that the voluntary participation of the two students in the three-qubit survey already involves a selection bias, as they were not turned off by the already quite complex one- and two-qubit surveys, such that these students could be seen as higher-performing students in general. The main interview results are shown in Tables XVIII (X and Z gate), XIX (H gate), XX (measurements), XXI (separability and entanglement), and XXII (CNOT gate and guess-the-gates) in the Appendix A. In the X- and Z- gate questions, student no. 4 adopted visual strategies when presented with the visualization. Student no. 2 compared coefficients in the symbolic notation to find the correct answer to the X-gate questions and had trouble with the Z-gate question because they thought the Z gate would switch the phases instead of just changing the relative phase. By contrast, student no. 2 used a visual strategy to solve the H-gate question. Here, student no. 4 observed the coefficients to arrive at an answer, even when presented with the visualization, and had trouble identifying the “partner” states for the H gate to combine into one.

In the measurement probability questions, both students calculated the answer with and without visualization. In the measurement state question, student no. 2 lacked conceptual understanding of renormalization after measurement and indicated that the states merged to one upon being measured. By contrast, student no. 4 arrived at the correct answer by visual comparison of the phases. In the separability and entanglement questions, student no. 2 lacked a strategy on two questions. However, they picked the correct answer by intuition on the separability question without visualization and adopted a visual strategy on the entanglement question with visualization. Student no. 4 adopted purely visual strategies successfully, imagining the DCN in their head to solve also the questions where the visualization was not present. Both students adopted visual strategies also in the CNOT gate question when presented with visualization, swapping states along the axis of the qubit in question but also managed to correctly solve the question without visualization. Both students had problems solving the guess-the-gates questions. Here, student no. 2 again had conceptual difficulties with measurements and was cognitively overloaded on the question without visualization. Student no. 4 applied the gates in the wrong order in the question with visualization.

At the end of the three-qubit survey, we asked for key takeaways. The students responded with positive attitudes toward DCN. Student no. 2 stated: “It was awful trying to do any switches and stuff in the non-DCN, the bra-ket notation. There was so much to manually think about in my head. In DCN I could just see it in front of me, it was comparatively easier so I didn’t have to keep a lot in my head. It was absurdly better in the three-qubit system than

in two. Three qubits needs DCN.” Student no. 4 said: “I found it humane that especially in those questions where the circle notation was not there anymore, that it didn’t get out of hand, that at least the amplitudes were equal and only two phase shifts were there. The questions without circle notation were not too complex. Especially entanglement the circle notation helps a lot. Sometimes I use with Hadamard for example the matrix or the direct calculations, but all in all the circle notation helps in many things significantly. [translated from German]” We see students particularly benefiting from clarification of the role of a single qubit to the whole system, utilizing strategies along the axis of that qubit in the X-, H-, and CNOT gate questions, while also utilizing visual strategies to compare coefficients to identify (partially) separable or (partially) entangled states. A requirement for the use of these strategies, however, is conceptual knowledge.

In summary, we can see that DCN can provide different strategies to solving these questions depending on context, previous knowledge, and representational competence of the students, in accordance with the DeFT framework. The questions were suitable for the students, as the problem-solving strategies used were—although very different—generally reasonable, and problems mostly lay in the execution rather than in conceptual knowledge. We can see that connectional understanding can play an important role in solving these questions, even if DCN is not present, as some students are able to imagine the representation in their heads. Interestingly, this could mean that connectional understanding reduces the difference in performance with and without visualization, suggesting that a measure of conceptual competence with visualization should be adopted in future studies in addition to measuring connectional understanding. Although the three-qubit survey had only two participants, further investigation in this realm may be fruitful.

## VI. QUANTITATIVE RESULTS AND DISCUSSION

In the following, we refer to the group “math-vis” as the group that was assigned questions without visualization in survey part A and questions with visualization in part B. Vice versa, the group “vis-math” was assigned questions with visualization in survey part A and questions without visualization in part B. The quantitative data are shown in Appendix B.

We discuss the results obtained regarding answer correctness and time taken for correct answers for the one-qubit (Sec. VIA), two-qubit (Secs. VIB 1 and VIB 2), and three-qubit (Secs. VIC 1 and VIC 2) surveys with the intention of learning about the suitability of the questions and identifying indications of effects to explore in future studies. As described in [81], the difficulty of an ideal item with three distractors is 0.74. We label an item as difficult at or below an average correctness of 0.5, moderate between 0.51 and 0.84, and easy at or above 0.85 [82]. We also

TABLE III. Average scores (the maximum possible is 1 for 100% correct) and time taken (in seconds) for correctly solved questions, average connectional understanding, and the difference of the performance metrics with visualization minus without in the different surveys. Means and standard deviations are calculated taking the averages over the participants individual results.

Number of qubits	One		Two		Three	
	CN	CN	DCN	CN	DCN	
Average connectional understanding	$0.81 \pm 0.23$	$0.78 \pm 0.21$	$0.7 \pm 0.23$	$0.85 \pm 0.17$	$0.89 \pm 0.12$	
Score without visualization	$0.67 \pm 0.24$	$0.59 \pm 0.25$	$0.58 \pm 0.29$	$0.62 \pm 0.26$	$0.73 \pm 0.28$	
Score with visualization	$0.7 \pm 0.24$	$0.61 \pm 0.23$	$0.54 \pm 0.31$	$0.68 \pm 0.25$	$0.67 \pm 0.24$	
Score difference	$0.03 \pm 0.25$	$0.01 \pm 0.14$	$-0.04 \pm 0.3$	$0.06 \pm 0.1$	$-0.06 \pm 0.13$	
Time (s) without visualization	$99 \pm 62$	$89 \pm 49$	$121 \pm 52$	$105 \pm 32$	$140 \pm 88$	
Time (s) with visualization	$87 \pm 49$	$114 \pm 73$	$117 \pm 75$	$110 \pm 53$	$121 \pm 53$	
Time (s) difference	$-12 \pm 59$	$25 \pm 56$	$-4 \pm 69$	$6 \pm 38$	$-18 \pm 51$	

TABLE IV. Average intrinsic cognitive load (ICL), germane cognitive load (GCL), and extraneous cognitive load (ECL) with and without visualization, and the corresponding differences of with visualization minus without. The maximum possible cognitive load score is 6. Means and standard deviations are calculated taking the averages over the participants individual results.

Number of qubits	One		Two		Three	
	CN	CN	DCN	CN	DCN	
ICL without visualization	$3.73 \pm 1.22$	$3.92 \pm 1.04$	$4.69 \pm 1.55$	$4.62 \pm 0.99$	$4.57 \pm 1.29$	
ICL with visualization	$3.3 \pm 1.27$	$4.0 \pm 1.13$	$4.0 \pm 1.46$	$4.4 \pm 1.11$	$3.5 \pm 1.8$	
ICL difference	$-0.42 \pm 1.32$	$0.18 \pm 1.11$	$-0.55 \pm 1.19$	$-0.25 \pm 1.2$	$-0.83 \pm 1.77$	
GCL without visualization	$4.2 \pm 1.39$	$4.73 \pm 0.86$	$4.84 \pm 1.15$	$4.89 \pm 1.37$	$4.86 \pm 1.36$	
GCL with visualization	$3.91 \pm 1.21$	$4.27 \pm 0.86$	$4.73 \pm 1.0$	$4.4 \pm 1.36$	$3.57 \pm 1.59$	
GCL difference	$-0.25 \pm 1.41$	$-0.5 \pm 0.92$	$-0.11 \pm 0.92$	$-0.44 \pm 1.26$	$-1.29 \pm 1.83$	
ECL without visualization	$2.67 \pm 1.35$	$2.48 \pm 0.82$	$3.98 \pm 1.34$	$3.05 \pm 1.17$	$3.21 \pm 1.46$	
ECL with visualization	$2.4 \pm 1.39$	$2.59 \pm 0.82$	$3.23 \pm 1.62$	$2.75 \pm 1.33$	$1.86 \pm 1.38$	
ECL difference	$-0.3 \pm 1.72$	$0.11 \pm 1.26$	$-0.75 \pm 1.66$	$-0.3 \pm 1.33$	$-1.36 \pm 2.03$	

examine the variance of the results for each type of question with and without visualization and the differences between the two groups to separate the questions into high or low variance. We consider questions with moderate difficulty and high variance to be of the highest value for further exploration. We provide such an overview of the discussion in Sec. VII. Table III summarizes the average connectional understanding, average scores, and average times for correctly solved questions. In addition, it shows the differences in these performance metrics (with visualization minus without visualization). Average intrinsic cognitive load (ICL), germane cognitive load (GCL) and extraneous cognitive load (ECL), and the corresponding differences (with visualization minus without visualization) are shown in Table IV. A more detailed analysis can be found in Appendix B.

### A. One-qubit systems

In the first survey, participants were asked to find

- a state that is equal to another state apart from a global phase,

- the resulting state after application of an X gate (only in part A) or a Z gate (only in part B), phase gate, and a Hadamard (H) gate,
- the probability of finding a particular result upon measurement of a given state, and
- the resulting state after measurement.

The average scores were  $0.67 \pm 0.24$  without visualization and  $0.70 \pm 0.24$  with visualization, at average times of  $104 \pm 74$  and  $88 \pm 56$  s, respectively. In general, the questions seemed suitable for the participants, as the mean scores were mostly above 50%, except for the X-gate and measurement probability questions. The necessity to think about global phase equivalence could be a reason for the low performance in the X-gate question, while many participants might have had no prior knowledge of or forgotten to square measurement probabilities. The average scores for single-qubit questions are shown in Table XXIII, and the average times for correctly solved questions are shown in Table XXIV. Figure 20 shows the combined average score per question, and Fig. 21 shows the combined average time taken (for all answers, not only the correct ones).

### 1. Representational understanding

Participants were asked to calculate Euler formula identities and similar questions to translate between circle notation (CN) and a complex number. In addition, they were asked to multiply the numbers written in DN and CN, as shown in Fig. 7. Then they were asked to rate their confidence for the question from 0 (random guess), 1 (very unsure), to 4 (very sure). The correct answers were weighted (multiplied by  $x/4$ , where  $x$  is the confidence). Detailed results are shown in Sec. VI B 1 a.

None of the participants scored higher with CN than with Dirac notation (DN) (weighted or unweighted), suggesting that the representational understanding of most participants of the visualization was lower than that of DN. Presenting the visualization to students with high representational understanding with CN but low representational understanding with the DN would probably show greater benefits, but these students might be difficult to find.

All in all, we see that possible benefits of visualization are learner and context dependent, and, if present, are probably not strongly dependent on the presence of the visualization, that is, we expect small effects if any. It is worth investigating further under which circumstances benefits occur. In one-qubit systems, the CN is merely a visual representation of complex numbers, providing an aid, in theory, complementary to the DN. In two- and three-qubit systems, dimensional circle notation (DCN) is an extension of CN that could provide additional benefits due to the introduction of dimensionality.

### 2. Connectional understanding and cognitive load

The average connectional understanding was  $0.81 \pm 0.23$ . Thirteen students achieved perfect scores, and ten students achieved a score between 0.5 and 0.75. The students who did not score perfectly also scored lower in the rest of the survey ( $0.57 \pm 0.10$  to  $0.74 \pm 0.20$  without visualization and  $0.57 \pm 0.25$  to  $0.78 \pm 0.21$  with visualization) and took less time on the correctly solved questions ( $82 \pm 74$  to  $114 \pm 71$  s without visualization and  $86 \pm 57$  to  $89 \pm 55$  s with visualization) than the students who solved the connectional understanding questions perfectly. The average perceived extraneous cognitive load (ECL) was  $2.4 \pm 1.39$  with visualization and  $2.67 \pm 1.35$  without, average intrinsic cognitive load (ICL) was  $3.3 \pm 1.27$  with visualization and  $3.73 \pm 1.22$  without, and germane cognitive load (GCL) was, on average,  $3.91 \pm 1.21$  with visualization and  $4.2 \pm 1.39$  without. In the difference of perceived ECL of survey parts supported by visualization minus parts not supported by visualization, the numbers vary more (as can be seen in the high variance of 1.72 compared to 1.32 and 1.41 in ICL and GCL, respectively; see Table III). This high variance could indicate a strong dependence of ECL reduction due to visualization on learner characteristics.

### B. Two-qubit systems

In both parts of the second survey, participants were split into two groups, one presented with CN and one presented with DCN, and were asked to find

- the resulting state after application of an X gate and a Hadamard (H) gate,
- the probability of finding a particular result upon measurement of a given state,
- a separable state and an entangled state (between nonseparable, or nonentangled states, respectively), and
- a combination of two gates that were applied to transform one given state into another given state (“guess-the-gates”) questions.

The average scores for two-qubit questions are shown in Table XXV and the average times for correctly solved questions are shown in Table XXVI. In the latter, the combined scores of part A and part B show only cases where participants managed to solve both questions, i.e., both questions on the same concept with and without visualization, correctly. Figure 20 shows the combined average score per question, and Fig. 21 shows the combined average time taken (for all answers, not only the correct ones).

#### 1. Two-qubit systems with circle notation

In the two-qubit survey with CN, the average scores were  $0.59 \pm 0.25$  without visualization and  $0.61 \pm 0.23$  with visualization, at average times of  $82 \pm 47$  and  $113 \pm 76$  s, respectively. In both survey parts, we see group math-vis scoring higher than group vis-math on all questions except the separability and entanglement questions, where group vis-math scored higher. This could be explained by learner characteristics, such as the ability to identify entanglement by visual means or using mathematical strategies. The H-gate question shows a large difference in score between the two groups, where group math-vis scored 0.86 without visualization and group vis-math scored 0.17. This could hint toward the Hadamard gate question performance being especially sensitive to learner prerequisites.

When presented with the visualization, participants took longer when solving questions correctly, at  $113 \pm 76$  s versus  $82 \pm 47$  s. This could be the result of alternative strategies to solve the question that the visualization provides, which brings the question’s difficulty closer to the level of the participant’s ability, according to the distance-difficulty hypothesis [69,70].

*Connectional understanding and cognitive load.* Similarly to the one-qubit survey, participants scored  $0.85 \pm 0.14$  in translating between CN and DN. Again, we see that participants who scored perfectly when solving these questions also scored higher in the rest of the survey ( $0.88 \pm 0.14$  without and  $0.83 \pm 0.17$  with visualization) than participants without a perfect score ( $0.45 \pm 0.15$  without visualization and  $0.50 \pm 0.16$  with visualization).

The group of participants with perfect connectional understanding scores also took more “reasonable” amounts of time to correctly solve questions, i.e., the variance in time taken was lower ( $100 \pm 39$  s without visualization and  $120 \pm 40$  with visualization) than in the group without a perfect connectional understanding score ( $68 \pm 47$  s without visualization and  $107 \pm 91$  with visualization). Presumably, students who did not know the answer and chose to take a guess took a shorter amount of time, while students who tried hard to find an answer in tasks where this would not be necessary for students with greater prior knowledge will take a long time. Therefore, a higher variance in the time taken for task solving could be expected for lower-performing students [69,70].

The average ICL was  $3.92 \pm 1.04$  without and  $4.0 \pm 1.13$  with visualization, GCL ( $4.73 \pm 0.86$  without and  $4.27 \pm 0.86$  with visualization), and ECL was  $2.48 \pm 0.82$  without and  $2.59 \pm 0.82$  with visualization. It seems that, on average, the questions succeeded in offering a high degree of intrinsic difficulty, as indicated by the comparatively high ICL.

## 2. Two-qubit systems with dimensional circle notation

In the two-qubit survey with DCN, the average scores were  $0.58 \pm 0.29$  without visualization and  $0.54 \pm 0.31$  with visualization, at average times of  $121 \pm 58$  and  $139 \pm 99$  s, respectively. In all question types except the X gate, one group performed better than the other group in both part A and part B (regardless of whether visualization is presented). This consistency can possibly be explained by learner prerequisites predicting performance better than the visualization condition. In addition, this could reflect context dependency of student competence, i.e., competence in one quantum process does not necessarily transfer to others. The differences between the scores of the groups are particularly apparent in the H gate and entanglement questions in favor of group vis-math (averaging over the question types: 0.56 vs 0.14, and 0.81 vs 0.43, respectively) and the guess-the-gates questions in favor of group math-vis (0.44 vs 0.1).

The H-gate and guess-the-gates questions seemed quite difficult, as the average scores in these questions were below 0.4. Feedback from some students after the test suggested that the guess-the-gates questions were particularly challenging and mentally demanding.

Considering the time spent on correctly solved questions, we see outliers of above 6 min in the group vis-math on the Z-gate A, H-gate A, and H-gate questions of part B. Participants took, on average, a similar amount of time when presented with visualization or without ( $121 \pm 52$  s without visualization vs  $117 \pm 75$  s without).

*Connectional understanding and cognitive load.* The average connectional understanding of participants in the two-qubit DCN survey was  $0.7 \pm 0.23$ . No one in the DCN

group correctly answered all the connectional understanding questions. The average scores in this survey were also lower than in all other surveys, at  $0.58 \pm 0.29$  without visualization and  $0.54 \pm 0.31$  with visualization.

For participants of the DCN group, ICL and ECL were higher than for participants of the CN group when solving questions without visualization, at  $4.7 \pm 1.6$  and  $4.0 \pm 1.3$ , respectively, with ECL higher than in all other surveys. We see indications of reductions in ICL and ECL when participants are presented with DCN of  $-0.6 \pm 1.2$  and  $-0.8 \pm 1.7$ . We do not see a similar reduction in GCL as we do in the CN group. The cognitive load results indicate potential reductions in ICL and ECL when students are presented with DCN, specifically, as opposed to CN. If students adopt new strategies to solve questions using DCN rather than Dirac Notation, a reduction in ICL could be theoretically explained if these strategies make the tasks easier to solve.

## C. Three qubits

In the third survey, participants were asked to find:

- the resulting state after application of an X gate, a Z gate, and a Hadamard (H) gate,
- the probability of finding a particular result upon measurement of a given state,
- the resulting state after a measurement,
- the resulting state after application of a CNOT gate,
- a fully separable state and a fully entangled state (between partially separable and entangled states), and
- a combination of two gates that were applied to transform one given state into another given state (“guess-the-gates” questions).

The average scores for the three-qubit questions are shown in Table XXVII and the average times for correctly solved questions are shown in Table XXVIII. In the latter, the combined scores of part A and part B show only cases where participants managed to solve both questions, i.e., two questions of the same type, correctly. Figure 22 shows the combined average score per question, and Fig. 23 shows the combined average time taken (for all answers, not just the correct ones). Cases where a participant took more than 10 min for one question were excluded.

### 1. Three-qubit systems with circle notation

In general, the participants scored marginally higher when presented with CN ( $0.68 \pm 0.25$ ) than when presented with the questions without visualization ( $0.62 \pm 0.28$ ). The overall score reflects that the questions were suitable for the participants in terms of difficulty. The results of the H-gate and guess-the-gates questions could indicate a benefit of presenting the visualization in these contexts, as both groups scored higher when the visualization was presented (0.9 vs 0.6 for the H-gate question and 0.6 vs 0.1 for the guess-the-gates question). In the entanglement question, group math-vis scored higher

(0.6 without visualization and 0.4 with visualization) than group vis-math (0.2 with and without visualization) in both parts of the survey. These results could indicate differences in context-dependent competence of the two groups, similar to what we observed in the two-qubit survey, in the identification of entangled states. In some cases, students seemed to benefit from the visualization, whereas, on average, there is no difference.

The results warrant further investigation in the regime of the more complex H-gate, guess-the-gates, separability, and entanglement questions to find the reasons for the increases in performance with visualization for some participant groups and large deviations. However, due to the special complexity of the guess-the-gates questions, it should probably not be included in future studies involving students not as well versed in QIST. To benefit from visualization in the context of separability and entanglement, students may need a thorough introduction to the strategy to identify separability and entanglement (see Fig. 6). We received feedback from some participants that they did not comprehend the process of identifying entangled states before or during the study. Perhaps the topic is too advanced as only a subtopic of a study like this and should be introduced more thoroughly in a study focused solely on this context.

*Connectional understanding and cognitive load.* In the three-qubit survey, participants presented with CN scored on average  $0.85 \pm 0.17$  on the connectional understanding questions, similar to the results of the one- and two-qubit CN surveys. Participants who scored perfectly in connectional understanding also scored comparatively high in the rest of the survey at  $0.80 \pm 0.16$  without visualization and  $0.84 \pm 0.17$  with visualization, compared to  $0.44 \pm 0.21$  and  $0.51 \pm 0.19$  for the other participants. However, we do not see this connectional understanding translate into an improvement in score when visualization is presented.

When participants were presented with CN, we observed indications of a reduction in ICL, GCL, and ECL of  $-0.25 \pm 1.28$ ,  $-0.44 \pm 1.33$ , and  $-0.3 \pm 1.4$ , respectively. These results justify further investigation into the circumstances under which a reduction in cognitive load occurs through presentation of CN.

## 2. Three-qubit systems with dimensional circle notation

Group math-vis consists of two students, and group vis-math consists of three participants. Hence, the resolution on the average score value per question is coarse. Due to the large impact a single person's performance can have on the results, it is difficult to find reliable indications of effects. Future studies could focus more on the impact of DCN in the domain of three qubits in particular. If there is any benefit of visualizing quantum states representing qubits as axes in space, one would expect the benefit to be the largest here, because the difference between the dimensional approach and standard CN increases with system size.

Similarly to the results for answer correctness, it would be unreasonable to make statements about the time taken for correct answers due to the small number of participants.

*Connectional understanding and cognitive load.* The impact of the results of one participant seems too large to make any firm assumptions about the connectional understanding and cognitive load in the three-qubit DCN context.

## VII. SUMMARY OF DISCUSSION AND LIMITATIONS

In the qualitative part of the study, we observed that students adopt a multitude of different strategies when solving questions with visualization. These strategies are learner and context dependent, as predicted by the Design, Functions, Tasks (DeFT) framework. They fall within the *constrain interpretation* and *complementing* domains, using symbolic notation to perform more accurate calculations or to check answers, and utilizing visualization especially to determine actions involving phase changes, to see where gates act along the qubit axis, or to identify (partially) separable or (partially) entangled states. This lays the foundation for expecting performance and cognitive load differences when students solve questions with visualization as compared to when they do not. However, we also observe hints of context dependency: for example, visualization seems to be used less when the question requires comparison of areas, as in the measurement probability questions. In general, we find the suitability of the questions for a heterogeneous target group, as these questions were neither too difficult nor too easy for the five students. The group was recruited from our quantum computing lecture and was relatively new to the field, and as such, can be seen as representative, although the size of the group means that the number of different strategies that could be identified is limited.

Considering the suitability of the questions in terms of performance alone and the variance between the two groups with or without visualization, we summarize the results of this study in Table V, where we assigned questions a label of easy (average score above 0.84), moderate (between 0.51 and 0.84) or hard (below 0.51), and labeled half the questions with differences between groups and/or whether visualization was presented above the median as high variance, and below the median as low variance. We consider questions of moderate difficulty and questions with higher variance to be of value for further investigation, as it is unclear whether this variance is due to learner characteristics or the presentation of the visualization. Overall, questions of moderate difficulty and high variance in at least two different surveys include the H- and the Z-gate (three-qubit systems), the CNOT gate [two-qubit circle notation (CN) and three-qubit dimensional circle notation (DCN)], and entanglement (two qubits) questions. However, in both three-qubit surveys,

TABLE V. Overview of question types across survey contexts [one-, two-, and three-qubits in circle notation (CN) and dimensional circle notation (DCN)]. Each cell lists a difficulty level (hard (h) for scores below 0.51, moderate (m) for scores between 0.51 and 0.84, and easy (e) for scores above 0.84 [82]) and a variance indicator (low variance (lv) for a standard deviation of the four group averages (with or w/o vis. and group math-vis or group vis-math) below the median standard deviation of 0.13, and high variance (hv) for standard deviation at or above 0.13). Questions with the m/hv combination are highlighted or in bold, as these questions might be of most interest in future studies.

Question type	One-qubit CN		Two-qubit CN		Three-qubit DCN		Three-qubit CN		Three-qubit DCN	
	Difficulty	Variance	Difficulty	Variance	Difficulty	Variance	Difficulty	Variance	Difficulty	Variance
Global phase equivalence	<b>m</b>	<b>hv</b>	...	...	...	...	...	...	...	...
X gate	h	...	<b>m</b>	<b>hv</b>	e	lv	e	lv	e	hv
Z gate	m	...	...	...	...	...	<b>m</b>	<b>hv</b>	<b>m</b>	<b>hv</b>
H gate	m	lv	e	hv	h	hv	<b>m</b>	<b>hv</b>	<b>m</b>	<b>hv</b>
Phase gate	m	lv	...	...	...	...	...	...	...	...
Measurement probability	h	hv	h	lv	h	hv	<b>m</b>	<b>hv</b>	e	hv
Measurement state	m	hv	...	...	...	...	m	lv	<b>m</b>	<b>hv</b>
CNOT gate	...	...	...	...	...	...	<b>m</b>	<b>hv</b>	<b>m</b>	<b>hv</b>
Guess-the-gates	...	...	h	lv	h	hv	h	hv	h	hv
Separability	...	...	<b>m</b>	<b>hv</b>	m	lv	h	lv	<b>m</b>	<b>hv</b>
Entanglement	...	...	<b>m</b>	<b>hv</b>	<b>m</b>	<b>hv</b>	h	hv	<b>m</b>	<b>hv</b>

one participant group answered the Z-gate question perfectly, indicating possible ceiling effects in that question type in particular.

In general, we do not see indications of a large change in performance when participants are presented with a visualization (see Table III), although there are indications of context-dependent performance increases with visualization that require further investigation. These increases are apparent in the global phase equivalency question of the one-qubit survey, the Z-gate, the measurement probability, and the separability question of the two-qubit CN survey, the H-gate question of the two-qubit DCN survey, and the H-gate and guess-the-gates questions of the three-qubit CN survey. However, context-dependent learner characteristics seem to have a greater impact on performance than the presence or absence of a visualization. This could be due to students with better background knowledge already knowing how to solve the questions regardless of whether visualization is present or not, while students with less background knowledge might not know how to utilize the visualization. Some groups seem to perform better than other groups, specifically in certain types of questions, as we see in the two-qubit CN survey in the X-gate, the H-gate, separability and entanglement questions, all questions in the two-qubit DCN survey, and in the Z-gate, measurement probability, entanglement, CNOT gate, and guess-the-gates questions in the three-qubit CN survey, further highlighting the dependency of possible effects on the interaction between context and task solvers.

In order to find indicators of complexity and suitability of questions, we considered the time taken for all answers, not just correct ones (see Figs. 16, 21, and 23 for the one-, two-, and three-qubit surveys, respectively). During the one-qubit survey, the Global Phase and the H gate questions took the participants the longest time. We would omit

the H-gate and X-gate questions, as they were associated with unnecessary confusion due to the incorporation of the global phase equivalency. During the two-qubit survey, the guess-the-gates and the measurement probability questions took participants in the CN survey the longest, and the guess-the-gates and H-gate questions for participants of the DCN survey (see Fig. 21). In the three-qubit survey, participants presented with both CN and DCN took the longest in the guess-the-gates and H gate questions (see Fig. 23), in the former of which some participant groups also performed quite poorly. Long times could in some cases be due to greater complexity and in some cases to the distance-difficulty hypothesis [69,70], which, in contrast, would suggest that these questions were quite suitable for the participants' level of ability, making performance on these questions an important indicator. Based on these results, we would omit the guess-the-gates questions in two- and three-qubit systems in future studies due to their high difficulty, as the interview results also suggest.

We also measured the cognitive load imposed on the students during the surveys. The generally high intrinsic cognitive load (ICL) and average scores between 0.5 and 0.7 indicate that the difficulty of the problems was adequate overall. We see reductions in cognitive load when students are presented with visualization in both CN and DCN surveys. One possibility is that the amount of space DCN takes up compared to CN leads to increased extraneous cognitive load (ECL), due to participants not being able to see all answers at once (which means page scrolling becomes necessary), resulting in a possible split-attention effect [83]. On the other hand, the lower connective understanding of participants in the two-qubit DCN survey could also be a reason for the higher ECL. To summarize the results in regard to RQ1, on average, the test items

were suitable, even for a heterogeneous target group, with a few exceptions.

Most of the participants were recruited from our quantum computing lecture. While it is an optional master's lecture in information science, students from other fields also select it as an elective subject. We expect and have experienced a higher reluctance towards mathematical notation among students from other fields than physics or mathematics, especially if they were not introduced to DN before. When answering the question of whether they had been introduced to DN before, we received feedback that some students thought that this referred to prelecture knowledge. We have also received positive feedback regarding the visualizations from students newer to the field, while more advanced physics students often do not see a benefit in the visualizations. A possible expertise-reversal effect remains to be investigated. The A-B survey design should mitigate previous knowledge as a relevant factor if the questions are neither too easy nor too difficult. For future studies with such heterogeneous groups, eye tracking or a think-aloud study design might allow us to gain a better understanding of the underlying thought processes. In studies with more participants, the analysis can be performed by separating the groups according to experience with DN.

In this study, we have seen how performance and cognitive load can be context and learner dependent. The results give an overview of the cognitive load with and without visualization, but context-specific assumptions cannot be made with these results. It can also be considered that measurements of cognitive load after multiple questions could be biased toward the most recently answered questions, which would be the measurement questions in the one-qubit survey and the guess-the-gates questions in the two- and three-qubit surveys. Performance on these questions is probably very dependent on previous knowledge. This, combined with the high expected heterogeneity, motivates separate analyses of the results for different types of questions in future studies.

The test instruments presented in this study are purely designed to measure the benefits of visualization in learning basic quantum operations and entanglement. They cover a broad range of quantum gates in one-, two-, and three-qubit systems, but do not address more complicated processes such as whole quantum algorithms or larger qubit systems, and they leave out real-world applications like different qubit implementations, discussions about consequences of quantum information theory like the no-cloning theorem, and decoherence, and omit discussion about various applications.

The measures of representational competence presented in this study are limited. The questions measure only a part of representational competence—representational and connectional understanding. Other factors for representational competence would include representational and connectional fluency and metarepresentational competencies,

considering the symbolic Dirac notation (DN) as a representation and therefore extending the framework in [5]. In the future, additional ideas for questions that provide a more general metric should be generated. For example, metarepresentational competence—choosing between DN, CN, or DCN to most efficiently solve different questions—could be investigated using eye tracking. Representational understanding is only measured in the first survey and includes only understanding of the phase in CN. In future studies, this should be extended to finding a measure of representational understanding of amplitudes as well, e.g., by asking students to compare numbers and sizes of circles. This type of understanding was only included in the connectional understanding questions.

Due to small sample sizes (especially in the three-qubit DCN survey) and large fluctuations (possibly due to large deviations of learner prerequisites), we can only identify indications of positive effects of presenting a visualization regarding performance and cognitive load. The circumstances and contexts in which these possible effects occur should be investigated more thoroughly. This study covered a broad range of different contexts, and, therefore, compromises had to be made with regard to the number of questions for each question type. To answer RQ3, reevaluation of the test instruments can be based on fine-tuning to make the questions in parts A and B of the surveys more similar. This can be done with respect to questions asking about, e.g., the state after measurement or entanglement, ensuring that all participants are presented with amplitude-type and phase-type questions in both parts of the survey. This would either mean an increase in the number of questions or a narrowing of question types. Due to the already large number and high complexity of the questions, we recommend the second option.

## VIII. CONCLUSION AND OUTLOOK

In general, the interviews, as well as the performance and cognitive load results, indicate that most of the questions were suitable for the participants. However, there were some notable exceptions. The X-gate and measurement state questions were possibly too easy, except when a global phase was also involved, as in the X-gate question of the one-qubit survey. The H-gate, measurement probability, guess-the-gates, separability, and entanglement questions were too difficult for some participant groups. Considering the overall time taken, the number of questions appears feasible for the participants. The A-B-crossover design allowed for within-subject as well as between-subject comparisons, while mitigating multiple testing effects due to randomization of the order (A-B or B-A) and we therefore see it as suitable to answer the overarching questions (see Sec. III).

The results of this study open up a wide range of questions to investigate in future studies. Some students performed better and experienced reduced cognitive load when presented with visualization, while others did not.

Therefore, it is important to investigate further under which circumstances exactly students benefit from visualizations of qubit systems, and which strategies are actually useful in this regard. In particular, contexts within the three-qubit dimensional circle notation (DCN) setting should be investigated more thoroughly. Here, DCN differs the most from the other two representations used, since the number of steps necessary to transform circle notation (CN) to DCN increases with the number of qubits. If, in fact, educators wish to teach students about systems of multiple qubits, future studies should also be restricted to multiqubit systems. If the goal is to teach only one-qubit systems, visualizations like the Bloch sphere or vector notation could also be included in the studies. The Bloch sphere could alleviate issues with the one-qubit X- and H-gate questions that the students encountered during this study.

The variance in results implies a substantial learner and context dependency. Therefore, future studies should control for these characteristics in a more detailed way to draw conclusions about the potential benefits of visualizations, as also proposed in [10]. This could also alleviate issues with question difficulty—for example, with the Hadamard and Entanglement/separability questions—where proper strategies to solve these questions could be emphasized more thoroughly in the introduction design. In order to support the understanding of cognitive processes involved in problem solving with visualization, the use of eye tracking techniques could be a valuable tool [84]. With eye tracking, it would be possible to observe which strategies the participants adopted to solve the questions and which strategies were used by more successful participants. This could result in further design guidelines for educators. It could also be investigated whether there are other benefits of visualization, such as a shorter fixation duration on incorrect answers when a visualization is presented [57].

Representational competence as a learner characteristic is theoretically the deciding factor for whether a visualization supports learning [5]. In addition, we saw one student adopting a strategy of imagining the DCN visualization in their head, which necessitates a considerable amount of representational competence. This suggests that high connective understanding could lead to smaller differences in performance for those students who adopt such strategies. Therefore, in future studies, it will be necessary to measure representational competence not only in terms of the ability to translate between representations, but also in terms of conceptual ability to reason within one representation. As we have seen from the qualitative results of this study, this reasoning is very context dependent, as the choice of representation and strategies differs per student depending on the question type. In the following studies, representational understanding should also be measured utilizing the same contexts as those that are tackled.

Many students answered the representational questions perfectly, and the number of different representational

competence scores was not high. Therefore, it should be considered to measure representational competence more thoroughly in future studies. This could be done, for example, by increasing the number of questions focusing on connective understanding, and finding other ways to measure representational competence that also include visual/conceptual understanding of CN or DCN and quantum operations in these visualizations. Future studies could further investigate the competence of students when presented with only a visualization and how this representational competence impacts the possible benefits of presenting a visualization in addition to DN. Perhaps also, mental-rotation ability can help explain possible beneficial effects of DCN [85].

In general, we saw indications of reductions in perceived cognitive load by presenting CN or DCN. Due to the consistency with which this occurs, it is possible that there are beneficial effects regarding a reduction of cognitive load when students are presented with a visualization. A decrease in extraneous cognitive load should, in theory, lead to improved learning outcomes when learning with visualization, provided this decrease coincides with an increase in germane cognitive load [86,87]. If we also find a decrease in intrinsic cognitive load, it could be that the questions become inherently easier when asked and solved in CN or DCN, which would mean they provide beneficial complementary strategies according to the DeFT framework. Another point to consider is that the cognitive load of learners could be context dependent. For example, for some question types, an effect on cognitive load from the presentation of a visualization may differ from other question types. In addition, learning with visualization could also improve the competency of participants when solving questions in uncompleted DN. Investigating further whether there is a benefit for students who are presented with visualization, and under which circumstances and in which contexts this occurs, could therefore be of great interest. Studies with higher numbers of participants should separate by experience with DN or quantum physics in general, as we expect higher reluctance toward mathematical notation for those who are less familiar with it.

Going beyond static multiple external representations, conveying processes in visualizations using video sequences and audio could be worth studying. Furthermore, investigating learning outcomes by including interactivity in the teaching process—for example, via gamification—could be an alternative path forward. There already exists a multitude of quantum games [88], and incorporating CN or DCN into these games is a possibility. Also, the possible beneficial effects of other visualization techniques and their context and learner dependency remain to be investigated in multiqubit systems.

In conclusion, there remains a large space to explore the benefits of visualizations in quantum information science and technology education. In this work, we show that it is worth exploring the benefits of visualizations of quantum

states due to a multitude of additional strategies that the visualizations enable. We presented test instruments to investigate the circumstances under which visualizations can support students in the field of QIST, as well as first results that narrow down promising contexts for future studies. In order to counteract ceiling effects, we identify possible adjustments in the measurement of representational competence. The other test instruments cover a broad range of contexts and can be adapted easily for further research, also with other visualizations. However, we recommend more focused studies that use eye tracking and are much more restricted in question types. Restricting the question types will decrease the influence of previous knowledge on the variance of the results. Promising contexts include entanglement, the CNOT gate, and the Hadamard gate in multiqubit systems, which are also some of the most important concepts of quantum computing. We expect future studies structured in this way to result in valuable takeaways for educators and consider this path forward to be highly relevant to the quantum ecosystem.

**ACKNOWLEDGMENTS**

We thank the research initiative Quantum Computing for Artificial Intelligence at the RPTU Kaiserslautern-Landau for covering the study expenses. E. R., J. B., P. L., M. K.-E., and A. W. acknowledge support by the German Federal Ministry of Education and Research (BMBF) in the

QuantTUK project under Grant No. 13N15995. E. R. and A. W. acknowledge support by the European Union via the Interreg Oberrhein programme in the project Quantum Valley Oberrhein (UpQuantVal). This work was supported by LMUexcellent, funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder. All procedures performed in the study were in accordance with the ethical standards of the national research committee and the Declaration of Helsinki of 1964 and its subsequent amendments. The study involved data collection in three online surveys that took about 45–60 min each and interviews with some of the students. Participation in this study was voluntary and anonymous, and under informed consent. The data collected included participants’ age, gender, field of study, and highest educational achievement. Personal data of participants is kept confidential and used solely for research purposes.

**DATA AVAILABILITY**

The data that support the findings of this article are openly available [89].

**APPENDIX A: QUALITATIVE RESULTS**

In this section, data collected from the five students in the qualitative part of the study are shown; in Appendix A 1 the

TABLE VI. Average scores for the students who partook in the qualitative study. Scores in brackets are weighted.

No. of qubits	One-qubit			Two-qubit			Three-qubit				
	Representational competence		Score	Representational competence		Score	Representational competence		Score		
	Symbolic	Visual		Connectional understanding	Without vis.		With vis.	Connectional understanding		Without vis.	With vis.
Student no.											
1	1(0.63)	0.5(0.25)	0.5	0.67	0.33	0.5	0.33	0.33			
2	0.5(0.5)	0.5(0.5)	0.25	0.83	0.83	0.33	0.8	0.8	0.75	0.67	0.67
3	0.75(0.5)	0(0)	0.5	0.33	0.5	0.33	0.5	0.2			
4	1(0.88)	0.75(0.63)	0.75	0.67	1	0.67	0.8	0.6	1	1	0.9
5	1(1)	1(0.88)	1	1	0.83	0.83	1	0.8			

TABLE VII. Perceived cognitive load (intrinsic cognitive load, germane cognitive load, extraneous cognitive load, respectively) of the students who partook in the qualitative study.

No. of qubits	One-qubit		Two-qubit		Three-qubit	
	Without visualization	With visualization	Without visualization	With visualization	Without visualization	With visualization
Student no.						
1	2, 6, 1	1, 5, 0	4, 1, 1.5	6, 6, 1.5		
2	1, 3, 2	1, 4, 3	1, 4, 1.5	3, 4, 3	4, 5, 4	3, 4, 1.5
3	5, 4, 3.5	3, 4, 2	6, 6, 5.5	5, 5, 2.5		
4	3, 3, 1.5	5, 5, 1.5	5, 5, 3	5, 6, 3	6, 6, 2	5, 4, 0
5	5, 6, 2.5	2, 3, 0.5	4, 4, 4	2, 4, 0.5		

TABLE VIII. Representational competence problems of the students who partook in the qualitative study.

Question type:	Complex numbers		Circle notation		Translation
	Problems and strategies	Example	Problems and strategies	Example	
1	Complex numbers cannot be real	“It has to be either $-i$ or $+i$ because a complex number has to have an imaginary part to it.”	$+1$ corresponds to line to the right association of left side with negative values	“It’s going to the right side so that means positive.” “ $-\frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}$ because [...] the angle is completely on the y axis so when it goes on the left side it is completely negative.”	Views the $60^\circ$ angle as a $45^\circ$ angle positive angles go clockwise wrong denominator comparison “This is the half of $90^\circ$ so that is $45^\circ$ but it makes zero sense if it is $45^\circ$ .” “This is $+45^\circ$ so it goes on the right side.” “ $1/\sqrt{2}$ has to be smaller than $1/\sqrt{6}$ and $1/\sqrt{3}$ .”
2	$e^{-i\pi/2} = 1$ $e^{i\pi/4} = \frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}$	“It’s 1. Somewhat sure.” “It doesn’t have a minus so it’s in the opposite direction so the angle is in the opposite so it gets a minus in the complex plane.”	Unsure whether polar coordinates or circle notation convention	“If it is the qubit notation, it would be measured from here, if it is normal circle, it is measured from here, but it says <i>normal circle</i> .” <i>selects correct answer</i>	“ $e^{i\pi/4}$ is on the right hand side so it’s this” (selects circle corresponding to $\frac{1}{\sqrt{2}} 0\rangle + \frac{1}{\sqrt{2}}e^{-i\pi/4} 1\rangle$ .)
3	The number $\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i$ is a quantum state problems with power laws	“It looks like here the probabilities are in question like how probable it is for this state to be measured, like if it’s a one or a zero.” “ $e^{i\pi/6} \cdot e^{-i\pi/3} = e^{-i\pi/18}$ , because the $1/3$ goes into the $1/6$ and the minus into the plus.”	-1 corresponds to line to the left one circle representing one state that the qubit is in negative and positive makes negative	“Left is I assume the negative side of the axis” <i>When asked about one single full circle representing <math>e^{-i\pi/4}</math></i> “How am I supposed to pick a complex number here? because based on what I know this circle notation represents which state a qubit is in.” “Because this is a negative complex number and this is $\pi/2$ , the answer is negative.”	$e^{i\pi/4}$ is equal to a circle corresponding to $e^{-i\pi/4}$ Positive angles are clockwise
4	...	...	Circle corresponding to $e^{-i\pi/4}$ ; is equal to $-\frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}$	“I have the double minus, for double plus I go left around and for double minus I go right around.”	...
5	...	...	...	...	...

TABLE IX. Identified problems and strategies in the one-qubit survey. Indices are connected to examples and quotations in Sec. V B.

Question type:	Z gate			Phase gate				
	With visualization	Without visualization	Problems and strategies	With visualization	Without visualization	Problems and strategies		
Condition:								
Student no.	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example		
1			Calculation by addition	“I just did the maths, if $\varphi$ is positive, you arrive at $-2\pi/3 + \pi = \pi/3$ .”	Positive angle means counterclockwise	“I know the $P(\pi/3)$ changes the state by eighty degrees, so that’s the only reason I picked this option.”	Correct strategy: adding the angles	“Simply because the phase change should be $-3\pi/4$ and the math says it’s $-7/20\pi$ .”
2			Easily solved	“It’s subtracting by, no it adds.”	Relative phase = global phase	“It’s applied to both, so it’s gonna stay the same [state].”	Easily solved	“Just adding $-3\pi/4$ and $2\pi/5$ is $-7\pi/20$ .”
3	Confusing Z with NOT gate	“This one $[\frac{\sqrt{3}}{2} 0\rangle + \frac{1}{2}e^{2\pi/3} 1\rangle]$ , because the Z gate swaps the values.”		Approximation		“If you do the calculations $-3\pi/4$ and $2\pi/5$ , $-3$ by $2$ and $4$ by $5$ , $-7\pi/20$ should be the closest.”	Multiplying phases instead of adding	“The $-$ is there because of the $-$ in $e^{-i\pi/5}$ .”
4	Easily solved	“We just flip the phase of $ 1\rangle$ by $\pi$ .”		Circle for approximation, maths for checking answer		“It flips the phase by almost $\pi$ in the y direction, and this circle seems like it fits, but I computed it in my head and yeah it fits.”	Easily solved	“ $-3\pi/15 + 5\pi/15$ is $2\pi/15$ .”
5	Visual strategy	“Just rotate this here by $\pi$ .”		Visual strategy		“I think I would rotate this somewhere here because this would be a rotation of $-\pi$ and then $3/4$ times.”	Easily solved	“Here I add $\pi/3$ to $-\pi/5$ and if I calculated correctly it should be $2/15$ .”

TABLE X. Identified problems and strategies in the one-qubit survey. Indices are connected to examples and quotations in Sec. VB.

Question type:	Global phase			X gate		
	With visualization	Without visualization	Without visualization	With visualization	Problems and strategies	Without visualization
Student no.	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example
1	Equal angle change to both lines in the circle	“This is something my professor taught me very well. Global phase means equal change to both circles.”	Strategy to solve unknown	“Don’t know the logic on how to solve this without having some kind of phase given.”	Amplitude blindness	“I know the NOT gate reverses everything, so I first looked at the amplitudes [...]” <i>picks option where the amplitudes are not swapped.</i>
2	Rotation in circles, mathematics to check	“So just these two are rotated the same. This one. It’s rotated by $\pi/10$ .”	Checks denominators first, correct calculation	“The denominators are all correct. This one. Just adding $-2\pi/5$ .”	Confusion due to nonobservance of global phase	“I initially thought it also changed the phase but apparently it does not change the phase and just swapping the bits directly?”
3	Multiples of adding	“This one because $-2\pi/5$ times $2\pi/7$ equals $-4\pi/35$ ”	Chooses answer with reasoning independent of question	“This one $[\frac{1}{\sqrt{2}}e^{i\pi/5} + \frac{3}{\sqrt{2}}e^{-i\pi/5}]$ , $\pi/5$ gets applied to both states”	Disregards phase	The X gate basically shifts the values” <i>points at <math>\sqrt{3}/2</math> and <math>1/2</math></i>
4	Symbolic strategy	“ $2/7$ is $10/35$ and $-2/5$ is $-14/35$ so if I apply $-14$ to $10$ I get $-4$ .”	Starting difficulties	“That $[\frac{1}{2}e^{i\pi/10} 0\rangle + \frac{\sqrt{3}}{2}e^{-i\pi/10} 1\rangle]$ doesn’t look like a global phase to me.”	Confusion due to nonobservance of global phase	“so in my opinion everything is wrong because in my head I just switch $ 0\rangle$ and $ 1\rangle$ but nothing matches.”
5	Visual strategy	“I think I would rotate this somewhere here because this would be a rotation of $-\pi$ .”	Factoring out	“Here you can factor out the $e^{i\pi/10}$ , then you would have $e^{-2i\pi/10}$ which is $1/5$ .”	Remembering global phase changes	“NOT gate flips the state of the qubits so I would I expect same state of 1 here at the 0 again so I think maybe the global phase has changed. If I would factor $e^{-i\pi/2}$ out then I would have $e^{i\pi/2}$ and it’s the first answer.”

TABLE XI. Identified problems and strategies in the H<sub>gate</sub> and measurements questions of the one-qubit survey (students nos. 1–3). For a summary, see Sec. VB.

Question type:	H <sub>gate</sub>			Measurements			
	With visualization	Without visualization	Without visualization	With visualization	Without visualization	Without visualization	
Student no.	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example	
1	Unidentified problem with H-gate on $\frac{1}{\sqrt{2}} 0\rangle + \frac{1}{\sqrt{2}}e^{i\pi/4} 1\rangle$	“Why waste time.”	Unidentified problem with H-gate on $\frac{1}{\sqrt{2}} 0\rangle + \frac{1}{\sqrt{2}}e^{i\pi/2} 1\rangle$	“I don’t get the question.”	Absolute values = measurement probabilities	“By just looking at the numbers, 0.45 and 0.9” <i>picks option that measuring 1 is twice as likely as measuring 0</i>	Approximation “0.5 is 0.025 and I would consider this at 0.9 squared is 0.081. [...] 3 times 0.025 is 0.075 I think that kind of makes sense.”
2	$H(\frac{1}{\sqrt{2}} 0\rangle + \frac{1}{\sqrt{2}}e^{i\pi/2} 1\rangle) =  1\rangle$	“I find it weird the  1⟩ is missing. [...] I am not sure what’s going on here I thought the H gate collapses the value.”	Knows correct strategy in principle, but too difficult to execute	“What happened here? Is it supposed to be this way? Ah the phase is gonna affect it. $\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} \cdot e^{i\pi/4}$ , so $\sqrt{2}$ goes out, and $\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} \cdot e^{i\pi/4}$ . But what is that? I have no clue.”	Derives probabilities from square roots in state depiction	“It’s supposed to be in the root format so it’s squared I think.”	“Measuring 0 is three times as likely as measuring 1.”
3	Hadamard gate creates an equally weighted superposition from any state	“The Hadamard gate is supposed to create an equally weighted superposition. None of them is equally weighted so I am picking a random one.”	Easily solved	“It’s on the y axis.”	Not squaring the amplitudes	“if you take out the $\sqrt{3}$ I think the chance of measuring 0 is $\sqrt{3}$ times measuring 1.”	“Because 0.9 is twice 0.45, measuring 1 is twice as likely as measuring 0.”

TABLE XII. Identified problems and strategies in the H gate and measurements questions of the one-qubit survey (students no. 4 and no. 5). For a summary, see Sec. VB.

Question type:	H gate				Measurements			
	With visualization		Without visualization		With visualization		Without visualization	
Student no.	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example
4	Matrix multiplication	“In my head I just applied the matrix of the Hadamard gate to this state and I had $1 + e^{i\pi/4}$ and $1 - e^{i\pi/4}$ and I flipped so the $ 0\rangle$ has no phase and this is then $-\pi/2$ .”	Lack of strategy	“I can’t really explain that. Intuition.”	Mathematical strategy	“The probabilities are 3/4 and 1/4 so 0 is three times more likely than 1.”	Easily solved by calculation	“Because if I take the absolute values squared it’s 20% vs 80% and 80% is four times 20%.”
5	Uncertain	“When I try to apply the H gate to each qubit individually then add the amplitudes. But I’m not sure.” <i>selects correct answer</i>	Cannot compute	“If I apply the H gate to each qubit then I get the equal superposition of $ 0\rangle$ and $ 1\rangle$ and here $ 0\rangle$ and $- 1\rangle$ . I can’t compute it in my head.”	Easily solved by calculation	“Measuring 0 is three times as likely as measuring 1 so I get 3/4 and 1/4.”	Comparing in root notation	“ $1/\sqrt{5}$ and $2/\sqrt{5}$ makes 4/5 so it’s four times as likely.”

TABLE XIII. Identified problems and strategies in the X- and H-gate questions of the two-qubit survey. For a summary, see Sec. V C.

Question type:	X gate				H gate			
	With visualization		Without visualization		With visualization		Without visualization	
Condition:	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example
1 (CN)	Interchanging coefficients next to each other	“They interchange, so what was $1/3$ will become $1/\sqrt{3}e^{i\pi/4}$ and this will become $1/3$ and similar in the second half.”	$X_2$ means two swaps	“The $X_2$ gate is different from the $X_1$ gate, it changes either with the second to last or the last one. Then if there is a NOT gate these two must change as well.”	Hadamard gate always creates equal superpositions	“I was thinking equal probability would go to all. But that’s not happening. Probability is being divided between the first two and the last two and I have no reason for that. So I am going to pick a random answer.”	Lack of strategy	“I’m going with this one.”
2 (CN)	Easily solved by coefficient comparison	“ $X_2$ so it’s switching these two $ 00\rangle$ and $ 10\rangle$ .”	Easily solved by coefficient comparison	“The ones were the first bit is changing are swapped.”	Visual strategy	“First one so next to each other. First one is state $ 1\rangle$ because it’s negative and second has 0 relative [phase] so it’s in the $ 0\rangle$ .”	Splitting states	“This $[1/\sqrt{3} 00\rangle]$ would be split in this and this $[1/\sqrt{6} 00\rangle$ and $1/\sqrt{6} 01\rangle]$ and here since we are starting from 1 this would get a negative.”
3 (DCN)	Mistaking $X_2$ gate for CNOT <sub>21</sub>	“When qubit 2 is one these two should switch so that happens over here.”	Easily solved by coefficient comparison	“this one because these two shift and these two shift values.”	Visual reasoning, mistaking $H_2$ for $H_1$ gate	“Along the $x$ axis the values have an equally weighted superposition.”	Lack of strategy	“It applies when the first qubit is 1. So that’s why, here the first qubit is 1.” <i>selects</i> $\frac{1}{\sqrt{2}} 00\rangle - \frac{1}{\sqrt{2}} 10\rangle$
4 (DCN)	Visual strategy	“ $X_1$ as in the introduction just flips the circles around the axis of qubit 1.”	Easily solved by coefficient comparison	“So the first number is the second qubit. So this [first 0 in $ 00\rangle$ ] is flipped to 1.”	Wrong qubit	“ $ 10\rangle$ should get this $e^{i\pi/4}$ .”	Canceling out of amplitudes	“ $ 00\rangle$ gets erased and $ 11\rangle$ gets erased so I think its $[\frac{1}{\sqrt{2}} 00\rangle - \frac{1}{\sqrt{2}} 10\rangle]$ ”
5 (DCN)	Visual strategy	“Should be this answer because here these colors are swapped [colored inner circles].”	Easily solved by coefficient comparison	“I have to flip the qubit 2, the $ 00\rangle$ would then correspond to the $ 10\rangle$ state and we have the same coefficients.”	Visual strategy	“Here we distribute the probability from the $ 00\rangle$ state to the $ 10\rangle$ state and not change the angle, [...] and here to the $ 01\rangle$ state and rotate the angle where the qubit is 1.”	Apply reverse gate	“If I apply the $H_1$ gate to this state because it is it’s own inverse, I should end up in this state.”

TABLE XIV. Identified problems and strategies in the guess-the-gates and measurements questions of the two-qubit survey (students nos. 1–3). For a summary, see Sec. V C.

Question type:	Guess-the-gates				Measurements			
	With visualization		Without visualization		With visualization		Without visualization	
Student no.	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example
1 (CN)	...	...	...	...	A circle as a qubit, measurement result 0 corresponds to measuring 00	“If it was just two qubits I would just square it and make an educated guess, but now the 1 probability is being divided between four circles.” “It would make sense if I just remove this clutter and only look at this [00] and [11].”	Lack of knowledge where qubit 1 can be found in the Dirac notation	“I don’t understand the question. Both these bits have 01 and 10. I can’t tell the difference.”
2 (CN)	Visual strategy	“Hadamard on the second one. [...] and phase shift.”	Intuition	“I feel like here its phase was shifted and it was a NOT” <i>picks correct answer</i> $H_2Z_1$	Calculations in symbolic notation	“1/3 + 2/9 is 5/9, 1/9 + 1/3 is 4/9. So it’s 5/4”	Calculating	“(1/2) <sup>2</sup> plus (1/2) <sup>2</sup> is 3/8 and on P(1) it was 5/8”
3 (DCN)	Z gate does a $\pi/2$ phase shift	“1/√2 was given to both sides for an equal superposition. And then there was a phase shift applied by $\pi/2$ .”	Disregard for gate indices, correct solution	“A Hadamard gate was applied because it delegated superpositions equally. And it’s a not gate because the states have switched and I am assuming it switched on the x axis again.”	Approximation, looking at wrong values	“Just looking at the values, measuring 1 is √2/3 or something it’s 1/√3 for 0”	Approximation, no squaring, wrong qubit	“0 has a probability of 1/2√2 plus √3/2√2 and 1 has a probability of 1/2 plus 1/2”

TABLE XV. Identified problems and strategies in the guess-the-gates and measurements questions of the two-qubit survey (students no. 4 and no. 5). For a summary, see Sec. VC.

Question type:	Guess-the-gates				Measurements			
	With visualization		Without visualization		Without visualization			
Student no.	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example		
4 (DCN)	Narrow down with patterns, but problem computing	“I don’t think that an X gate was applied because we have this $1/\sqrt{6} 00\rangle$ which we get by applying a Hadamard to the first or the second qubit. I also have the impression somewhere was a Z gate. I don’t know. First intuition was $H_1Z_2$ .”	Coefficient comparison	“I can’t get rid of this $\sqrt{3}$ in this state $ 00\rangle$ so there has to be an X gate somewhere after applying the Hadamard. So then maybe a flip of the second one.”	Issue with percentages	“This is $1/8 + 1/4 = 3/8$ and this should add up to $5/8$ . So what is $5/8$ over $3/8$ ? So I think measuring 1 is 33% more likely than measuring 0, because 66% seems too much.”	Incorrect squaring	“Here I have $1/3$ and here I have $2/6$ so they add up to $4/6$ . I don’t remember the formula [...] but if the probabilities are doubled then 50% more is more likely than 25% more.”
5 (DCN)	Visual strategy	“ $Z_1$ would rotate this angle to the other side and then $H_2$ would distribute the amplitudes around the qubit 2 axis.”	Focussing on one basis state in the superposition	“If I flip the second qubit $ 01\rangle$ becomes $ 11\rangle$ and then the H gate would give this expression then.”	Use of insufficient small amount of information	“I think it’s 33% more likely because of this $\sqrt{3}$ here but I’m not sure.”	Approximation	“I’m not completely sure but we have only $\sqrt{2}$ more so I think it’s only 25% more likely [to measure 0 than 1].

TABLE XVI. Identified problems and strategies in the separability and entanglement questions of the two-qubit survey (students no. 1 and no. 2). For a summary, see Sec. V C.

Question type:	Separability			Entanglement		
	With visualization	Without visualization	Problems and strategies	With visualization	Without visualization	Problems and strategies
Student no.	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example
1 (CN)	Lack of strategy, separability as a process rather than a property	“If I was given the input and output of two qubits, and ask a question regarding is that separable or not, that would have been a far easier problem to solve. This I cannot fathom how to solve.”	Lack of strategy, separability as a process rather than a property	“As I explained previously. If there was some change made to the qubits and shown what the results was I would have been able to tell.”	Entanglement means that action on one also is action on the other, four basis states = four qubits	Lack of strategy, entanglement means that action on one also is action on the other impact on the other qubit.”
2 (CN)	Visual strategy	“[Going through answer possibilities] Not following the trend. Not following the trend. This is separable because it is following the trend, its a shift on this and a similar shift on this.”	Lack of strategy	“I don’t know how it’s written in the bracket notation, I’m going to pick the third [correct] one, no reason.”	Checking for separability instead of entanglement	Incorrect phase shift application “If I do a $-1/3$ on this $[-\pi/6]$ it’s gonna change, then it won’t be separable.”

TABLE XVII. Identified problems and strategies in the separability and entanglement questions of the two-qubit survey (students nos. 3–5). For a summary, see Sec. V C.

Question type:	Separability			Entanglement		
	With visualization	Without visualization	Problems and strategies	With visualization	Without visualization	Problems and strategies
3 (DCN)	<p>Problems and strategies</p> <p>Example</p> <p>“I am probably not getting the idea of the separable states correctly. It feels like if I calculate the first bit somehow this will also impact the second bit.”</p>	<p>Problems and strategies</p> <p>Entanglement = dependency</p> <p>Example</p> <p>“I cannot determine how they will not depend on each other. It feels like they are all entangled. No matter how I calculate, measurement on the first one is going to impact qubit two as well, because their states are together.”</p>	<p>Problems and strategies</p> <p>Intuition</p>	<p>Problems and strategies</p> <p>Example</p> <p>“I don’t know how am I supposed to determine if I calculate this whether I get a separable state or not. Both the separable and entangled states are still a vague concept to me. [...] Feels like these two states might be entangled based on the coefficients.”</p>	<p>Problems and strategies</p> <p>Lack of strategy</p> <p>Example</p> <p>“All of them look entangled. I probably don’t understand the topic that well.”</p>	
4 (DCN)	<p>Problems and strategies</p> <p>Visual strategy: amplitude comparisons</p> <p>Example</p> <p>“This looks very nice maybe. Because from this to this and from this to this. The others don’t make sense.”</p>	<p>Problems and strategies</p> <p>Imagining circle notation and phase shift comparison</p> <p>Example</p> <p>“I just imagined the circle notation. If I have <math>-\pi/2</math> and going <math>\pi/3</math> to the left direction I should end up in <math>-\pi/6</math>.”</p>	<p>Problems and strategies</p> <p>Visual strategy: phase shifts</p>	<p>Problems and strategies</p> <p>Imagining circle notation and amplitude comparison</p> <p>Example</p> <p>“Here you need a negative shift and here you need a positive shift.”</p>	<p>Problems and strategies</p> <p>Imagining circle notation. I just computed the probabilities of the left side and tried to find a factor to get these on the right side. And it worked for these three and not for this.”</p>	
5 (DCN)	<p>Problems and strategies</p> <p>Visual strategy</p> <p>Example</p> <p>“Here we could multiply the amplitudes by the same factor from the left to the right.”</p>	<p>Problems and strategies</p> <p>Using the rule <math>\alpha_{00}\alpha_{11} = \alpha_{01}\alpha_{10}</math></p> <p>Example</p> <p>“If I multiply <math>1/2</math> with the factor of this then I get <math>1/4e^{-i\pi/6}</math> and if I multiply these I get the same.”</p>	<p>Problems and strategies</p> <p>Visual strategy</p>	<p>Problems and strategies</p> <p>Using the rule <math>\alpha_{00}\alpha_{11} = \alpha_{01}\alpha_{10}</math></p> <p>Example</p> <p>“We would rotate from the left to the right the angles in this direction and here in the other direction and it would not be separable.”</p>	<p>Problems and strategies</p> <p>Using the rule <math>\alpha_{00}\alpha_{11} = \alpha_{01}\alpha_{10}</math></p> <p>Example</p> <p>“If I multiply these two numbers I do not get the same as when I multiply these two.”</p>	

TABLE XVIII. Identified problems and strategies in the X- and Z-gate questions of the three-qubit survey. For a summary, see Sec. VD.

Question type:	X gate			Z gate		
	With visualization	Without visualization	Problems and strategies	With visualization	Without visualization	Problems and strategies
2 (DCN)	<p>Problems and strategies</p> <p>Example</p> <p>“ 000⟩ is 1/4. This should be <math>1/\sqrt{2}</math>.”</p>	<p>Problems and strategies</p> <p>Example</p> <p>“There is no  100⟩. So this one is the right one.” <i>checks the other terms.</i></p>	<p>Key feature search</p>	<p>Problems and strategies</p> <p>Example</p> <p>“ 101⟩ and  111⟩ should just stay the same. Does it not switch? The option I picked <math>\frac{1}{\sqrt{6}}e^{2i\pi/3} 001\rangle + \frac{1}{\sqrt{6}} 010\rangle + \frac{1}{\sqrt{6}}e^{3i\pi/4} 110\rangle - \frac{1}{\sqrt{6}} 101\rangle + \frac{1}{\sqrt{6}}e^{i\pi/2} 110\rangle - \frac{1}{\sqrt{6}} 111\rangle</math> had this and this [ 001⟩,  011⟩, and  101⟩,  111⟩] both on the opposite side because they had similar phases.”</p>	<p>Problems and strategies</p> <p>Example</p> <p>Intuition, Z gate switches phases instead of a <math>\pi</math> relative phase shift</p>	<p>Example</p> <p>“This one seems more correct. I don’t know. This one [ 000⟩] and this one  001⟩ should have swapped the phase but I don’t see that anywhere.”</p>
4 (DCN)	<p>Visual strategy: action in direction of axis</p> <p>“The <math>X_3</math> gate swaps this axis. So  000⟩ should get 0 probability and the third qubit is 1 so from left to right it gets bigger so I go with this.”</p>	<p>Problems and strategies</p> <p>Example</p> <p>“The <math>X_1</math> gate flips for example this  000⟩ with this  001⟩ so the amplitudes for this state is the only one where this matches.”</p>	<p>Coefficient comparison</p>	<p>Problems and strategies</p> <p>Example</p> <p>“If the phase of the first qubit is flipped than I have to look at these four and this is the only one where all are flipped.”</p>	<p>Problems and strategies</p> <p>Example</p> <p>Phase shift examination</p>	<p>Example</p> <p>“We add <math>\pi</math> if the second qubit is 1. We get the — here [ 010⟩] we get from <math>-\pi/4</math> to <math>3\pi/4</math>, from <math>\pi/2</math> to <math>-\pi/2</math> that’s also correct and here is just the minus.”</p>

TABLE XIX. Identified problems and strategies in the H-gate questions of the three-qubit survey. For a summary, see Sec. VD.

Condition:	With visualization		Without visualization	
Student no.	Problems and strategies	Example	Problems and strategies	Example
2 (DCN)	Visual strategy along qubit axis	Because superposition in this $[ 000\rangle]$ and superposition in this $[ 011\rangle]$ , these two are reduced $[ 001\rangle,  101\rangle]$ and these two are reduced $[ 010\rangle,  110\rangle]$	Key feature search followed by answer validation	“Because it is on the second bit, this one $[ 000\rangle]$ and this one $[ 010\rangle]$ come together. This one doesn’t have anything, so it makes another one and same over here.”
4 (DCN)	Coefficient analysis	“If I apply H on the second, the amplitude of 0 just be $\sqrt{2}/3$ , so this or this. $ 100\rangle$ has a positive sign. If I apply H here $[\sqrt{2}/3 110\rangle]$ I get a positive sign. Only 1 gets a negative sign.”	Problem identifying “partner” state	“Hadamard to third I have only one times $ 000\rangle$ after that with amplitude $1/\sqrt{12}$ so $1/2\sqrt{3}$ so [one of] these two. This $[ 001\rangle]$ produces $-1/\sqrt{12}$ or not? And this also. Here $[1/\sqrt{6} 011\rangle]$ I would get $1/\sqrt{12}$ and that $[ 111\rangle]$ is the only possible appearance.

performance scores and cognitive load, in Appendix A 2 the problems with the representational competence questions, and in Appendixes A 3–A 5, their problems and problem-solving strategies in the main parts of the one-, two-, and three-qubit survey, respectively.

### 1. Overall scores

Table VI shows the average performance scores of the students who took part in the qualitative study. Table VII shows the perceived cognitive load.

### 2. Representational competence

Table VIII shows the problems that emerged in the representational competence parts of the survey of those students who took part in the qualitative study.

### 3. One-qubit survey

Table IX shows the problems and problem-solving strategies in the Z-gate and phase-gate questions of the one-qubit survey of the students who took part in the qualitative study. Further, Table X shows the problems and problem-solving strategies in the global phase and X-gate

questions of the one-qubit survey, and Tables XI and XII show those of the H-gate and measurement questions.

### 4. Two-qubit survey

Table XIII shows the problems and problem-solving strategies in the X-gate and H-gate questions of the two-qubit survey of the students who took part in the qualitative study. Further, Tables XIV and XV show the problems and problem-solving strategies in the guess-the-gates and measurement questions, and Tables XVI and XVII show those of the separability and Entanglement questions.

### 5. Three-qubit survey

Table XVIII shows the problems and problem-solving strategies in the X-gate and Z-gate questions of the three-qubit survey of the students who took part in the qualitative study. Table XIX shows those of the Hadamard gate question and Table XX those of the measurement questions. Further, Table XXI shows the problems and problem-solving strategies in the separability and entanglement questions, and Table XXII shows that of the CNOT and guess-the-gates questions.

TABLE XX. Identified problems and strategies in the measurement probabilities and measurement state questions of the three-qubit survey. For a summary, see Sec. VD.

Question type:	Measurement probability			Measurement state			
	With visualization	Without visualization	With visualization	Without visualization	With visualization	Without visualization	
Student no.	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example	
2 (DCN)	Calculations	“0 is this $1/\sqrt{6} 000\rangle$ and this $1/\sqrt{6} 100\rangle$ so 1/3 and 1 2/3. Measuring 1 is twice as likely as measuring 0.”	Calculations	“From taking the probabilities where this one [qubit 2] is 0, it came 1/3 and this one [where qubit 2 is 1] was 2/3, so 1 is two times as likely as 0.”	Intuition, misconception: basis states are merged and not erased	Lack of understanding of normalization	“I don’t know how it’s normalized.” <i>picks a random answer</i>
4 (DCN)	Coefficient calculations	“The probability of getting 0 for the second qubit should be 1/3 and for getting 1 it should be 2/3 so measuring 1 is twice as likely as measuring 0.”	Coefficient calculations	“So the probability of 0 should be 1/3 and for 1 2/3 so 1 is twice as likely as 0.”	Visual strategy: phase comparison	Probability calculations to get normalization constant	“We have to divide the probabilities I think it was the square root of 1/2 so 1/4. No it’s $1/\sqrt{2}$ , so the amplitude over here [001] should be $1/\sqrt{6}$ , that’s fine. And here [100] it should be $1/\sqrt{2}I$ guess. I can check the last one $1/\sqrt{3}$ yeah, should be correct.”

TABLE XXI. Identified problems and strategies in the separability and entanglement questions of the three-qubit survey. For a summary, see Sec. V D.

Question type:	Separability			Entanglement		
	With visualization	Without visualization	Problems and strategies	With visualization	Without visualization	Problems and strategies
2 (DCN)	<p>Problems and strategies</p> <p>Example</p> <p>“I don’t really know what fully separable means. One of the first three. They’re all similar.”</p>	<p>Problems and strategies</p> <p>Example</p> <p>“I don’t know this one either. This one <math>\frac{1}{2\sqrt{2}} 000\rangle + \frac{1}{2\sqrt{2}} 001\rangle + \frac{1}{2\sqrt{2}}e^{i\pi/2} 010\rangle + \frac{1}{2\sqrt{2}}e^{i\pi/2} 011\rangle + \frac{1}{2\sqrt{2}} 100\rangle + \frac{1}{2\sqrt{2}} 101\rangle + \frac{1}{2\sqrt{2}}e^{i\pi/2} 110\rangle + \frac{1}{2\sqrt{2}}e^{i\pi/2} 111\rangle</math> has common values in a weird way.”</p>	<p>Lack of strategy</p> <p>Lack of strategy, picks correct answer by intuition</p>	<p>Problems and strategies</p> <p>Example</p> <p>“This one <math>\frac{1}{2\sqrt{2}} 000\rangle + \frac{1}{2\sqrt{2}} 001\rangle + \frac{1}{2\sqrt{2}} 010\rangle - \frac{1}{2\sqrt{2}} 011\rangle + \frac{1}{2\sqrt{2}} 100\rangle - \frac{1}{2\sqrt{2}} 101\rangle + \frac{1}{2\sqrt{2}} 110\rangle + \frac{1}{2\sqrt{2}} 111\rangle</math> is different in all directions.”</p>	<p>Lack of strategy</p> <p>lack of strategy</p>	<p>Example</p> <p>“Nothing specifically, I don’t know which it could be.”</p>
4 (DCN)	<p>Problems and strategies</p> <p>Example</p> <p>This is <math>\pi/2</math>, this should be <math>\pi/2</math>, so no. [next answer possibility] For the second qubit this is <math>\pi/2</math> and this is <math>-\pi/2</math>, so no. This is <math>\pi/2</math>, this is 1, and this is 1, seems fully separable.”</p>	<p>Problems and strategies</p> <p>Example</p> <p>“Should be this one when I think about the circle notation then they all should, ah yeah the second qubit is always 1. So is factor 0 allowed? Okay let me check the others again. It does not work for this. This does also not work.”</p>	<p>Visual strategy: comparison along planes</p> <p>Visual strategy: comparison along planes</p>	<p>Problems and strategies</p> <p>Example</p> <p>“Here if I draw the plane for the first qubit it’s just 1. This plane does not work. This plane does not work. Here the plane for the third qubit has factor one. I think this is the fully entangled one. There is no plane that has a symmetry.”</p>	<p>Imagining visualization</p> <p>Imagining visualization</p>	<p>Example</p> <p>“So I imagined the circle representation in my head and checked it for the first one where only the <math> 011\rangle</math> and <math> 101\rangle</math> have a negative sign and they are direct neighbors in this area for the third qubit and then I looked at the other states, both states have a minus sign and they only differ in one qubit. If I draw a plane again I think they would be separable and this state differs in two different qubits which means I cannot draw a plane there.”</p>

TABLE XXII. Identified problems and strategies in the CNOT gate and guess-the-gates questions of the three-qubit survey. For a summary, see Sec. VD.

Question type:	CNOT gate			Guess-the-gates		
	With visualization	Without visualization	Problems and strategies	With visualization	Without visualization	Problems and strategies
Student no.	Problems and strategies	Example	Problems and strategies	Example	Problems and strategies	Example
2 (DCN)	Visual strategy	“Switch in the 2 direction this and this $ 100\rangle$ and $ 110\rangle$ and this and this $ 101\rangle$ and $ 111\rangle$ . This one seems correct.”	Key feature search followed by answer validation	“These two $ 001\rangle$ and $ 101\rangle$ switch. This one $ 111\rangle$ with that $ 011\rangle$ .”	X gate correctly identified, doesn't know how measurements work	“3 would be this [moves cursor along axis 3], 2 would be this [moves cursor along axis 2]. Could be [answer option] 2 or 3 [ $X_3 M_2$ or $X_3 M_1$ ]. I don't know what the difference would be.”
4 (DCN)	Visual strategy	“If the first qubit is 1 then the third gets flipped so if I have here 1 I flip on axis 3 so I flip this and this $ 001\rangle$ and $ 101\rangle$ and this and this $ 011\rangle$ and $ 111\rangle$ .”	Bit flip calculations	“Nothing should happen with this state $ 001\rangle$ so one of these two. So we flip if the third is 1. So it should be this one.”	Gates applied in wrong order	“It's a Hadamard on the first one, because I don't get another $ 000\rangle$ here [in the rest of the state]. And now a CNOT. Okay I have everything somehow [no amplitude is 0]. If I flip the first if the second is 1 then this $ 001\rangle$ state would do nothing but the $ 001\rangle$ state has a phase. So I guess I flip the second if the first is 0. Then it's this $ H_1\text{CNOT}_{12}\rangle$ .”

**APPENDIX B: QUANTITATIVE SURVEY RESULTS**

The quantitative data supporting the findings of this article is openly available [89].

Table III summarizes the average connectional understanding, average scores, and average times for correctly solved questions. In addition, it shows the differences in these performance metrics with visualization minus without visualization. Average intrinsic cognitive load (ICL), germane cognitive load (GCL), and extraneous cognitive load (ECL) and the corresponding differences (between with visualization and without visualization) are shown in Table IV. Figure 10 shows the average score per

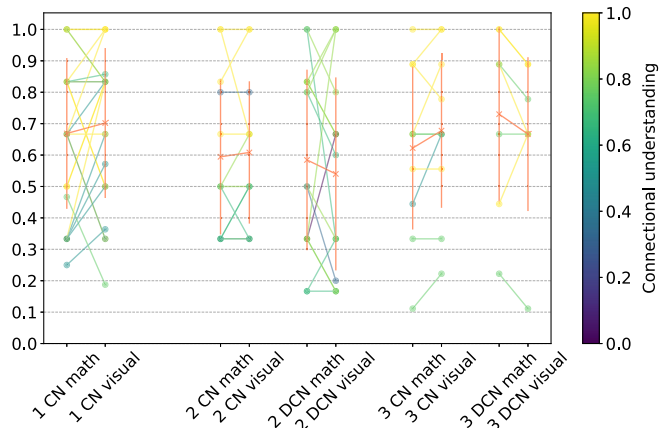


FIG. 10. Average scores for different numbers of qubits. Two points connected by a line represent an individual participant, colored according to the connectional understanding score of the participant in that survey. Mean and standard error are shown in orange.

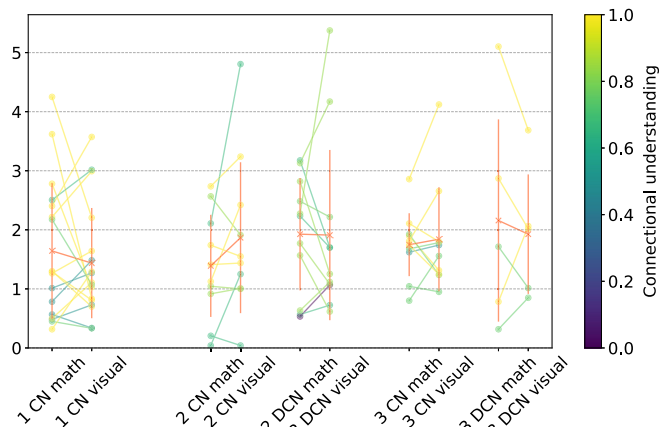


FIG. 11. Average time taken per participant for correctly solved questions and connectional understanding. Two points connected by a line represent an individual participant, colored according to the connectional understanding score of the participant in that survey. Mean and standard error are shown in orange. Times above 20 min were excluded.

participant, and Fig. 11 the average time taken for correctly solved questions. The influence of connectional understanding on ICL, GCL, and ECL as well as the average cognitive load is shown in Figs. 12–14 for tasks with one, two, and three qubits.

In the following, we show detailed results for single-qubit (Appendix B 1), two-qubit (Appendix B 2) and three-qubit studies (Appendix B 3). We describe the average correctness per question with and without visualization and survey parts A and B and the combined score of part A and B. Then, we do the same for average time taken for correct answers, considering for the combination of parts A and B only cases where both questions of each topic, with and without visualization, were answered correctly.

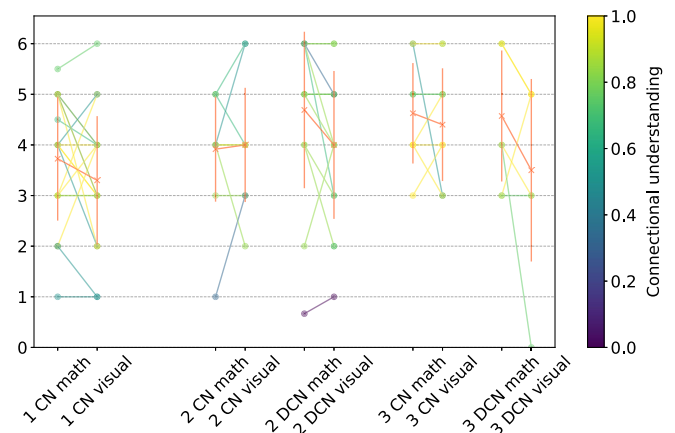


FIG. 12. Intrinsic cognitive load by number of qubits, without visualization (math) and with visualization. Two points connected by a line represent single participants, colored according to the connectional understanding score of the participant in that survey. Mean and standard error are shown in orange.

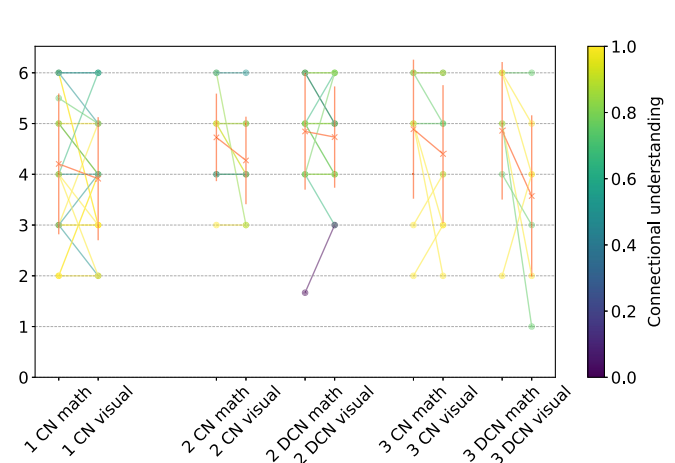


FIG. 13. Germane cognitive load by number of qubits, without visualization (math) and with visualization. Two points connected by a line represent single participants, colored according to the connectional understanding score of the participant in that survey. Mean and standard error are shown in orange.

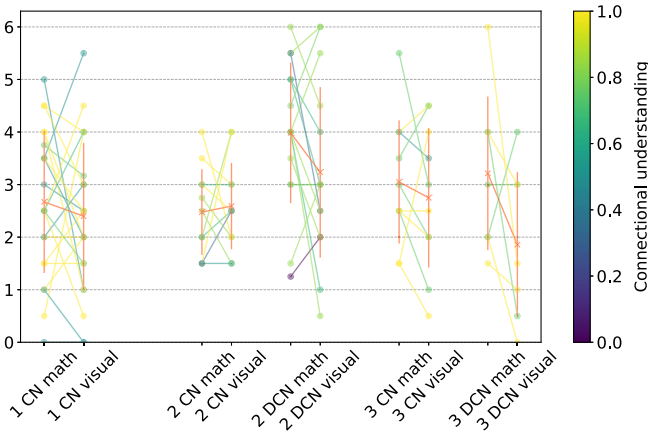


FIG. 14. Extraneous cognitive load by number of qubits, without visualization (math) and with visualization. Two points connected by a line represent single participants, colored according to the connectional understanding score of the participant in that survey. Mean and standard error are shown in orange.

### 1. One-qubit survey

The average scores for single-qubit questions are shown in Table XXIII and average times for correctly solved questions in Table XXIV. In the latter, the combined scores of part A and part B show only cases where participants managed to solve both questions, i.e. both questions on the same concept with and without visualization, correctly. Figure 15 shows the combined average score per question and Fig. 16 shows the combined average time taken (for all answers, not just the correct ones). Cases in which a participant took more than 20 min for one question were excluded.

During part A, group math-vis solved questions without visualization with an average score of  $0.74 \pm 0.14$  and an average time of  $1.60 \pm 0.64$  min for correctly solved

questions, while group vis-math answered the same questions with visualization with an average score of  $0.66 \pm 0.28$  and an average time of  $2.28 \pm 1.06$  min. During part B, the average scores were  $0.57 \pm 0.17$  (group vis-math) and  $0.6 \pm 0.11$  (group math-vis), respectively, at average times for correctly solved questions of  $1.57 \pm 0.63$  and  $1.78 \pm 1.17$  min. When combining parts A and B of the survey, the average scores were  $0.66 \pm 0.12$  without visualization and  $0.63 \pm 0.12$  with visualization, at  $1.55 \pm 0.51$  and  $2.07 \pm 1.1$  min.

In survey part A, group math-vis scored higher than group vis-math on the global phase, X-gate, phase-gate, H-gate, measurement probability, and measurement state questions with an average score of  $0.67 \pm 0.13$  to  $0.5 \pm 0.24$  on these questions and did not score lower on any questions. Group math-vis took longer on correct answers to the X-gate, phase-gate, H-gate, and measurement state questions with an average of  $151 \pm 74$  to  $56 \pm 22$  s while taking less time on the global phase equivalence and measurement probability questions with  $84 \pm 4$  to  $75 \pm 1$  s.

In part B of the survey, group math-vis scored higher than group vis-math on the global phase equivalence, Z-gate, H-gate, measurement probability, and measurement state questions, with an average of  $0.84 \pm 0.13$  to  $0.62 \pm 0.14$ . Group vis-math scored higher on the phase gate question, with 0.88 to 0.78. Group math-vis took more time on the global phase, Z-gate, and measurement state questions with  $178 \pm 120$  to  $140 \pm 85$  s and group vis-math took longer on phase-gate, H-gate and measurement probability questions with  $77 \pm 23$  to  $63 \pm 19$  s.

When presented with visualization, group math-vis scored higher on the global phase, H-gate, measurement probability, and measurement state questions with average scores of  $0.86 \pm 0.14$  to  $0.67 \pm 0.14$ . When presented questions with visualization, group math-vis took longer

TABLE XXIII. Average scores for one-qubit questions, with and without visualization ( $N = 17$ ). Note: Standard deviation is used for confidence intervals.

Survey part:	A		B		Combined	
	Without visualization <sup>a</sup>	With visualization <sup>b</sup>	Without visualization <sup>b</sup>	With visualization <sup>a</sup>	Without visualization <sup>c</sup>	With visualization <sup>c</sup>
Question						
Global phase	0.62	0.67	0.57	1.0	0.59	0.84
X gate	0.46	0.29	...	...	...	...
Z gate	...	...	0.54	0.75	...	...
Phase gate	0.77	0.64	0.92	0.83	0.85	0.73
H gate	0.69	0.57	0.54	0.67	0.62	0.62
Measurement probability	0.62	0.21	0.5	0.67	0.56	0.42
Measurement state	0.92	0.71	0.64	1.0	0.78	0.85
Mean	$0.68 \pm 0.14$	$0.52 \pm 0.19$	$0.62 \pm 0.14$	$0.82 \pm 0.14$	$0.68 \pm 0.11$	$0.69 \pm 0.16$

<sup>a</sup> $N = 12$ .

<sup>b</sup> $N = 11$ .

<sup>c</sup> $N = 23$ .

TABLE XXIV. Average time taken for correct answers to one-qubit questions, with and without visualization in minutes ( $N = 17$ ). Note: The number of participants  $N$  that solved the question correctly and are therefore included in the average is shown in brackets. Standard deviations are shown where  $N > 1$ . The confidence interval of the mean is the mean of the standard deviations.

Survey part:	A		B		Combined	
	Without visualization <sup>a</sup>	With visualization <sup>b</sup>	Without visualization <sup>b</sup>	With visualization <sup>a</sup>	Without visualization <sup>c</sup>	With visualization <sup>c</sup>
Global phase equivalence	$1.28 \pm 0.63(6)$	$1.48 \pm 1.57(5)$	$2.94 \pm 2.99(6)$	$2.95 \pm 2.2(10)$	$1.63 \pm 1.08(10)$	$2.26 \pm 2.27(10)$
X gate	$4.31 \pm 2.61(5)$	$1.52 \pm 0.92(3)$	...	...	...	...
Z gate	...	...	$3.69 \pm 5.56(5)$	$5.43 \pm 4.11(7)$	...	...
Phase gate	$1.65 \pm 2.49(8)$	$1.01 \pm 0.44(7)$	$1.19 \pm 1.05(9)$	$0.87 \pm 0.42(7)$	$1.47 \pm 1.99(11)$	$0.93 \pm 0.46(11)$
H gate	$2.93 \pm 2.9(7)$	$0.68 \pm 0.41(6)$	$1.78 \pm 2.74(6)$	$1.5 \pm 0.83(7)$	$2.13 \pm 2.59(10)$	$0.84 \pm 0.4(10)$
Measurement probability	$1.23 \pm 1.0(6)$	$1.33(1)$	$0.86 \pm 0.59(4)$	$0.81 \pm 0.58(6)$	$1.13 \pm 0.96(6)$	$0.91 \pm 0.61(6)$
Measurement state	$1.14 \pm 1.39(9)$	$0.55 \pm 0.24(7)$	$0.38 \pm 0.17(6)$	$0.54 \pm 0.68(9)$	$0.83 \pm 1.14(14)$	$0.55 \pm 0.55(14)$
Mean	$2.09 \pm 1.27$	$1.09 \pm 0.41$	$1.81 \pm 1.28$	$2.01 \pm 1.88$	$1.44 \pm 0.49$	$1.1 \pm 0.67$

<sup>a</sup> $N = 9$ .  
<sup>b</sup> $N = 8$ .  
<sup>c</sup> $N = 17$ .

on the global phase equivalence question with 177 to 77 s and took a shorter amount of time on the phase gate, H-gate, measurement probability, and measurement state questions with  $104 \pm 43$  to  $56 \pm 21$  s than when the group was not presented with a visualization. Group vis-math scored higher when presented without visualization on the global phase, phase gate, H gate, and measurement probability question with scores of  $0.66 \pm 0.18$  to  $0.5 \pm 0.23$ . Group vis-math took longer on the measurement probability and measurement state questions when presented with visualization with  $57 \pm 23$  to  $37 \pm 15$  s and took less time on the global phase, phase- gate, and H-gate questions with  $118 \pm 43$  to  $63 \pm 20$  s.

**a. Representational competence**

Participants were asked to calculate Euler formula identities and similar questions to translate between CN to a complex number. In addition, they were asked to multiply numbers written in DN and CN as shown in Fig. 7. Then they were asked to rate their confidence in the question from 0 (random guess), 1 (very unsure) to 4 (very sure). The correct answers were weighted (multiplied by  $x/4$ , where  $x$  is the confidence). The weighted (unweighted) average symbolic representational understanding was  $0.76 \pm 0.25$  ( $0.87 \pm 0.23$ ). The weighted (unweighted) average CN representational understanding was  $0.47 \pm 0.34$  ( $0.69 \pm 0.32$ ). The average weighted

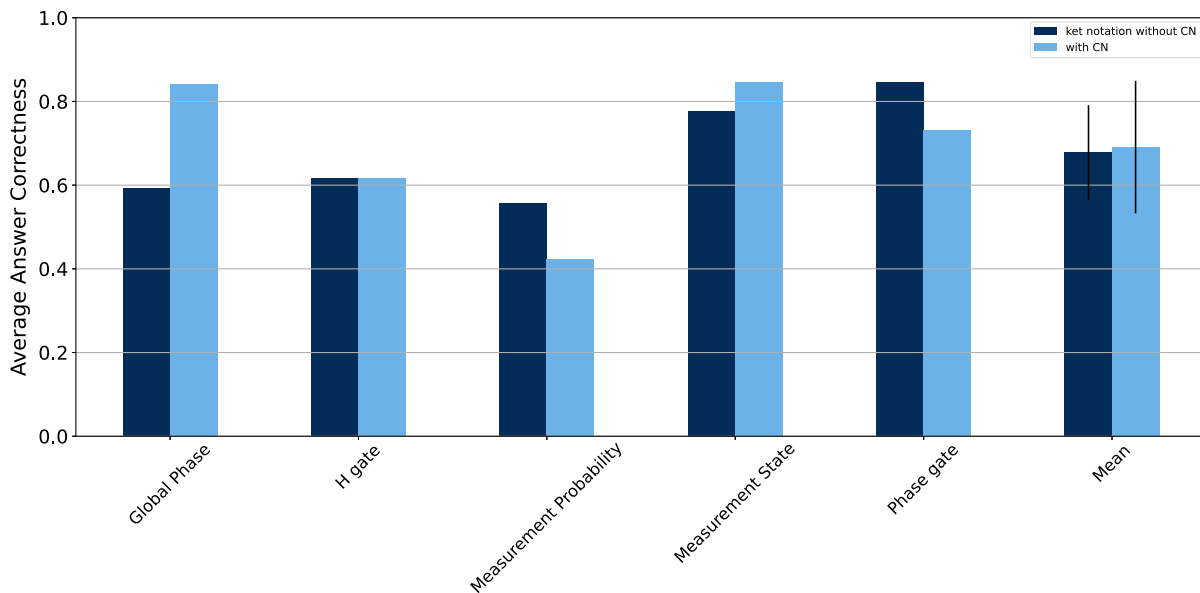


FIG. 15. Average score for each question in the one-qubit survey, for questions were participants were presented with circle notation (CN) (dark blue) and questions were the participants were not presented with CN (light blue). ( $N = 23$ ).

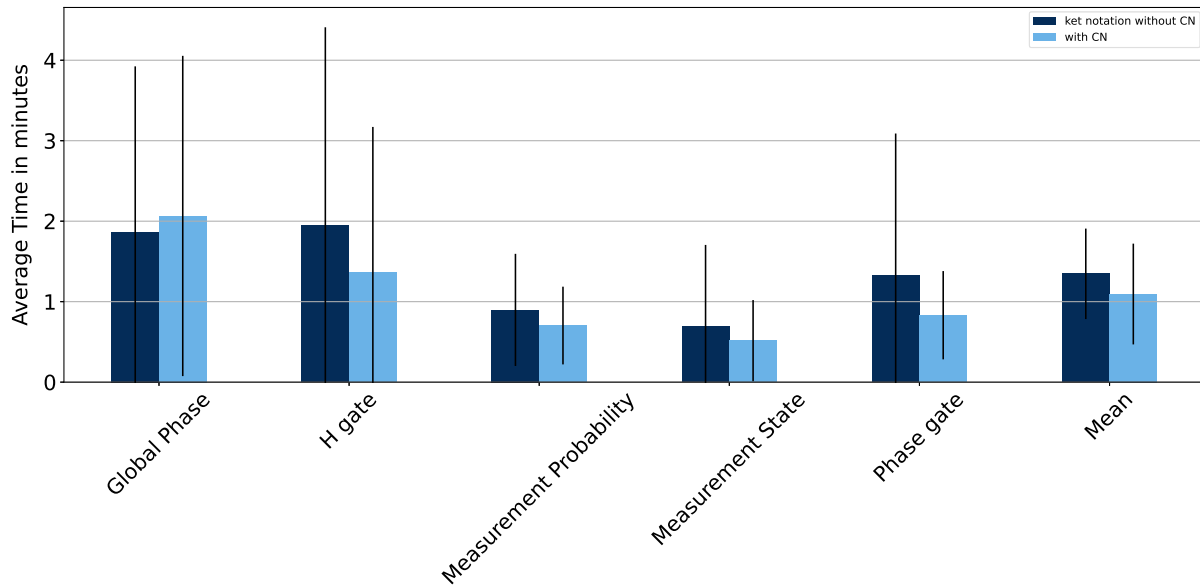


FIG. 16. Average time taken for correct answers to each question in minutes in the one-qubit survey, for questions were participants were presented with circle notation (CN) (dark blue) and questions were the participants were not presented with CN (light blue). ( $N = 18$ ).

(unweighted) difference of CN representational understanding minus mathematics representational understanding was  $-0.29 \pm 0.23$  ( $-0.18 \pm 0.19$ ).

Figure 17 shows, for each participant in the first survey, the difference in scores (with visualization minus without visualization) and Fig. 18 the difference in time, both plotted against the difference in representational understanding between visualization and math. Figure 19 shows the dependency of the difference in intrinsic, germane, and ECL on the representational understanding difference.

**b. Connectional understanding and cognitive load**

On average connectional understanding, 12 students achieved perfect scores and 5 students achieved a score of 0.5. The students who did not score perfectly also scored lower in the rest of the survey ( $0.5 \pm 0.21$  to  $0.75 \pm 0.21$  without visualization and  $0.63 \pm 0.19$  to  $0.69 \pm 0.25$  with visualization) and took a smaller amount of time on the correctly solved questions ( $82 \pm 74$  to  $114 \pm 71$  s without visualization and  $86 \pm 57$  to  $89 \pm 55$  s with visualization) than the students who solved the connectional understanding questions perfectly.

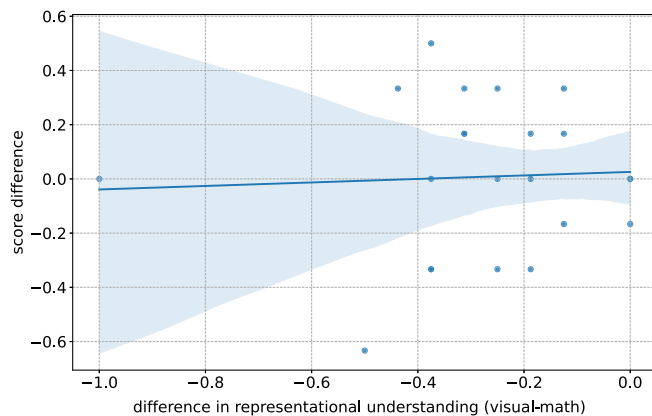


FIG. 17. Difference of average score (with CN minus without CN) plotted against difference of representational understanding (with CN minus without CN) in the one-qubit survey. The linear regression has slope  $b = 0.04$ , intercept  $a = 0.01$ , Pearson  $r = 0.03$  at  $p = 0.91$ , standard error  $\sigma = 0.32$ . The shaded areas represents a 95% confidence interval.

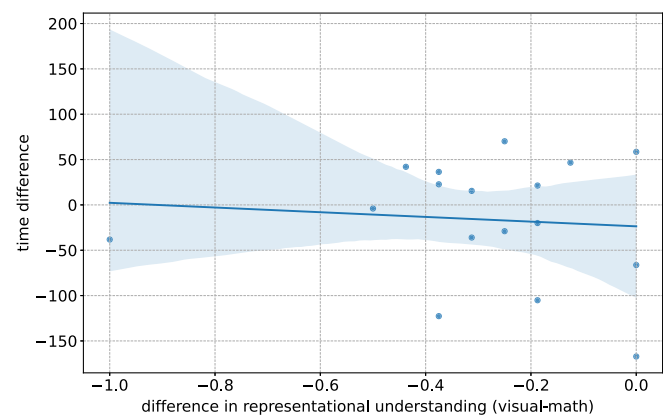


FIG. 18. Average time taken for correct answers (with CN minus symbolic notation) plotted against difference of representational understanding (with CN minus without CN) in the one-qubit survey. The linear regression has slope  $b = -26$ , intercept  $a = -24$ , Pearson  $r = -0.09$  at  $p = 0.73$  and standard error  $\sigma = 73$ . The shaded areas represents a 95% confidence interval.

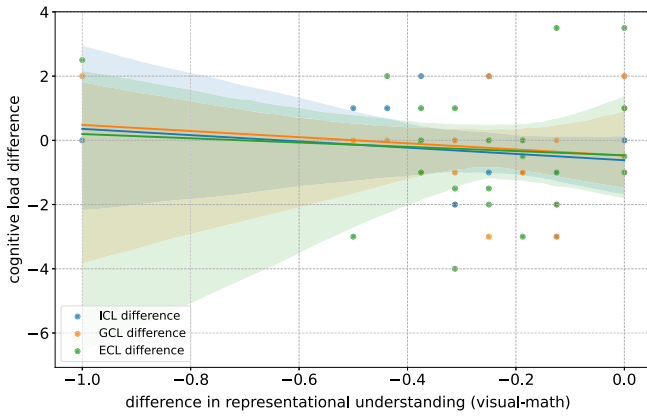


FIG. 19. Perceived intrinsic, germane and ECL difference (with CN minus without CN) and their dependency on the representational understanding difference (again, with CN minus symbolic notation) in the one-qubit survey. Shown are also the linear regressions between the three variables and the representational understanding difference with slope  $b$  and intercept  $a$ . Pearson's  $r$ ,  $p$  value, and standard error  $\sigma$  are as follows. ICL:  $b = -0.68$ ,  $a = -0.27$ ,  $r = -0.16$ ,  $p = 0.59$ ,  $\sigma = 1.2$  GCL:  $b = -0.83$ ,  $a = -0.36$ ,  $r = -0.16$ ,  $p = 0.56$ ,  $\sigma = 1.4$  ECL:  $b = -0.87$ ,  $a = -0.37$ ,  $r = -0.09$ ,  $p = 0.73$ ,  $\sigma = 2.5$ . The shaded areas represent 95% confidence intervals.

Average ECL ( $2.88 \pm 1.67$  with visualization and  $2.76 \pm 1.54$  without) was lower than the average ICL ( $3.75 \pm 0.97$  with visualization and  $3.64 \pm 1.04$  without) and GCL ( $3.94 \pm 1.26$  with visualization and  $4.13 \pm 1.41$  without).

## 2. Two-qubit survey

Table XXV lists the average scores and Table XXVI lists the average time taken for correctly solved one-qubit questions, with and without visualization, for surveys that were taken supported by CN and DCN. Figure 20 shows the combined average score per question and Fig. 21 shows the combined average time taken (for all answers, not just correct ones). Cases where a participant took more than 20 min for one question were excluded.

### a. Circle notation

During part A, participants assigned to the CN group solved the questions without visualization with an average score of  $0.72 \pm 0.14$  with an average time of  $1.54 \pm 0.62$  min for correctly solved questions, while the other group answered the same questions with visualization with an average score of  $0.65 \pm 0.26$  with an average time of  $2.18 \pm 1.02$  min. During part B, the mean scores with and without visualization were  $0.55 \pm 0.17$  and  $0.62 \pm 0.12$ , respectively, at average times for correctly solved questions of  $1.46 \pm 0.67$  and  $1.68 \pm 1.12$  min. When combining parts A and B of the survey, the average scores were  $0.64 \pm 0.13$  without CN and  $0.64 \pm 0.12$  with CN, at  $1.48 \pm 0.51$  and  $1.98 \pm 1.05$  min.

While group math-vis, that was assigned questions without CN first, scored higher than group vis-math on the X-gate and H-gate questions in survey part A with, on average,  $0.9 \pm 0.1$  to  $0.3 \pm 0.1$ , the inverse is apparent for the separability and entanglement questions, where group vis-math scored higher than group math-vis  $1.0$  to  $0.7 \pm 0.1$ . Group vis-math, with CN, took longer to correctly solve the X-gate, H-gate, measurement probability, separability, entanglement, and CNOT gate questions with  $132 \pm 61$  to  $86 \pm 32$  s and took less time on the guess-the-gates questions with 140 to 125 s.

In survey part B, group math-vis (that was now given questions with CN) scored higher than group vis-math on the X-gate, H-gate, measurement probability, and guess-the-gates questions with respective average scores  $0.65 \pm 0.09$  to  $0.45 \pm 0.09$  and group vis-math scored higher than group math-vis on the separability, entanglement, and CNOT gate questions with  $0.73 \pm 0.09$  to  $0.53 \pm 0.09$ . In this survey part, group math-vis, with CN, took longer on the measurement probability, entanglement, and CNOT gate questions with, on average,  $120 \pm 23$  to  $65 \pm 9$  s while taking less time on the X-gate, H-gate, separability, and guess-the-gates questions with  $102 \pm 38$  to  $65 \pm 22$  ss on these questions on average.

Comparing within each group, group math-vis scored worse with CN than without CN on the X-gate, H-gate, entanglement, and CNOT gate questions with  $0.85 \pm 0.09$  to  $0.6 \pm 0.14$ . When solving questions correctly, group math-vis took longer on the X-gate, H-gate, measurement probabilities, and CNOT gate questions when presented with CN, with  $102 \pm 38$  to  $65 \pm 22$  s, while taking a shorter amount of time on the separability, entanglement, and Guess-the-Gates questions with  $120 \pm 23$  to  $65 \pm 9$  s.

When presented with CN, group vis-math scored higher than when not presented with a visualization on the measurement probabilities, separability, entanglement, and guess-the-gates questions with, on average, with  $0.8 \pm 0.2$  to  $0.55 \pm 0.17$ . The group scored higher when solving questions without CN on the X-gate and H-gate questions with  $0.5 \pm 0.1$  to  $0.3 \pm 0.1$ . They took longer with CN when solving the measurement probabilities, separability, entanglement, and CNOT gate questions correctly, with an average of  $155 \pm 70$  to  $65 \pm 11$  s, and took less time on the X-gate, H-gate, and guess-the-gates questions when presented with CN, with  $134 \pm 8$  to  $112 \pm 22$  s.

### b. Dimensional circle notation

During part A, participants assigned to the DCN group solved questions without visualization with an average score of  $0.57 \pm 0.23$  with an average time of  $2.39 \pm 0.64$  min for correctly solved questions, while the same questions with visualization were answered with an average score of  $0.63 \pm 0.25$  with an average time of

TABLE XXV. Average scores for two-qubit questions, with and without visualization ( $N = 27$ ). Note: Standard deviation is used for confidence intervals.

Survey part:	Circle notation					
	A		B		Combined	
	Without visualization <sup>b</sup>	With vis. <sup>b</sup>	Without visualization <sup>b</sup>	With visualization <sup>b</sup>	Without visualization <sup>a</sup>	With visualization <sup>a</sup>
Question						
X gate	0.86	0.5	0.67	0.86	0.77	0.69
H gate	0.86	0.17	0.33	0.57	0.62	0.38
Measurement probability	0.57	0.5	0.33	0.57	0.46	0.54
Separability	0.57	1.0	0.83	0.57	0.69	0.77
Entanglement	0.57	0.83	0.5	0.29	0.54	0.54
Guess-the-gates	0.5	0.5	0.33	0.5	0.42	0.5
Mean	$0.65 \pm 0.15$	$0.58 \pm 0.27$	$0.5 \pm 0.19$	$0.56 \pm 0.17$	$0.58 \pm 0.12$	$0.57 \pm 0.13$
Survey part:	Dimensional circle notation					
	A		B		Combined	
	Without visualization <sup>d</sup>	With visualization <sup>c</sup>	Without visualization <sup>c</sup>	With visualization <sup>d</sup>	Without visualization <sup>c</sup>	With visualization <sup>c</sup>
Question						
X gate	1.0	0.88	0.88	0.71	0.93	0.8
H gate	0.14	0.62	0.5	0.14	0.33	0.4
Measurement probability	0.43	0.5	0.62	0.29	0.53	0.4
Separability	0.57	0.62	0.75	0.71	0.67	0.67
Entanglement	0.43	0.88	0.75	0.43	0.6	0.67
Guess-the-gates	0.71	0.2	0.0	0.17	0.42	0.18
Mean	$0.55 \pm 0.27$	$0.62 \pm 0.23$	$0.58 \pm 0.29$	$0.41 \pm 0.24$	$0.58 \pm 0.19$	$0.52 \pm 0.21$

<sup>a</sup> $N = 12$ .

<sup>b</sup> $N = 6$ .

<sup>c</sup> $N = 7$ .

<sup>d</sup> $N = 8$ .

<sup>e</sup> $N = 15$ .

$2.68 \pm 1.65$  min. During part B, the average scores were  $0.54 \pm 0.26$  and  $0.45 \pm 0.26$ , respectively, at average times for correctly solved questions of  $2.44 \pm 1.87$  and  $3.71 \pm 2.27$  min. When combining the survey parts A and B, the average scores were  $0.56 \pm 0.2$  without visualization and  $0.53 \pm 0.2$  with visualization, at  $2.95 \pm 2.44$  and  $4.64 \pm 6.26$  min.

In both survey parts, on the X-gate, separability, and guess-the-gates questions, group math-vis scored higher than group vis-math. In part A, this was  $0.72 \pm 0.21$  to  $0.47 \pm 0.25$  and in part B, with  $0.61 \pm 0.31$  to  $0.47 \pm 0.34$ . On the other hand, group vis-math scored higher than group math-vis on the H gate, Measurement probability, Entanglement, and CNOT gate questions  $0.75 \pm 0.17$  to  $0.46 \pm 0.18$  in part A and  $0.6 \pm 0.14$  to  $0.33 \pm 0.12$  in part B.

In survey part A, group math-vis took longer on the H-gate, measurement probabilities, entanglement, CNOT gate, and guess-the-gates questions with  $158 \pm 20$  to  $99 \pm 47$  s and less time on the X-gate and separability questions with  $145 \pm 39$  to  $107 \pm 39$  s. In survey part B, group math-vis took longer on the X-gate, measurement probabilities, entanglement, and CNOT gate questions with

$170 \pm 83$  to  $90 \pm 9$  s and group vis-math took longer on the H-gate and separability questions with  $259 \pm 111$  to  $227 \pm 126$  s.

Group math-vis had a higher score on the separability questions with DCN with 0.83 to 0.5 while performing worse with DCN on the X-gate, measurement probabilities, entanglement, CNOT, and guess-the-gates questions with  $0.67 \pm 0.18$  to  $0.43 \pm 0.23$ . The group took longer with DCN when correctly solving the X-gate, H-gate, measurement probabilities, entanglement, and guess-the-gates questions with  $271 \pm 119$  to  $143 \pm 42$  s and took less time on the separability and CNOT questions with  $146 \pm 0$  to  $103 \pm 3$  s.

Supported by DCN, group vis-math scored higher on the H-gate, entanglement, CNOT, and guess-the-gates questions with  $0.65 \pm 0.3$  to  $0.45 \pm 0.3$  while scoring worse on the separability question with 0.6 to 0.4. On the X-gate, measurement probabilities, separability, and CNOT questions, group vis-math took longer with DCN, with  $140 \pm 40$  to  $106 \pm 26$  s, while taking longer without DCN on the H-gate and entanglement questions with  $227 \pm 143$  to  $56 \pm 28$  s.

TABLE XXVI. Average time taken for correct answers to two-qubit questions in minutes ( $N = 22$ ). Note: The number of participants  $N$  that solved the question correctly and are therefore included in the average is shown in brackets. Standard deviations are shown where  $N > 1$ . The confidence interval of the mean is the mean of the standard deviations.

Survey part:	Circle notation <sup>a</sup>					
	A		B		Combined	
	Without visualization <sup>b</sup>	With visualization <sup>c</sup>	Without visualization <sup>c</sup>	With visualization <sup>b</sup>	Without visualization <sup>a</sup>	With visualization <sup>a</sup>
Question						
X gate	0.71 ± 0.16(5)	1.34 ± 0.06(2)	2.05 ± 1.06(3)	0.8 ± 0.35(5)	0.95 ± 0.39(6)	1.08 ± 0.23(6)
H gate	0.93 ± 0.96(5)	2.18(1)	2.28 ± 0.83(2)	2.2 ± 0.64(3)	1.15 ± 0.96(3)	2.2 ± 0.64(3)
Measurement probability	2.38 ± 0.33(3)	4.58 ± 2.83(3)	1.38 ± 0.34(2)	4.12 ± 4.42(3)	2.0 ± 0.67(4)	4.39 ± 3.42(4)
Separability	1.78 ± 1.34(3)	1.87 ± 0.61(5)	0.95 ± 0.94(4)	0.84 ± 0.25(3)	1.3 ± 1.2(7)	1.38 ± 0.7(7)
Entanglement	1.47 ± 0.39(4)	1.73 ± 1.21(5)	1.09 ± 0.33(3)	1.24 ± 0.12(2)	1.36 ± 0.43(5)	1.86 ± 1.13(5)
Guess-the-gates	2.34 ± 2.01(3)	2.08 ± 0.61(3)	2.34 ± 2.29(2)	1.17 ± 1.04(3)	2.41 ± 2.37(4)	1.48 ± 0.91(4)
Mean	1.6 ± 0.64	2.28 ± 1.06	1.57 ± 0.63	1.78 ± 1.17	1.55 ± 0.51	2.07 ± 1.1

Survey part:	Dimensional circle notation <sup>a</sup>					
	A		B		Combined	
	Without visualization <sup>b</sup>	With visualization <sup>c</sup>	Without visualization <sup>c</sup>	With visualization <sup>b</sup>	Without visualization <sup>a</sup>	With visualization <sup>a</sup>
Question						
X gate	1.13 ± 0.38(6)	1.77 ± 0.67(4)	1.35 ± 0.65(4)	1.7 ± 0.66(5)	1.29 ± 0.49(9)	1.73 ± 0.67(9)
H gate	2.73(1)	6.12 ± 7.98(3)	6.17 ± 1.65(2)	5.88(1)	7.82(1)	17.4(1)
Measurement probability	3.22 ± 1.7(3)	2.91 ± 1.62(3)	1.71 ± 0.26(3)	5.13 ± 3.27(2)	1.86 ± 0.34(4)	2.65 ± 1.48(4)
Separability	2.43 ± 1.85(3)	3.08 ± 1.99(2)	2.48 ± 2.22(3)	1.68 ± 0.9(5)	2.83 ± 2.03(5)	2.14 ± 1.51(5)
Entanglement	2.27 ± 0.49(3)	1.39 ± 1.4(4)	1.41 ± 0.95(3)	2.77 ± 2.48(2)	1.81 ± 0.95(5)	2.09 ± 2.04(5)
Guess-the-gates	2.53 ± 0.68(4)	1.9(1)	(0)	7.07(1)	(0)	(0)
Mean	2.39 ± 0.64	2.68 ± 1.65	2.44 ± 1.87	3.71 ± 2.27	2.95 ± 2.44	4.64 ± 6.26

<sup>a</sup> $N = 11$ .

<sup>b</sup> $N = 6$ .

<sup>c</sup> $N = 5$ .

### 3. Three-qubit survey

Table XXVII lists the average scores and Table XXVIII lists the average time taken for correctly solved one-qubit questions, with and without visualization, for surveys that were taken supported by CN and DCN. Figure 22 shows the combined average score per question and Fig. 23 shows the combined average time taken (for all answers, not just the correct ones). Cases in which a participant took more than 20 min for one question were excluded.

#### a. Circle notation

During part A, participants assigned to the CN group solved questions without visualization with an average score of  $0.58 \pm 0.26$  and an average time of  $1.77 \pm 0.94$  min for correctly solved questions, while the same questions with visualization were answered with an average score of  $0.76 \pm 0.28$  and an average time of  $3.0 \pm 1.63$  min. During part B, the average scores were  $0.67 \pm 0.31$  and  $0.6 \pm 0.25$ , respectively, at average times for correctly solved questions of  $2.43 \pm 1.26$  and

$1.25 \pm 0.45$  min. When combining parts A and B of the survey, the average scores were  $0.62 \pm 0.28$  without visualization and  $0.68 \pm 0.25$  with visualization, at  $2.38 \pm 1.33$  and  $2.37 \pm 1.51$  min.

In survey part A, group math-vis, that was presented questions without CN first, scored higher than group vis-math on the entanglement question with 0.6 to 0.2. Group vis-math scored higher on the Z gate, the H gate, the Measurement probability, the CNOT, and the guess-the-gates questions with  $0.92 \pm 0.1$  to  $0.52 \pm 0.27$ . In this survey part, group math-vis took longer when solving the entanglement question correctly, with 152 to 126 s, while group vis-math took longer for correct solutions to the X- gate, Z-gate, H-gate, measurement probabilities, measurement state, separability, and CNOT gate questions, with  $183 \pm 101$  to  $100 \pm 53$  s.

In survey part B, group math-vis scored higher on the H gate, entanglement, and guess-the-gates questions with  $0.6 \pm 0.28$  to  $0.33 \pm 0.19$ , while group vis-math scored higher on the Z-gate, measurement probabilities, measurement state, Separability, and CNOT gate questions with

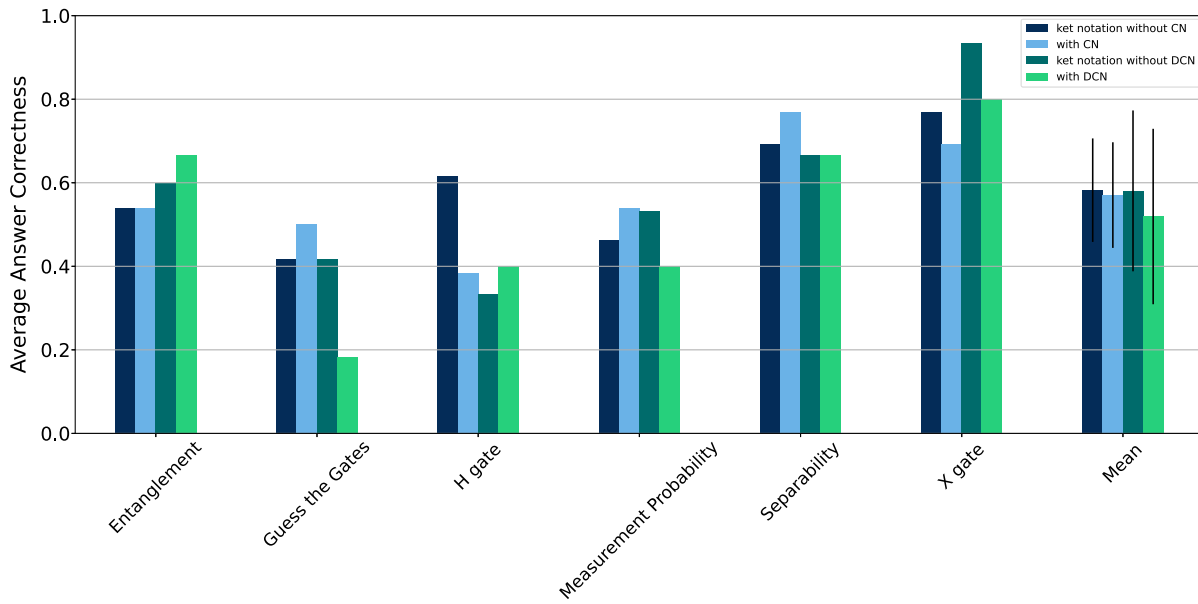


FIG. 20. Average score for each question in the two-qubit survey, separated by participants who were presented with CN (blue,  $N = 12$ ) and participants who were presented with DCN (green,  $N = 15$ ). Darker colors are for questions where participants were not presented with a visualization and lighter colors for questions where participants were presented with a visualization.

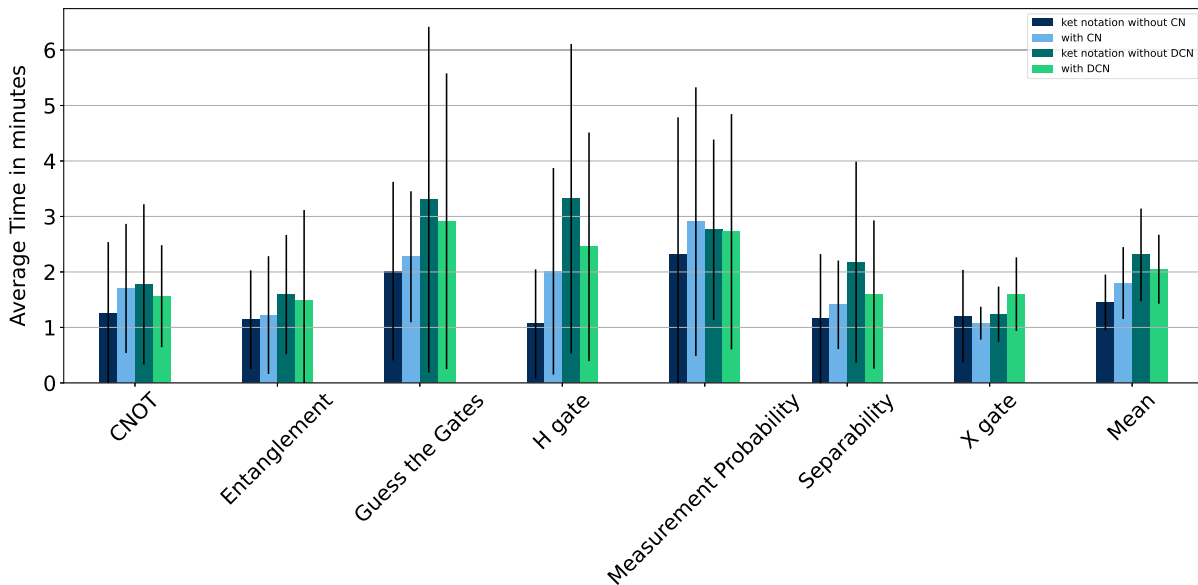


FIG. 21. Average time taken for correct answers to each question in minutes in the two-qubit survey, separated by participants who were presented with CN (blue,  $N = 11$ ) and participants who were presented with DCN (green,  $N = 11$ ). Darker colors are for questions where participants were not presented with a visualization and lighter colors for questions where participants were presented with a visualization.

$0.8 \pm 0.22$  to  $0.52 \pm 0.16$ . Group vis-math took longer on all questions in survey part B with  $146 \pm 71$  to  $75 \pm 26$  s.

Group math-vis, that was presented questions without CN first, scored higher on the H gate and guess-the-gates questions with  $0.7 \pm 0.3$  to  $0.3 \pm 0.3$  with CN. The group scored worse with CN on the separability, entanglement,

and CNOT gate questions by a score difference of  $0.6 \pm 0.16$  to  $0.4 \pm 0.16$ . With CN. Participants in the group took longer when solving the X-gate, Z-gate, and separability questions correctly,  $70 \pm 5$  to  $63 \pm 6$  s, while taking less time on the H gate, measurement probabilities, measurement state, entanglement, CNOT gate, and guess-the-gates questions with  $132 \pm 52$  to  $69 \pm 27$  s.

TABLE XXVII. Average scores for three-qubit questions, with and without visualization ( $N = 17$ ). Note: Standard deviations are used for confidence intervals.

Survey part:	Circle notation <sup>a</sup>					
	A		B		Combined	
	Without visualization <sup>b</sup>	With visualization <sup>b</sup>	Without visualization <sup>b</sup>	With visualization <sup>b</sup>	Without visualization <sup>a</sup>	With visualization <sup>a</sup>
Question						
X gate	1.0	1.0	1.0	1.0	1.0	1.0
Z gate	0.6	1.0	1.0	0.6	0.8	0.8
H gate	0.6	0.8	0.6	1.0	0.6	0.9
Measurement probability	0.6	1.0	1.0	0.6	0.8	0.8
Measurement state	0.6	0.6	0.8	0.6	0.7	0.6
Separability	0.4	0.4	0.4	0.2	0.4	0.3
Entanglement	0.6	0.2	0.2	0.4	0.4	0.3
CNOT gate	0.8	1.0	0.8	0.6	0.8	0.8
Guess-the-gates	0.0	0.8	0.2	0.4	0.1	0.6
Mean	$0.58 \pm 0.26$	$0.76 \pm 0.28$	$0.67 \pm 0.31$	$0.6 \pm 0.25$	$0.62 \pm 0.28$	$0.68 \pm 0.25$

Survey part:	Dimensional circle notation <sup>b</sup>					
	A		B		Combined	
	Without visualization <sup>c</sup>	With visualization <sup>d</sup>	Without visualization <sup>d</sup>	With visualization <sup>c</sup>	Without visualization <sup>b</sup>	With visualization <sup>b</sup>
Question						
X gate	1.0	0.75	0.75	1.0	0.86	0.86
Z gate	1.0	0.75	0.5	0.67	0.71	0.71
H gate	1.0	0.25	0.75	0.67	0.86	0.43
Measurement probability	1.0	0.75	0.75	1.0	0.86	0.86
Measurement state	0.33	1.0	0.5	1.0	0.43	1.0
Separability	1.0	0.5	0.75	0.0	0.86	0.29
Entanglement	0.67	0.75	0.75	1.0	0.71	0.86
CNOT gate	1.0	0.75	0.5	0.67	0.71	0.71
Guess-the-gates	0.33	0.25	0.75	0.33	0.57	0.29
Mean	$0.81 \pm 0.28$	$0.64 \pm 0.24$	$0.67 \pm 0.12$	$0.7 \pm 0.33$	$0.73 \pm 0.14$	$0.67 \pm 0.25$

<sup>a</sup> $N = 10$ .<sup>b</sup> $N = 7$ .<sup>c</sup> $N = 3$ .<sup>d</sup> $N = 4$ .

Group vis-math, that was presented questions with CN first, scored higher on the H gate, CNOT gate, and guess-the-gates questions with CN with  $0.87 \pm 0.09$  to  $0.53 \pm 0.25$ . The group scored lower with CN on the measurement state question with 0.6 as compared to 0.8. When presented with CN, participants in the group took longer on the Z-gate, H-gate, measurement probabilities, measurement state, separability, and CNOT gate questions, with  $198 \pm 101$  to  $130 \pm 49$  s. They took less time on the X gate, entanglement, and guess-the-gates questions, with  $178 \pm 93$  to  $143 \pm 53$  s.

### b. Dimensional circle notation

During part A, participants assigned to the DCN group solved questions without visualization with an average

score of  $0.89 \pm 0.21$  with an average time of  $4.75 \pm 1.77$  min for correctly solved questions, while the same questions with visualization were answered with an average score of  $0.6 \pm 0.24$  with an average time of  $2.32 \pm 0.63$  min. During part B, the average scores were  $0.56 \pm 0.16$  and  $0.72 \pm 0.34$ , respectively, at average times for correctly solved questions of  $2.0 \pm 1.34$  and  $2.29 \pm 1.33$  min. When combining parts A and B of the survey, the average scores were  $0.69 \pm 0.15$  without visualization and  $0.62 \pm 0.29$  with visualization, at  $3.03 \pm 0.78$  and  $2.44 \pm 0.9$  min.

Group math-vis, that was assigned questions without DCN first, scored higher than group vis-math on the X-gate, Z-gate, H-gate, measurement probabilities, separability, entanglement, CNOT gate, and guess-the-gates questions in survey part A, with  $0.94 \pm 0.17$  to  $0.5 \pm 0.24$ . The inverse is

TABLE XXVIII. Average time taken for correct answers to three-qubit questions in minutes ( $N = 15$ ). Note: The number of participants  $N$  that solved the question correctly and are therefore included in the average is shown in brackets. Standard deviations are shown where  $N > 1$ . The confidence interval of the mean is the mean of the standard deviations.

Survey part:	Circle notation <sup>a</sup>					
	A		B		Combined	
	Without visualization <sup>b</sup>	With visualization <sup>b</sup>	Without visualization <sup>b</sup>	With visualization <sup>b</sup>	Without visualization <sup>a</sup>	With visualization <sup>a</sup>
Question						
X gate	0.94 ± 0.36(5)	1.46 ± 1.16(5)	1.52 ± 0.82(5)	1.05 ± 0.34(5)	1.23 ± 0.7(10)	1.26 ± 0.88(10)
Z gate	1.03 ± 0.04(3)	2.02 ± 1.18(5)	1.53 ± 0.87(5)	1.17 ± 0.29(3)	1.35 ± 0.73(8)	1.7 ± 1.04(8)
H gate	3.72 ± 1.74(3)	6.82 ± 5.99(4)	3.64 ± 1.59(3)	1.85 ± 0.53(5)	3.68 ± 1.67(6)	4.79 ± 5.26(6)
Measurement probabilities	1.42 ± 0.66(3)	2.7 ± 1.17(5)	1.27 ± 0.87(5)	0.44 ± 0.2(3)	1.32 ± 0.8(8)	1.85 ± 1.43(8)
Measurement state	1.92 ± 0.1(3)	3.53 ± 1.2(3)	2.75 ± 1.71(4)	1.28 ± 0.44(3)	2.14 ± 1.31(6)	2.41 ± 1.44(6)
Separability	1.19 ± 0.36(2)	3.1 ± 1.92(2)	2.18 ± 0.28(2)	1.27(1)	2.01 ± 0.46(2)	1.23 ± 0.04(2)
Entanglement	2.53 ± 2.24(3)	2.1(1)	2.27(1)	1.16 ± 0.14(2)	3.07 ± 1.89(3)	1.47 ± 0.46(3)
CNOT gate	1.41 ± 0.46(4)	1.68 ± 1.72(5)	1.6 ± 0.94(4)	1.06 ± 0.04(3)	1.53 ± 0.8(7)	1.48 ± 1.49(7)
Guess-the-gates	(0)	3.59 ± 1.35(4)	5.13(1)	1.98 ± 0.42(2)	5.13(1)	5.13(1)
Mean	1.77 ± 0.94	3.0 ± 1.63	2.43 ± 1.26	1.25 ± 0.45	2.38 ± 1.33	2.37 ± 1.51

Survey part	Dimensional circle notation <sup>b</sup>					
	A		B		Combined	
	Without visualization <sup>c</sup>	With visualization <sup>d</sup>	Without visualization <sup>d</sup>	With visualization <sup>c</sup>	Without visualization <sup>b</sup>	With visualization <sup>b</sup>
Question						
X gate	1.83 ± 0.42(2)	3.11 ± 2.38(2)	1.49 ± 0.43(2)	1.94 ± 0.69(2)	1.66 ± 0.45(4)	2.52 ± 1.84(4)
Z gate	1.23 ± 0.55(2)	0.97 ± 0.1(2)	1.95(1)	1.88 ± 1.25(2)	1.47 ± 0.56(3)	1.61 ± 1.09(3)
H gate	8.02 ± 3.47(2)	(0)	1.79 ± 0.51(2)	1.32(1)	4.55(1)	1.32(1)
Measurement Probabilities	2.43 ± 1.0(2)	1.12 ± 0.04(2)	0.52 ± 0.09(2)	2.35 ± 1.9(2)	1.48 ± 1.19(4)	1.74 ± 1.48(4)
Measurement state	1.05(1)	1.08 ± 0.2(3)	1.77(1)	2.0 ± 1.4(2)	1.41 ± 0.36(2)	0.83 ± 0.23(2)
Separability	3.66 ± 2.66(2)	1.57(1)	3.51 ± 3.28(2)	(0)	6.78(1)	1.57(1)
Entanglement	6.12 ± 4.09(2)	1.88 ± 0.52(2)	2.1 ± 1.93(2)	2.76 ± 1.43(2)	4.11 ± 3.78(4)	2.32 ± 1.16(4)
CNOT gate	0.89 ± 0.18(2)	1.55 ± 0.62(2)	2.68(1)	4.52(1)	1.88 ± 0.81(2)	2.72 ± 1.79(2)
Guess-the-gates	17.48(1)	7.28(1)	2.18 ± 1.78(2)	1.52(1)	3.97(1)	7.28(1)
Mean	4.75 ± 5.37	2.32 ± 2.12	2.0 ± 0.82	2.29 ± 1.01	3.03 ± 1.91	2.44 ± 1.92

<sup>a</sup> $N = 10$ .

<sup>b</sup> $N = 5$ .

<sup>c</sup> $N = 2$ .

<sup>d</sup> $N = 3$ .

apparent for the measurement state question, where group vis-math scored higher than group math-vis with 1 to 0.5. Group vis-math, with DCN, took longer to correctly solve the Z-gate, measurement probabilities, separability, entanglement, and guess-the-gates questions, with 371 ± 353 to 154 ± 143 s, whereas group math-vis took longer for correctly solving the X-gate, measurement state, and CNOT gate questions, 115 ± 52 to 76 ± 25 s.

In survey part B, group math-vis (that now was presented questions with DCN) scored higher than group vis-math on the X-gate, Z-gate, measurement probabilities, measurement state, entanglement, and CNOT gate questions with 0.92 ± 0.19 to 0.5 ± 0.17 and group vis-math scored higher than group math-vis on the H-gate, separability, and guess-the-gates questions with 0.67 ± 0.0 to

0.33 ± 0.24. In this survey part, group math-vis, with DCN, took longer on the X-gate, measurement probabilities, measurement state, entanglement, and CNOT gate questions, with 163 ± 57 to 103 ± 43 s, while taking less time on the Z-gate, H-gate, and guess-the-gates questions, with 118 ± 10 to 94 ± 14 s.

Comparing within each group, group math-vis scored higher with DCN on the measurement state question with 1 to 0.5, and scored lower with DCN than without DCN on the H-gate, separability, and CNOT gate questions by averages of 1.0 ± 0.0 to 0.33 ± 0.24. When solving questions correctly, group math-vis took longer on the X-gate, Z-gate, measurement state, and CNOT gate questions when presented with DCN, with 155 ± 67 to 75 ± 21 s, while taking less time on the H-gate, measurement probabilities,

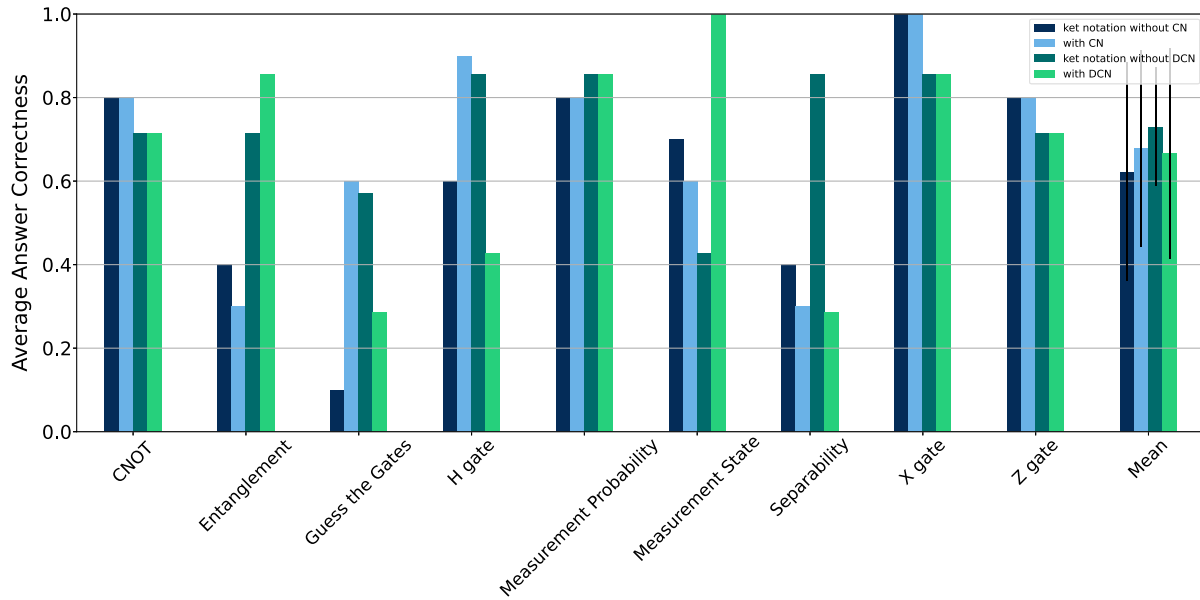


FIG. 22. Average score for each question in the three-qubit survey, separated by participants who were presented with CN (blue,  $N = 10$ ) and participants who were presented with DCN (green,  $N = 7$ ). Darker colors are for questions where participants were not presented with a visualization and lighter colors for questions where participants were presented with a visualization.

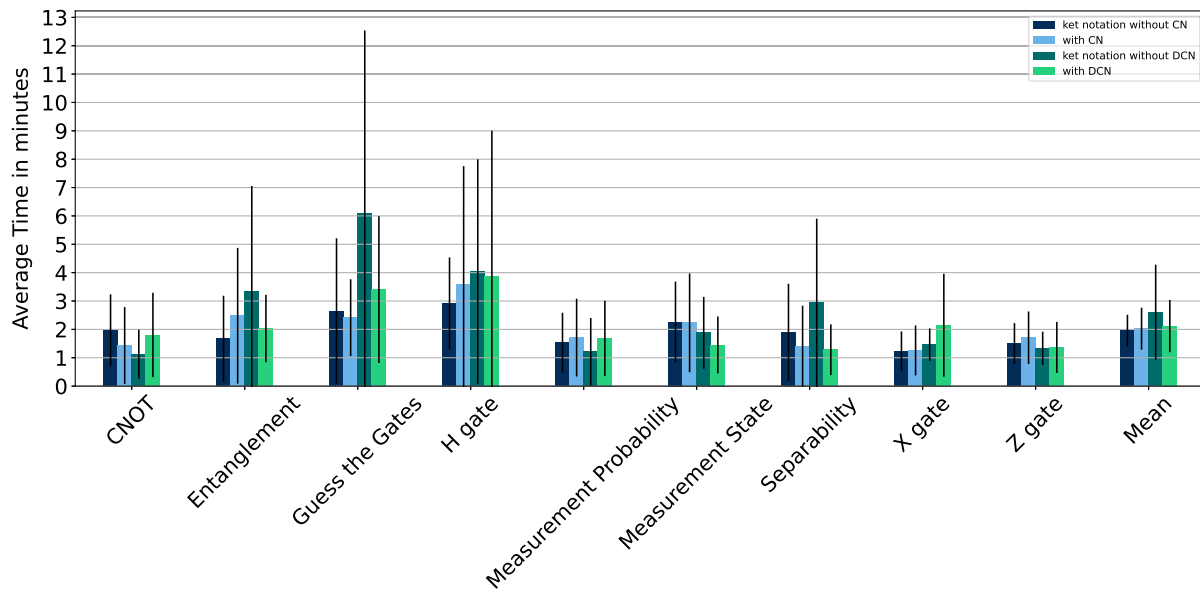


FIG. 23. Average time taken for each question in minutes in the three-qubit survey, separated by participants who were presented by CN (blue,  $N = 10$ ) and participants who were presented with DCN (green,  $N = 5$ ). Darker colors are for questions where participants were not presented with a visualization and lighter colors for questions where participants were presented with a visualization.

entanglement, and guess-the-gates questions,  $511 \pm 333$  to  $119 \pm 35$  s.

When presented with DCN, group vis-math scored higher than itself on the Z-gate, measurement state, and CNOT gate questions by, on average,  $0.78 \pm 0.16$  to  $0.33 \pm 0.0$ . The group scored higher when solving questions without DCN on the H gate, Separability, and

guess-the-gates questions with  $0.67 \pm 0.0$  to  $0.22 \pm 0.16$ . They took longer with DCN when solving the X-gate, measurement probabilities, and guess-the-gates questions correctly,  $230 \pm 154$  to  $84 \pm 41$  s, and took less time on the Z gate, measurement state, separability, entanglement, and CNOT gate questions when presented with DCN, on average  $144 \pm 38$  to  $85 \pm 20$  s.

- [1] J. C. Meyer, G. Passante, S. J. Pollock, and B. R. Wilcox, Today's interdisciplinary quantum information classroom: Themes from a survey of quantum information science instructors, *Phys. Rev. Phys. Educ. Res.* **18**, 010150 (2022).
- [2] *Visualization in Science Education*, edited by J. K. Gilbert (Springer Netherlands, Dordrecht, 2005).
- [3] M. Evagorou, S. Erduran, and T. Mäntylä, The role of visual representations in scientific practices: From conceptual understanding and knowledge generation to “seeing” how science works, *Int. J. STEM Educ.* **2**, 11 (2015).
- [4] M. Kim and Q. Jin, Studies on visualisation in science classrooms: A systematic literature review, *Int. J. Sci. Educ.* **44**, 2613 (2022).
- [5] M. A. Rau, Conditions for the effectiveness of multiple visual representations in enhancing stem learning, *Educ. Psychol. Rev.* **29**, 717 (2017).
- [6] M. Opfermann, A. Schmeck, and H. E. Fischer, Multiple representations in physics and science education—Why should we use them?, in *Multiple Representations in Physics Education* (2017), pp. 1–22, [10.1007/978-3-319-58914-5](https://doi.org/10.1007/978-3-319-58914-5).
- [7] N. Munfaridah, L. Avraamidou, and M. Goedhart, The use of multiple representations in undergraduate physics education: What do we know and where do we go from here?, *Eurasia J. Math. Sci. Technol. Educ.* **17**, em1934 (2021).
- [8] J. Liu and D. Franklin, Introduction to quantum computing for everyone: Experience report, in *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1, SIGCSE 2023* (Association for Computing Machinery, New York, NY, 2023), pp. 1157–1163, [10.1145/3545945.3569836](https://doi.org/10.1145/3545945.3569836).
- [9] S. Goorney, J. Bley, S. Heusler, and J. Sherson, A framework for curriculum transformation in quantum information science and technology education, *Eur. J. Phys.* **45**, 065702 (2024).
- [10] L. Schewior and M. A. Lindner, Revisiting picture functions in multimedia testing: A systematic narrative review and taxonomy extension, *Educ. Psychol. Rev.* **36**, 49 (2024).
- [11] J. Sweller, J. J. G. van Merriënboer, and F. G. W. C. Paas, Cognitive architecture and instructional design, *Educ. Psychol. Rev.* **10**, 251 (1998).
- [12] K. L. Daniel, C. J. Bucklin, E. Austin Leone, and J. Idema, Towards a definition of representational competence, in *Towards a Framework for Representational Competence in Science Education* (Springer International Publishing, Cham, 2018), pp. 3–11, [10.1007/978-3-319-89945-9\\_1](https://doi.org/10.1007/978-3-319-89945-9_1).
- [13] S. Ainsworth, The multiple representations principle in multimedia learning, in *The Cambridge Handbook of Multimedia Learning*, Cambridge Handbooks in Psychology, edited by R. E. Mayer and L. Fiorella (Cambridge University Press, Cambridge, England, 2021), pp. 158–170.
- [14] L. Hu, G. Chen, P. Li, and J. Huang, Multimedia effect in problem solving: A meta-analysis, *Educ. Psychol. Rev.* **33**, 1717 (2021).
- [15] R. E. Mayer, Cognitive theory of multimedia learning, in *The Cambridge Handbook of Multimedia Learning*, Cambridge Handbooks in Psychology, edited by R. E. Mayer and L. Fiorella (Cambridge University Press, Cambridge, England, 2021), pp. 57–72.
- [16] W. Schnotz, Integrated model of text and picture comprehension, in *Cambridge Handbook of Multimedia Learning*, Cambridge Handbooks in Psychology, edited by R. E. Mayer and L. Fiorella (Cambridge University Press, Cambridge, England 2021), pp. 82–99.
- [17] S. Ainsworth, DeFT: A conceptual framework for considering learning with multiple representations, *Learn. Instr.* **16**, 183 (2006).
- [18] J. L. Cooper, P. G. Sidney, and M. W. Alibali, Who benefits from diagrams and illustrations in math problems? Ability and attitudes matter, *Appl. Cogn. Psychol.* **32**, 24 (2018).
- [19] T. Fredlund, C. Linder, J. Airey, and A. Linder, Unpacking physics representations: Towards an appreciation of disciplinary affordance, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020129 (2014).
- [20] M. Wawro, K. Watson, and W. Christensen, Students' metarepresentational competence with matrix notation and Dirac notation in quantum mechanics, *Phys. Rev. Phys. Educ. Res.* **16**, 020112 (2020).
- [21] S. Seegerer, T. Michaeli, and R. Romeike, Quantum computing as a topic in computer science education, in *Proceedings of the 16th Workshop in Primary and Secondary Computing Education, WiPSCE '21* (Association for Computing Machinery, New York, NY, 2021), [10.1145/3481312.3481348](https://doi.org/10.1145/3481312.3481348).
- [22] A. Kohnle, C. Baily, and S. Ruby, Investigating the influence of visualization on student understanding of quantum superposition, *presented at PER Conf. 2014, Minneapolis, MN*, pp. 139–142, [10.1119/perc.2014.pr.031](https://doi.org/10.1119/perc.2014.pr.031).
- [23] A. Kohnle, C. Baily, A. Campbell, N. Korolkova, and M. J. Paetkau, Enhancing student learning of two-level quantum systems with interactive simulations, *Am. J. Phys.* **83**, 560 (2015).
- [24] E. Marshman and C. Singh, Interactive tutorial to improve student understanding of single photon experiments involving a Mach–Zehnder interferometer, *Eur. J. Phys.* **37**, 024001 (2016).
- [25] P. Hu, Y. Li, R. S. K. Mong, and C. Singh, Student understanding of the Bloch sphere, *Eur. J. Phys.* **45**, 025705 (2024).
- [26] F. Hennig, K. Tóth, M. Förster, and P. Bitzenbauer, A new teaching-learning sequence to promote secondary school students' learning of quantum physics using Dirac notation, *Phys. Educ.* **59**, 045007 (2024).
- [27] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, England, 2006).
- [28] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, England, 2018).
- [29] D. C. Brody and L. P. Hughston, Geometric quantum mechanics, *J. Geom. Phys.* **38**, 19 (2001).
- [30] K. G. H. Vollbrecht and R. F. Werner, Entanglement measures under symmetry, *Phys. Rev. A* **64**, 062307 (2001).
- [31] S. Heusler, P. Schlummer, and M. S. Ubben, The topological origin of quantum randomness, *Symmetry* **13**, 581 (2021).
- [32] H. Mäkelä and A. Messina, N-qubit states as points on the Bloch sphere, *Phys. Scr.* **2010**, 014054 (2010).

- [33] M. Boyer, R. Liss, and T. Mor, Geometry of entanglement in the Bloch sphere, *Phys. Rev. A* **95**, 032308 (2017).
- [34] O. Gamel, Entangled Bloch spheres: Bloch matrix and two-qubit state space, *Phys. Rev. A* **93**, 062320 (2016).
- [35] C. J. D. Kemp, N. R. Cooper, and F. N. Ünal, Nested-sphere description of the  $n$ -level Chern number and the generalized Bloch hypersphere, *Phys. Rev. Res.* **4**, 023120 (2022).
- [36] M. Backens, The ZX-calculus is complete for stabilizer quantum mechanics, *New J. Phys.* **16**, 093021 (2014).
- [37] A. Hadzihasanovic, A diagrammatic axiomatisation for qubit entanglement, in *Proceedings of the 30th Annual ACM/IEEE Symposium on Logic in Computer Science* (2015), pp. 573–584, [10.1109/LICS.2015.59](https://doi.org/10.1109/LICS.2015.59).
- [38] K. F. Ng and Q. Wang, A universal completion of the ZX-calculus, [arXiv:1706.09877](https://arxiv.org/abs/1706.09877).
- [39] M. Backens and A. Kissinger, ZH: A complete graphical calculus for quantum computations involving classical non-linearity, *Electron. Proc. Theor. Comput. Sci.* **287**, 23 (2019).
- [40] B. Coecke, D. Horsman, A. Kissinger, and Q. Wang, Kindergarden quantum mechanics graduates ... or how I learned to stop gluing LEGO together and love the ZX-calculus, *Theor. Comput. Sci.* **897**, 1 (2022).
- [41] H. Bombin and M. A. Martin-Delgado, Topological quantum distillation, *Phys. Rev. Lett.* **97**, 180501 (2006).
- [42] A. Kubica, B. Yoshida, and F. Pastawski, Unfolding the color code, *New J. Phys.* **17**, 083026 (2015).
- [43] Y. Aadel, A. Belhaj, M. Bensed, Z. Benslimane, M. B. Sedra, and A. Segui, Qubit systems from colored toric geometry and hypercube graph theory, *Commun. Theor. Phys.* **68**, 285 (2017).
- [44] M. Vasmer and A. Kubica, Morphing quantum codes, *PRX Quantum* **3**, 030319 (2022).
- [45] A. Garon, R. Zeier, and S. J. Glaser, Visualizing operators of coupled spin systems, *Phys. Rev. A* **91**, 042122 (2015).
- [46] D. Leiner, R. Zeier, and S. J. Glaser, Symmetry-adapted decomposition of tensor operators and the visualization of coupled spin systems, *J. Phys. A.* **53**, 495301 (2020).
- [47] B. Koczor, R. Zeier, and S. J. Glaser, Time evolution of coupled spin systems in a generalized Wigner representation, *Ann. Phys. (Amsterdam)* **408**, 1 (2019).
- [48] D. Huber and S. J. Glaser, Beads: A canonical visualization of quantum states for applications in quantum information processing, *New J. Phys.* **27**, 094509 (2025).
- [49] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, England, 2010).
- [50] L. Nita, L. Mazzoli Smith, N. Chancellor, and H. Cramman, The challenge and opportunities of quantum literacy for future education and transdisciplinary problem-solving, *Res. Sci. Technol. Educ.* **41**, 564 (2023).
- [51] S. Felloni, A. Leporati, and G. Strini, Diagrams of states in quantum information: An illustrative tutorial, [arXiv:0904.2656](https://arxiv.org/abs/0904.2656).
- [52] L. Nita, N. Chancellor, L. M. Smith, H. Cramman, and G. Dost, Inclusive learning for quantum computing: Supporting the aims of quantum literacy using the puzzle game quantum odyssey, [arXiv:2106.07077](https://arxiv.org/abs/2106.07077).
- [53] E. R. Johnston, N. Harrigan, and M. Gimeno-Segovia, *Programming Quantum Computers: Essential Algorithms and Code Samples* (O'Reilly Media, Incorporated, Sebastopol, CA, 2019).
- [54] B. Just, *Quantum Computing Compact*, 2nd ed. (Springer, Berlin, Germany, 2023), first edition released in 2020 in German.
- [55] G. Hellstern, J. Hettel, and B. Just, Introducing quantum information and computation to a broader audience with MOOCS at OpenHPI, *Eur. Phys. J. Quantum Technol.* **11**, 59 (2024).
- [56] J. Bley, E. Rexigel, A. Arias, N. Longen, L. Krupp, M. Kiefer-Emmanouilidis, P. Lukowicz, A. Donhauser, S. Küchemann, J. Kuhn, and A. Widera, Visualizing entanglement in multiqubit systems, *Phys. Rev. Res.* **6**, 023077 (2024).
- [57] M. A. Lindner, A. Eitel, B. Strobel, and O. Köller, Identifying processes underlying the multimedia effect in testing: An eye-movement analysis, *Learn. Instr.* **47**, 91 (2017).
- [58] M. Brang, H. Franke, F. Greinert, M. S. Ubben, F. Hennig, and P. Bitzenbauer, Spooky action at a distance? A two-phase study into learners' views of quantum entanglement, *Eur. Phys. J. Quantum Technol.* **11**, 33 (2024).
- [59] J. Sweller, Cognitive load during problem solving: Effects on learning, *Cognit. Sci.* **12**, 257 (1988).
- [60] S. Kalyuga, P. Ayres, P. Chandler, and J. Sweller, The expertise reversal effect, *Educ. Psychol.* **38**, 23 (2003).
- [61] Y.-c. Yeh, Aptitude-treatment interaction, in *Encyclopedia of the Sciences of Learning* (Springer US, Boston, MA, 2012), pp. 295–298.
- [62] S. M. Downing, Twelve steps for effective test development, in *Handbook of Test Development*, edited by S. M. Downing and T. M. Haladyna (Lawrence Erlbaum Associates, Mahwah, NJ, 2006), pp. 3–25.
- [63] J. L. Little and E. L. Bjork, Optimizing multiple-choice tests as tools for learning, *Mem. Cognit.* **43**, 14 (2015).
- [64] D. Thissen, L. Steinberg, and A. R. Fitzpatrick, Multiple-choice models: The distractors are also part of the item, *J. Educ. Meas.* **26**, 161 (1989).
- [65] D. F. Styer, Common misconceptions regarding quantum mechanics, *Am. J. Phys.* **64**, 31 (1996).
- [66] P. Hu, Y. Li, and C. Singh, Investigating and improving student understanding of the basics of quantum computing, *Phys. Rev. Phys. Educ. Res.* **20**, 020108 (2024).
- [67] J. K. Freericks, Focus on conceptual ideas in quantum mechanics for teacher training, *J. Phys. Conf. Ser.* **2490**, 012004 (2023).
- [68] M. J. Gierl, O. Bulut, Q. Guo, and X. Zhang, Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review, *Rev. Educ. Res.* **87**, 1082 (2017).
- [69] D. Thissen, 9—timed testing: An approach using item response theory, in *New Horizons in Testing* (Academic Press, San Diego, 1983), pp. 179–203, [10.1016/B978-0-12-742780-5.50019-6](https://doi.org/10.1016/B978-0-12-742780-5.50019-6).
- [70] P. J. Ferrando and U. Lorenzo-Seva, An item response theory model for incorporating response time data in binary personality items, *Appl. Psychol. Meas.* **31**, 525 (2007).

- [71] F. Goldhammer, J. Naumann, A. Stelter, K. Tóth, H. Rölke, and E. Klieme, The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment, *J. Educ. Psychol.* **106**, 608 (2014).
- [72] M. Tancoš, E. Chvojka, M. Jabůrek, and Š. Portešová, Faster smarter: Children with higher levels of ability take longer to give incorrect answers, especially when the task matches their ability, *J. Intell.* **11**, 63 (2023).
- [73] B. Hoffman and G. Schraw, Conceptions of efficiency: Applications in learning and problem solving, *Educ. Psychol.* **45**, 1 (2010).
- [74] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/hdpv-frft> for the introductory slides and test items of all surveys.
- [75] M. D. Wolcott and N. G. Lobczowski, Using cognitive interviews and think-aloud protocols to understand thought processes, *Curr. Pharm. Teach. Learn.* **13**, 181 (2021).
- [76] B. C. O. F. Fehringer, Spatial thinking from a different view: Disentangling top-down and bottom-up processes using eye tracking, *Open Psychol.* **2**, 138 (2020).
- [77] R. Lockhart, Low-rank separable states are a set of measure zero within the set of low-rank states, *Phys. Rev. A* **65**, 064304 (2002).
- [78] M. Klepsch, F. Schmitz, and T. Seufert, Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load, *Front. Psychol.* **8**, 1997 (2017).
- [79] K. E. DeLeeuw and R. E. Mayer, A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load, *J. Educ. Psychol.* **100**, 223 (2008).
- [80] A. Raidvee, M. Toom, K. Averin, and J. Allik, Perception of means, sums, and areas, *Attent. Percept. Psychophys.* **82**, 865 (2020).
- [81] F. M. Lord, The relation of the reliability of multiple-choice tests to the distribution of item difficulties, *Psychometrika* **17**, 181 (1952).
- [82] U. of Washington Understanding item analyses, <https://www.washington.edu/assessment/scanning-scoring/scoring/reports/item-analysis/> [accessed 14th February, 2025].
- [83] P. Chandler and J. Sweller, The split-attention effect as a factor in the design of instruction, *Br. J. Educ. Psychol.* **62**, 233 (1992).
- [84] L. Hahn and P. Klein, Eye tracking in physics education research: A systematic literature review, *Phys. Rev. Phys. Educ. Res.* **18**, 013102 (2022).
- [85] H. A. Taylor, H. Burte, and K. T. Renshaw, Connecting spatial thinking to stem learning through visualizations, *Nat. Rev. Psychol.* **2**, 637 (2023).
- [86] J. J. G. van Merriënboer and J. Sweller, Cognitive load theory and complex learning: Recent developments and future directions, *Educ. Psychol. Rev.* **17**, 147 (2005).
- [87] P. Gerjets, K. Scheiter, and G. Cierniak, The scientific value of cognitive load theory: A research agenda based on the structuralist view of theories, *Educ. Psychol. Rev.* **21**, 43 (2009).
- [88] Z. C. Seskir *et al.*, Quantum games and interactive tools for quantum technologies outreach and education, *Opt. Eng. (N.Y.)* **61**, 081809 (2022).
- [89] J. Bley, E. Rexigel, A. Arias, L. Krupp, S. Steinert, N. Longen, P. Lukowicz, S. Küchemann, J. Kuhn, M. Kiefer-Emmanouilidis, and A. Widera, Multiqubit state visualisations to support problem solving—pilot study results, 2024, [10.5281/zenodo.14051040](https://zenodo.org/record/14051040).