Testing quantum reasoning: Developing, validating, and application of a questionnaire

Moritz Waitzmann,* Ruediger Scholz, and Susanne Wessnigk

CRC 1227 Designed Quantum States of Matter, Leibniz University Hannover, Welfengarten 1, 30167 Hannover, Germany

(Received 2 January 2023; revised 13 December 2023; accepted 28 February 2024; published 5 April 2024)

Clear and rigorous quantum reasoning is needed to explain quantum physical phenomena. As pillars of true quantum physical explanations, we suggest specific quantum reasoning derived from quantum physical key ideas. An experiment is suggested to support such a quantum reasoning, in which a quantized radiation field interacts with an optical beam splitter, leading to experimental results conflicting with classical physical predictions. The results, however, can be explained consistently with a quantum reasoning based on the key ideas of probability, superposition, and interference (PSI). In this quantum optical key experiment the optical beam splitter prepares a superposition of single photon states and a Michelson interferometer is used to detect the superposition via controlled propagation phases. Although different single photon experimental setups (aimed at helping students to gain access to foundational issues in quantum physics) have been discussed in the past, the wave-particle dualism bound to classical physics maintains its predominance as an explanation pattern for the interpretation of these experiments. The study presented here investigates the effect of the quantum optical key experiment on the ability of students to use quantum reasoning based on the key ideas of PSI to overcome the naive wave-particle dualism. The current state of relevant studies that test student access to quantum physics can roughly be divided into two distinct areas: one tests how mathematical abilities help them to understand quantum physics and one tests how nonmathematical representations of a set of specific quantum theoretical traits (“Wesenszüge”) lead to a deeper understanding of quantum physics. There is a lack of questionnaires that focus on the idea of developing quantum reasoning based on superposition, probability, and interference of quantum states combined with a real experiment using true quantum light. In the first part of the article, we describe the physical modeling and present the development of the questionnaire. The set of items has been constructed from newly developed items and combined with well-tested ones. The validation of the set addresses qualitative and quantitative methods. In the second part, we give a pre- and poststudy examination of the impact of the quantum optical key experiment on students’ quantum reasoning. A significant increase in the number of students using quantum arguments is based on PSI reasoning for the explanation of an interference, such as the behavior of single photon states. Though the increase is significant, we found only minor changes in a particular issue to the students’ reasoning when approaching quantum physics as illustrated by a sample of answers given in the second part of the article. The concept of quantum states and the principle of superposition still appear particularly difficult.

DOI: 10.1103/PhysRevPhysEducRes.20.010122

I. INTRODUCTION: QUANTUM TEACHING NEEDS CLEAR QUANTUM REASONING

In the same preface to his famous textbook on quantum mechanics “… a book on the new physics (quantum physics; R. S.), if not purely descriptive of experimental work, must be essentially mathematical.” [[1], p. VIII], Dirac points out that the laws of quantum theory do not “govern the world as it appears in our mental picture” (*ibid*.) but need to be interpreted. When these interpretations largely dispense with the formal-mathematical background of quantum theory, a clear and rigorous quantum reasoning-based diction is needed to avoid misleading classical bonds [2,3]. However, in upper secondary school physics curricula, the semiclassical wave-particle dualism is a prevalent basis for explaining phenomena in quantum physics [4], although it is known that from a physical and an educational point of view, semiclassical concepts should be replaced by theoretical constructs much closer to quantum theory [3]. Our suggestion is to get closer to true quantum explanations, by reducing quantum physical argumentation rigorously to quantum physical key ideas: the probabilistic nature of quantum physics, the principle of superposition,

*moritz.waitzmann@dq-mat.unihannover.de

and the quantum interference to detect superposition (PSI) [5–7]. Restricting the framework of explanation to just a few experiment-based key ideas seems suitable for explaining the fundamentals of quantum physics [8,9].

In this article, we define quantum reasoning by using PSI, the quantum physical key ideas of *probability*, *superposition*, and *interference* to interpret and explain quantum phenomena. One goal of this article is to present the process of developing and validating a test instrument to evaluate students' reasoning for the explanation of the quantum optical key experiment based on PSI. Another goal is to evaluate the questionnaires' suitability and to examine the effect of the quantum optical key experiment on students' quantum reasoning. This article is not designed to complement existing approaches to quantum physics [2,10].

II. THEORETICAL FRAMEWORK

A. Shortcomings in particle-wave dualism

Following the OECD Program for International Student Assessment (PISA) 2015, scientific literacy includes the ability to perceive, provide, and evaluate scientific explanations for natural and technical phenomena [11]. For quantum phenomena, this objective is challenging because the well-trained and widely successful real-world explanations and quantum physical rationales are crucially different [2,8,12]. The real physical meaning of quantum theoretical sentences is not self-evident and the interpretation of the state of a physical system is not that obvious in quantum physics and especially not characterized by a shared understanding, as it is in classical physics [13]. Historically, the *Copenhagen Interpretation* of quantum physics invented a particular complementarity to physical variables, as positionlike variables (x) were assigned to a particle scenario and momentumlike variables (p) to a wave scenario [14]. Attributing the wave-particle dualism directly to the quantum objects seemingly compensates for the internal discrepancy of the Copenhagen Interpretation. The behavior of single quantum objects, such as electrons or photons, can be explained by merging theoretical approaches of classical waves and classical particles [15].

The basic criticisms of wave-particle-dualism are as follows:

- Dualistic theories are intrinsically inconsistent. Sometimes they refer to an ontologically unsatisfactory "either-or"-scheme, where the quantum objects change their attributes during the experiment or they fall back to the "as-well-as"-scheme leading to objects completely losing their specific character [16,17].
- The mathematical formalism of quantum physics does not contain any waves or particles. Quantum theory only deals with quantum states [12,17].
- In quantum theory, two noncommutating observables (such as x and p) are complementary and a formal duality occurs if the observables are assigned to classical scenarios ($x \rightarrow$ particle and $p \rightarrow$ wave). However, there is no physical reason to do this (for the uncertainty relation, see Refs. [9,17]).

Wave-particle dualism also seems to be problematic from an educational point of view. Due to the visual evidence of the particle scenario and the wave scenario, students tend to adopt the misleading naive wave-particle duality. Here, quantum objects are seen as "as-well-as"-hybrids of classical waves and classical particles. Depending on the experimental setup, students can switch between two classical scenarios. All classical physical concepts can thus be maintained [3,12,18]. To overcome these problems, a successful, clear, and rigorous nonclassical diction is necessary [3,8].

B. The PSI-guided interpretation of experimental results

Clarity and rigor can be derived from the fundamental principles of quantum physics itself [19,20]. These principles, however, are not in line with classical physical models.

- A state function $|\psi\rangle$ represents the maximum information we have about the physical system. There is no classical physical analog for $|\psi\rangle$.
- The quantum theoretical superposition principle says that with two state functions $|\psi_1\rangle$ and $|\psi_2\rangle$, the linear superposition $|\psi\rangle = c_1 \cdot |\psi_1\rangle + c_2 \cdot |\psi_2\rangle$ is also a state function of the system.
- Quantum theory deals with state functions and operators instead of physical quantities. Born's rule links quantum entities and physical reality, enabling predictions from theory. One formulation of the rule states that the probability $\mathrm{P}(|\psi\rangle, |\psi_1\rangle)$ for a system in state $|\psi\rangle$ to be found in the state $|\psi_1\rangle$ is given by the absolute square of the inner product of the state functions $P(|\psi_1\rangle, |\psi\rangle) = |\langle\psi|\psi_1\rangle|^2$.

The quantum optical key experiment shown in Figs. 1 and 2 combines the central quantum arguments based on PSI. An optical beam splitter prepares a single photon superposition state superposing the two states that represent the classical possibilities of reflection (D_4) and

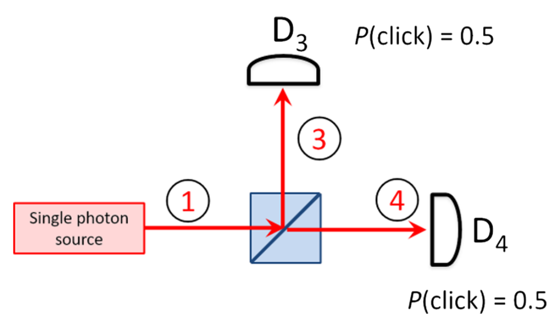

FIG. 1. The optical beam splitter prepares a single photon superposition state. A $p = 0.5$ detection probability assumes an ideal 50/50 beam splitter. No coincident clicks will occur [9].

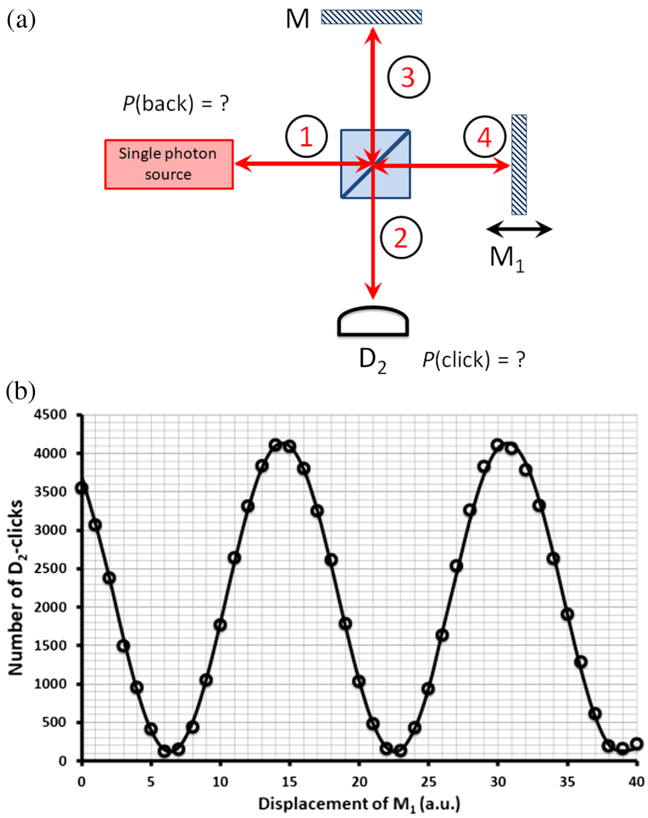

FIG. 2. (a) Setting up a Michelson interferometer by placing mirrors instead of detectors D_3 and D_4. (b) Quantum interference: The counting rate of single photon clicks of D_2 depends on the position of mirror M_1 [9].

transmission (D_3). The experiment shows no coincident clicks; thus, the results of D_4-click and D_3-click are complementary [9]. The PSI-based explanation relies on the calculation of the probabilities.

In this process, a single photon input state $|1_1\rangle$ [one photon at mode (1)] is transformed into a linear superposition of the two possible single photon output states. This is one of the simplest quantum-mechanical processes:

$$|1_1\rangle \rightarrow |\psi_{\text{out}}\rangle = R|1_3\rangle + T|1_4\rangle.$$

Following Born's rule, the probability of a click of the detector D_3 or D_4 can be determined (inserting $|R|^2 = |T|^2 = 0.5$ for a 50/50 beam splitter)

$$\begin{aligned} \mathrm{P}(\mathrm{D}_3) &= \mathrm{P}(|\psi_{\text{out}}\rangle, |1_3\rangle) = |\langle 1_3|\psi_{\text{out}}\rangle|^2 \\ &= |R|^2 \underbrace{|\langle 1_3|1_3\rangle|}_{=1}^2 + |T|^2 \underbrace{|\langle 1_3|1_4\rangle|}_{=0}^2 = 0.5 \\ \mathrm{P}(\mathrm{D}_4) &= \mathrm{P}(|\psi_{\text{out}}\rangle, |1_4\rangle) = |\langle 1_4|\psi_{\text{out}}\rangle|^2 \\ &= |R|^2 \underbrace{|\langle 1_4|1_3\rangle|}_{=0}^2 + |T|^2 \underbrace{|\langle 1_4|1_4\rangle|}_{=1}^2 = 0.5. \end{aligned}$$

Clicking of D_3 or D_4 destroys the superposition, leading to the unique measurement. The mirrors in the output of the optical beam splitter [Fig. 2(a)] will not destroy the superposition. The beam splitter prepares a new single photon state, transforming the two input states from the mirrors into a state superposition at D_2. The counting rate and therefore the probability of a D_2-click now depends on the position of the mirror M_1 [Fig. 2(b)]. This is due to the different propagation phases of the superposed states, which are caused by different separation between the mirrors and the beam splitter.

$$|\psi_{\text{out}}\rangle \rightarrow |\psi_{\mathrm{D}_2}\rangle = c_L|1_L\rangle + c_{L_1}|1_{L_1}\rangle.$$

Born's rule gives the probability of finding the photon state $|\psi_{\text{out},\mathrm{D}_2}\rangle$ and thus for a D_2-click:

$$P(\text{click}) = ||\psi_{\mathrm{D}_2}\rangle|^2 = (1 + \cos\Delta\varphi)/2 = \cos^2(\Delta\varphi/2),$$
$$\Delta\varphi = \varphi_L - \varphi_{L_1} = 2\frac{\omega}{c}L - 2\frac{\omega}{c}L_1 = 2\frac{\omega}{c}(L - L_1),$$

in keeping with the results of the experiment. For technical details of the experiment, original measurement results, and the complete quantum state pointer algebra, see Refs. [9,21].

The experiment (Figs. 1 and 2) is set up to simultaneously demonstrate two attributes of single photon states as a consequence of their superposition with stable phases at the beam splitter: an absence of coincidences at the "naked" beam splitter and a phase-dependent counter output in a Michelson setup. Though the events that allow for finding the photon at mirror M or M_1 can further be assumed to be noncoincidental, different phases of single photon states now lead to maxima or minima of the measured counting rate [probability $P(D_2)$; Fig. 2(b)] of a click of D_2. This phenomenon is referred to as quantum interference. The important role of PSI reasoning is apparent.

A pure quantum phenomenon (missing coincidences at the beam slitter) changes to a seemingly commonplace phenomenon (interferences in the Michelson interferometer) only by adding the mirrors. Naive dualism would describe the experiment with classical particles and classical waves by switching between these models for no reason other than the inability of classic particles to interfere. There is no need for a switch like this if one relies on quantum reasoning to discuss the probability of detector clicks in both cases derived from photon states. In the latter, a superposition of these states causes a phase sensitivity of the probability.

The implementation of quantum physics in secondary school curricula entailed the development of a range of relevant educational concepts. Across the board, these curricula have been geared to demystify the image of quantum theory and to ensure a permanent part of teaching physics [4]. To illustrate the breadth of the spectrum, two clearly different and well-evaluated concepts are mentioned briefly here. The Erlanger teaching concept aims to raise student awareness of the importance of modern quantum technologies ("Quantum world as the world of technology") [22]. The methodological focus is considerably more pronounced in the quantum interactive learning tutorials for undergraduate courses in quantum mechanics proposed in Ref. [23]. There, students are helped to build links between quantum phenomena and formal theory through computer-based visualization tools.

C. Construction of explanations in science

Although apparently focused on science learning for younger students, the CER scheme (*claim*, *evidence*, and *reasoning*) proposed by McNeill and Krajcik [24] well illustrates the task of explaining the quantum optical key experiment (Table I). To give an idea of how quantum reasoning may be integrated into a concept of learning and teaching, a few comments have been added. For the development of an appropriate set of physical (even quantum physical) explanations, it is necessary to decontextualize explanations to enable students to explain similar physical phenomena, always using the same physical principles and laws. This leads to the development of cognitive heuristics for the identification and recognition of the relevant physical principles for an explanation. The structure of this approach strongly refers to the idea of knowledge in pieces [25,26].

The challenging explanation task is to link the classical dualistic phenomenon (evidence) with a dualism free quantum physical claim (neither as classical particles nor as classical waves). The components of PSI reasoning deliver the perfect resource for explaining the phenomenon without recourse to classical bonds.

D. Reasoning patterns

In order to illustrate student conceptions (reasoning patterns) to explain quantum physical phenomena, three approaches are presented here:

1. Classical reasoning

Explaining interference phenomena with the wave model of light and the absence of coincidences in the quantum interaction of light with the beam splitter with the particle model.

This classical physical modeling approach to explain interference phenomena follows secondary school curricula and is strongly connected to the mental picture of water waves. Assuming the light particles are carried by a wrapping light wave, particlelike behavior of light (missing coincidences at the beam splitter) is compatible with this mental picture. Such a connection to mental pictures is described as a necessity for conceptual development [27].

Example: Single photons are detected either in the reflected or transmitted output because photons are particles. In the Michelson setup, one instead observes interference because the photons are components of the light waves.

2. Semiclassical reasoning

The experiment stipulates the reasoning pattern. Depending on the particular phenomenon, the rules of the concept of classical particles or those of classical waves apply.

As already mentioned, the formation of analogies to classical waves and classical particles (wave-particle dualism) is a popular nonmathematical approach to solving the contradiction in the Copenhagen Interpretation of Quantum Physics. This dualistic concept is considered naive because the mathematical formalism of quantum theory is not accessible in secondary school, and the strict regularities of the Copenhagen interpretation for the establishment of

TABLE I. Application of the CER-scheme.

The *claim* addresses the problem or the phenomenon, which should be explained.	*Single photons are quantum non-waves, non-classical entities with a clear and exclusive quantum physical behavior.*
The *evidence* supports the claim. Evidence arises from observation of things that could be seen and thus uses terms and arguments from the real world.	E1: *Either* D_3 *clicks or* D_4 *but never both detectors together.* E2: *A dependency of the counting rate Z on the mirror displacement* (Δx) *can be observed. Analyzing the dataset, one finds the pattern for the counting rate typical of interferences.*
The *reasoning* links the claim and the evidence by reducing the phenomenon to fundamental scientific principles. For quantum attitudes of the claim, the reasoning obviously must exhibit quantum features.	*Probability: The click of a detector is random. The click probability p of detection is derived from the probability amplitude by Born's rule (squared absolute values of the probability amplitude).* *Superposition: The complete quantum state of the system is described by the superposition of two single-photon states (transmission state and reflection state). The probability of a* D_2*-click is derived from Born's rule.* *Interference: The pattern of the counting rate is interpreted as a quantum interference due to the phase difference* $\Delta\varphi$ *proportional to the displacement of the mirror.*

an analogy are incomprehensible for students. Due to their lack of comprehension, students tend to produce visual evidence and base their arguments on naive wave-particle dualism as semiclassical reasoning for the explanation of quantum phenomena (see Ref. [27]).

In naive wave-particle dualism, students maintain their concepts of classical waves and particles while adding context sensitivity as an apparent quantum physical concept. This context sensitivity determines whether the quantum object occurs as a classical particle (naked beam splitter) or a classical wave (Michelson setup). The classical key ideas of locality and determinism can thus be maintained.

Example: Single photons are detected either in the reflected or transmitted output because photons occur as particles in this experimental setup. In the Michelson setup, on the other hand, they occur as waves and one observes interference.

3. Quantum reasoning

Using the quantum physical key ideas of probability, superposition, and interference (PSI) for the explanation of a quantum physical phenomenon.

This procedure is rather close to formal mathematical reasoning in quantum physics because the state function and its temporal development are the foundation of this quantum reasoning.

Example: Single photons are detected either in the reflected or transmitted output because the optical beam splitter creates a superposition state, which can be interpreted as reflecting and transmitting a photon concurrently. Whether a photon is reflected or transmitted will be realized randomly. In the Michelson setup, on the other hand, it is not possible to distinguish between reflected or transmitted via detection. Thus, the complete superposition state will be detected and quantum interference will be observed.

E. Quantum questionnaires

In order to investigate the effect of the quantum optical key experiments on quantum reasoning, a suitable questionnaire is needed. Different instruments can be found today, testing student understanding of quantum physics [2,10]. Most of the quantum *mechanical* oriented instruments emphasize mathematics and focus on student understanding of the solution of the Schrödinger equation, the de Broglie wavelength, the measurement process, and the behavior of electrons and atoms [10,28–30]. Examples of the quantum *optical* type are approaches that are used to evaluate student understanding of the behavior of single photons in a Mach-Zehnder interferometer [23,31]. The study that evaluates the *Erlanger Unterrichtskonzept* (Erlanger quantum physics course) is settled around single photon experiments [22]. However, the questionnaire mainly focuses on the technical realization of the experiments.

Existing questionnaires in the field of quantum education are mainly restricted to content knowledge about quantum physical principles (probabilistic interpretation, superposition), entities (wavefunction and operators), or models (atomic physics). Among them, only a few items seem to be suitable for measuring students' reasoning with regard to the quantum optical key experiment (for details, see Refs. [10,22,23,32]). A questionnaire focusing on procedural knowledge, in the sense of using the PSI concept of the quantum reasoning to explain phenomena of the quantum optical key experiment, is needed.

III. DEVELOPMENT AND VALIDATION OF THE QUESTIONNAIRE METHODOLOGY

The following section summarizes our understanding of the validation, measurement, and methodology used for the development of the item pool and its validation.

A. Basic requirements of the questionnaire and item content

To measure students' quantum reasoning ability, four different categories of items are necessary:

1. Content knowledge about PSI: The use of quantum physical key ideas for quantum reasoning requires a sufficient knowledge of PSI.
2. Phenomena: To investigate students' reasoning for quantum phenomena, it is necessary to know whether they perceive the essential details of the phenomena (e.g., electron double slit diffraction pattern, Debey-Scherrer diffraction pattern observed on the screen of the electron tube).
3. Explanations: To analyze students' reasoning, items are required that invite students to explain and interpret quantum phenomena.
4. Previous knowledge about wave optics: Due to traditional standard physics curricula, wave optics is assumed to be a student's previous knowledge of interference.

To make the adoption of well-tested items from the literature (e.g., explanation of quantum phenomena [10,22,23] resp. classical wave optics [32]) transparent and comprehensible, three selection criteria were formulated.

1. The item should address physical reasoning (classical reasoning, semiclassical reasoning, or quantum).
2. The item should address the student's reasoning (PSI, quantum state function, mathematics, or classical optics).
3. The item should be suitable for upper secondary and undergraduate students (matching the language or students' estimated ability).

About 28 items from already existing questionnaires from the field of quantum optics fit these criteria, and 12 additional items were also developed. To get information about student concepts of physics, the single- and multiple-choice items contain issues focusing on PSI, semiclassical reasoning, and/or classical reasoning (e.g., classical wave optics). The instrument was completed by a multiformat questionnaire comprising open-ended, single-, and multiple-choice items. Single- and multiple-choice items have been dichotomously categorized (right or wrong answer) and have high objectivity [33]. They capture a student's tendency to use classical, quantum, or semiclassical reasoning. Three open-ended items were developed and added to help confirm the results.

B. Validation

Validity is the degree to which interpretations of the test scores for proposed uses are based on evidence and theory. Validation means a critical evaluation of the interpretation or use argument (IUA) and its assumptions [34]. The interpretation of the results of the test scores is thought to be valid when all assumptions can be supported by evidence or theory, and when the IUA seems to be coherent and complete. Even if a single assumption cannot be supported by evidence and/or the IUA seems incoherent or incomplete, the score and its interpretation are invalid and the IUA or score must be revised [34].

It is therefore necessary to write down the IUA and the incorporated, inherent assumptions of the measurement scale's use. Subsequent sufficient methods to prove its validity can then be identified. For the developed measurement scale, the following is assumed:

1. The questionnaire measures quantum reasoning on the level of knowledge of upper secondary and university undergraduate students.
2. The number of items and their estimated difficulty adapts to the performance of the target group.
3. The items activate content knowledge.
4. The correct answer cannot be derived from the formulation of items or distractors.
5. The items are scientifically sound and use common quantum physical terms.

C. Methodology

To provide empirical evidence that supports the assumptions, different qualitative and quantitative studies were conducted (Fig. 3.). With the qualitative studies, the scientific correctness and student's understanding of the items were checked, the quantitative (Rasch analyzed) field testing with 84 undergraduate physics students generated data to check for unidimensionality. In order to get more information about the item interplay, an explorative factor analysis was conducted.

1. First item tryout—group discussion

To get an initial impression of the item pool, six physics students (five of them students of teaching) were asked to carry out the pool of items and to make comments

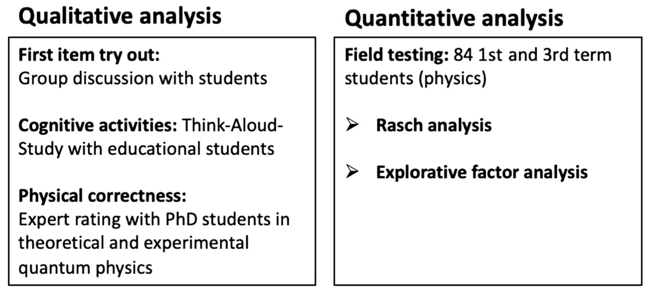

FIG. 3. Qualitative and quantitative studies used to check for empirical evidence.

(e.g., challenging diction). Afterward, items that gave contradictory answers were selected for a subsequent group discussion aimed at getting information about difficulties with answering the items correctly.

2. Cognitive activities—think-aloud interviews

To evaluate attempts to answer the items (assumptions 3 and 4), think-aloud interviews are a common method [32,35,36]. Students of teaching were asked to work on the item pool and to evaluate how confident they were in their choice. The interviewer took notes when the students had a problem with an item or found mistakes in the stem or the answers. To support thinking aloud, the participants were asked to work on the test items in pairs. This cooperative format was thought fruitful for encouraging the thinking-aloud process. Two groups were interviewed (duration of the interviews: 104:54 and 55:26 min).

To interpret the interviews, the strategy for solving the items was looked at in a subsequent qualitative content analysis. The coding system developed by Groth [37] and summarized in Berger *et al.* [35] was adapted for this purpose (Table II).

3. Scientific correctness—expert rating

To prove the scientific correctness of the items (assumption 5), Ph.D. students from the CRC 1227 Designed Quantum States of Matter (DQ-mat), as experts in the field of quantum optics, were asked to answer the items and to comment on each item. About 13 Ph.D. students participated. To analyze the rating, the answers were collected and analyzed descriptively. If differences between the anticipated correct answer for an item and the favored answer by the Ph.D. students were seen, the comments were used to reformulate the item stem or the answer options.

4. Field testing and quantitative analysis

Measurement and item-response theory. To ensure a biunique correspondence between theory and reality, measurement relies on unidimensional and well-defined measurement scales [36,37]. The item-response theory (IRT)

TABLE II. Coding system for qualitative content analysis.

Type	Certainty in choice	Item	Relevant strategy	Nonrelevant strategy	Interaction with the partner	Other
Description	Certainty of students' choice	Students reading the item or discuss its formulation	Students use content knowledge for item solution	Students guess the correct answer or the choice is without a reason.	Students talk to each other without a connection to the item.	Anything else like laughing or sounds

describes the relationship between the test takers' response to a given item and an underlying latent trait, which should be measured [36,38]. To reduce the complexity of the measurement, the dichotomous IRT assumes that the underlying trait, like quantum reasoning, can be expressed by a unidimensional measurement scale. The test taker's success is assumed to be only determined by their ability and the item difficulty [39].

A well-known dichotomous IRT model is the 1D Rasch model [40]. It gives the opportunity to determine the extent to which the latent trait is pronounced and thus to rank students' ability [41]. The model fit also indicates the extent to which a unidimensional measurement scale can be defined [40,42]. The data for the Rasch analysis were generated in a field test with first and third term physics students. Since quantum physics is a topic in third term physics, the abilities of the first term students in quantum physics and of upper secondary level high school students can be assumed to be virtually identical, as student knowledge comes mostly from upper secondary school. The situation is a little different for third term students because they should have learned something about quantum phenomena in basic university experimental physics lessons. The questionnaire was provided online via the *Limesurvey* platform and 84 students participated, with 39 students from the first term. According to Linacre [43], such a sample size is sufficient for the construction of a Rasch model.

Preparation of data. The *R*-package *eRm* [44], which was used for the Rasch analysis, works with dichotomic items. Therefore, the four open-ended items were recoded: "1" for a correct answer, "0" for everything else. Four items, such as Item 9 (Fig. 4) had three different answer options

Item 9

Indicate whether the following statements are true (T) or false (F)

1. If an electron is observed at a certain location, the electron was already there before measurement.
2. If a certain electron velocity is measured, the measurement has determined the velocity.
3. If the electron's kinetic energy is measured, it will maintain the energy if it is in uniform motion.

FIG. 4. Item 9: Coding three different answer options.

in the expert rating. These were coded with 1, when all of the item's statements were evaluated correctly and with 0 otherwise.

Explorative factor analysis. Quantum reasoning can be influenced by several (psychological and educational) constructs, such as knowledge about models and a well-trained change in perspectives. The item fit-statistics of the Rasch analysis allow for the evaluation of the assumption of unidimensionality, with the explorative factor analysis (EFA) generating more information about some possible item interplay (Fig. 5). A unidimensional measurement scale corresponds to the identification of a single factor in the dataset. However, if items could be identified which only load marginally to this factor or which load on an additional factor, specific content analysis for the affected items will provide more information, especially regarding quantum key ideas.

To perform the EFA, the R-Package *Psych* was used [45]. To evaluate the suitability of the dataset for EFA, the *measure of sample adequacy (MSA)* (*Kaiser-Meyer-Olkin criterium*) can be used, which indicates the correlation between the measured variables. A factor analysis is reasonable if the *overall MSA value* is greater than 0.5 [46]. To estimate the number of factors, a parallel analysis was conducted, discriminating the correlations between the items and the factor from random correlations [47].

The EFA results, the results of the Rasch analysis, and a careful view of the content frame of the questionnaire help to decide which items should be excluded from the item pool and which items should be maintained. As stated at the beginning of Sec. III B, validity refers to the degree to which interpretations of the test scores for proposed uses are based on evidence and theory, while content analysis rather refers to the well-targeted nature of the items. The final decision on how to use the items ultimately lies with the principal investigator alone.

IV. RESULTS RESEARCH GOAL 1

Following the guidelines for the item selection, 28 items were selected for an initial item pool (pool28). The item selection process started with the group discussion and uncovered two problems with pool28:

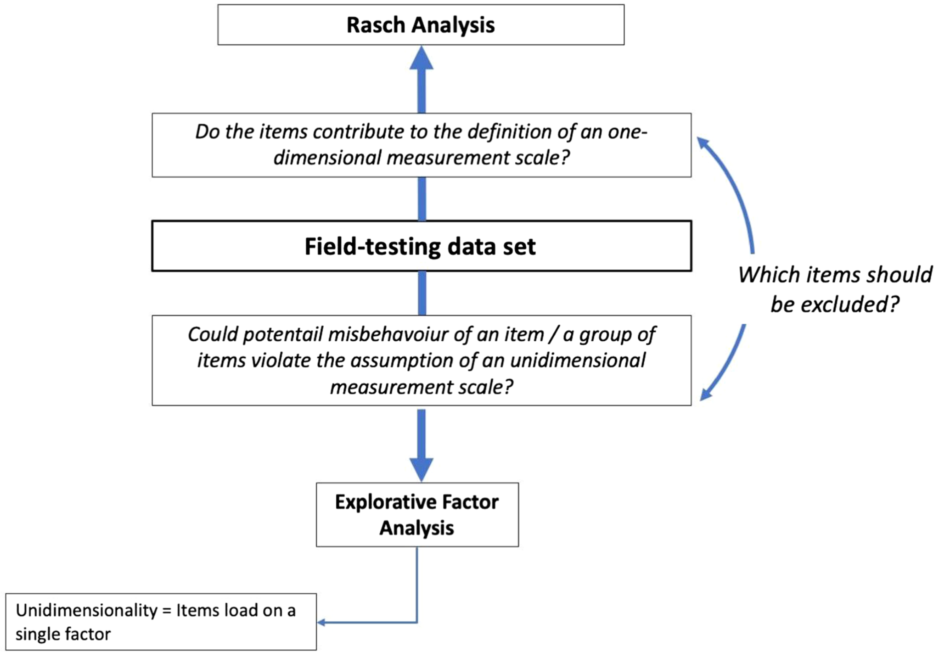

FIG. 5. Methods to evaluate the quality of the items.

- pool28 is challenging because the items are text intensive, particularly the items from Singh [23].
- Several items are challenging due to disproportionate technical language requirements (e.g., the *collapse of states*).

When considering these two results, it was necessary to exclude seven items and to reformulate several others. For a revised pool, 12 additional items were developed, focused on the prototypical course content of secondary school quantum teaching: (1) an optical beam splitter, (2) the Michelson interferometer, (3) the electron diffraction tube, (4) classical wave optics, and (5) deterministic argumentation of classical physics. The revised pool was used for further validation studies. Here, we found clear arguments for the exclusion of the items VW 2 (*identify different interference phenomena*) and VW 5 (*interference of white light using a CD*) as well as supportive evidence for the IUA; an overview is shown in Table III.[1]

To get additional insights into the item interplay, an explorative factor analysis was conducted with pool 31 revealing a slightly different picture. Although the Rasch analysis gave a clear and supportive argument for the assumption of a unidimensional measurement scale, the EFA identified two underlying factors in the dataset. Twenty items load on the first factor and seven items on the second factor (Table IV). Due to their small loadings, these items are less relevant from a rigorous statistical point of view. However, it can be seen that, with the exception of Item 3, these items are from the self-developed stem targeting naive wave-particle dualism (e.g., *Item 18: option 2 seems to be the correct answer for semiclassical reasoning;* Fig. 6). The occurrence of the second factor illustrates what has been already stated above: dualistic reasoning is highly attractive to students.

A. Further item selection

Table V shows seven items that do not load on any factor, although they show no misfit in the Rasch model. To decide whether the items should be excluded from the item pool or maintained, the items' content was again compared with the research goal [48]. As Table V shows, the items have different difficulties in the scope of −0.395 to 2.188 and, except for *Item 10* and *Item 15*, the MNSQ infit and outfit values are lower than 1.5. Two of these items were maintained due to their physical significance:

- In *Item 8*, a single photon Mach-Zehnder-like setup without the second beam splitter is considered. The results are thus the same as in a simple beam splitter experiment.
- *VW3* asks students, why it is reasonable to use the wave model of light in classical optics, which is assumed to be the student's previous concept of light.

The final pool (pool26) now contains 26 items.[2]

B. Parsimony

The Rasch analysis and the EFA provide two results that are somehow complementary: While a one-dimensional

[1] A comprehensive description of the methods and results is available from the author M. W., moritz.waitzmann@dq-mat.uni-hannover.de.

[2] See the Supplemental Material [49] for the complete item list pool26.

TABLE III. Main results of the think-aloud study, the expert rating, and the Rasch analysis.

Method	Aim	Main results
Think-aloud	Analysis of students' strategies for solving the items (assumptions 1, 3, and 4).	Most of the utterances are related to physics. Thus, the items activate cognitive activities. Problems with the items $VW2$ and $VW5$ were uncovered.
Expert rating	Proof of the items' physical correctness (assumption 5).	Most of the items are physically correct. *Item 26's* stem was revised. Again, problems with $VW2$ and $VW5$ were uncovered.
Rasch analysis	Analysis of the items' suitability for the target group and the measurement scale's unidimensionality (assumptions 1 and 2).	In general, the items fit the target group. However, $VW2$ and $VW5$ were excluded, due to the previously uncovered problems and an item misfit. In addition, one person must be excluded, due to a misfit. → pool 31

Rasch model shows no misfitting items and sufficient suitability between the items and the target group (especially for pool26), the EFA uncovers items which load on a second factor and further items which load on any factor. To argue for whether a one-dimensional or a two-dimensional model is preferable, the philosophy of science provides the epistemological principle of parsimony, known as *Ockham's Razor*. It estimates the simplicity of one- and more-dimensional models [48]. The principle of parsimony tells us that if there are different models for the same explanation, the model with the lowest number of variables should be used [50,51]. For Rasch, parsimony can be evaluated from the number of model parameters n_p. Typically, coefficients of the information theory can be used to compare the model's simplicity. Akaike's information criterion (AIC) is common for large sample sizes ($\mathrm{AIC} = \mathrm{variance} + 2n_p$) and the BIC (Bayes' information criterion; $\mathrm{BIC} = \mathrm{variance} + \log(N \cdot n_p)$) for a small sample of size N. The lowest AIC or BIC value indicates the simplest model [50,51]. The final decision for or against a model must be made by the researcher [48].

For this work, a one-dimensional Rasch model is compared with a two-dimensional one, to respect the items that load on a second factor (Table VI). For the AIC values, a marginal difference between the two models can be seen, which prefer the one-dimensional model. However, taking the BIC values into account, it becomes very obvious that a remarkable difference between a one-dimensional Rasch model and a two-dimensional one can be seen. Following the principle of parsimony, the use of a two-dimensional Rasch model has no benefit over the one-dimensional

TABLE IV. Loadings on the two factors above the cutoff value of 0.3.

	Category	MR1	MR2
Item 1	Quantum phenomena	0.464	
Item 2	Quantum phenomena	0.408	
Item 3	Content PSI	0.398	0.304
Item 4	Content PSI	0.408	
Item 5	Content PSI	0.434	
Item 6	Observation of phenomena	0.522	
Item 7	Content PSI	0.409	
Item 8	Quantum phenomena		
Item 9	Content PSI		
Item 10	Content PSI		
Item 11	Content PSI	0.393	
Item 12	Content PSI	0.624	
Item 13	Quantum phenomena	0.610	
Item 14	Quantum phenomena		
Item 15	Explanation of phenomena		
Item 16	Explanation of phenomena	0.376	
Item 17	Explanation of phenomena		0.759
Item 18	Explanation of phenomena	0.301	0.716
Item 19	Content PSI	0.368	
Item 20	Explanation of phenomena	0.543	
Item 21	Content PSI	0.473	
Item 22	Quantum phenomena	0.591	
Item 23	Content PSI		
Item 24	Explanation of phenomena	0.312	
Item 25	Explanation of phenomena	0.358	0.316
Item 26	Explanation of phenomena		0.441
Item 27	Explanation of phenomena		0.557
VW 1	Preknowledge waves	0.389	
VW 3	Preknowledge waves		
VW 4	Preknowledge waves	0.561	
VW 6	Preknowledge waves	0.432	

Item 18

Evaluate the following statement

Photons must be particles, because they were either reflected or transmitted by an optical beam splitter

1. The statement is true. Photons are tiny particles which are covered by the light wave and propagate with it. The beam splitter distributes them equally.
2. The statement is false. Photons are both particle and waves. Depending on the experiment, they are either waves or particles.
3. The statement is false. Photons are neither particles nor waves. The beam splitter transmits and reflects the photon concurrently. Until measurement its location is undefined.

FIG. 6. Item 18 as an example of the naive wave-particle dualism.

TABLE V. Discussed items' difficulty and fit parameter.

Item	Difficulty	Infit/outfit MNSQ
Item 8	1.226	1.591/1.247
Item 9	0.784	1.092/1.089
Item 10	2.188	1.143/1.122
Item 14	−0.395	1.313/1.236
Item 15	1.955	1.064/0.953
Item 23	0.371	1.152/1.140
VW3	−0.095	1.318/1.205

TABLE VI. AIC/BIC values for 1D-, 2D-, and 3D-Rasch model.

Dimensions	AIC	BIC
1	2460	2525
2	2479	2606

Rasch model. The assumption of a unidimensional measurement scale is therefore still reasonable.

C. Final questionnaire

The final questionnaire is a mixed-format test with 26 items which can be placed into four categories (see page 11 and Table VII). To illustrate these categories and link them to the defined construct of quantum reasoning, prototypical items will be presented in the following.[3]

1. Content knowledge about probability, superposition, and interference

In order to be able to use PSI as tools for quantum reasoning, basic knowledge of these key ideas is required. To analyze students' knowledge of PSI, seven items focusing on these key ideas, its interpretation, or its experimental observability, can be found in the pool.

A prototypical item from this category is *Item 19* (Fig. 7), regarding the key idea *probability*.

The optical beam splitter experiment perfectly demonstrates the quantum randomness: due to the indeterminism of quantum physics, there is no way to predict the experimental outcome, when a single photon impinges on the optical beam splitter. One can only make probabilistic statements and predict each detector's click with a probability of 50% and a coincident click of both detectors with a probability of 0% (option 3). A common misconception is that a current random event will influence the randomness of future events [52]. For example, in the beam splitter experiment, either the detector in the reflected output clicks or the one in the transmitted output, each with a probability of 50%. For a very large number of measurements, the measured frequencies will converge to the probabilities. The typical law-of-large-number-misconception assumes that when the first photon causes a click of the detector in the reflected output, the probability of detecting the second photon in the transmitted output increases enormously (option 1). Similar items were formulated for the principle of superposition and interference.

2. Quantum phenomena

To ensure that students correctly expect observations in experiments with single quantum objects, the questionnaire contains three items that focus on experimental phenomena. Item 2 (Fig. 8) suggests three different observations, conceivable in principle in the double-slit experiment with single photons (answer no. 1 is correct).

3. Explanations

Items are necessary for analyzing students' reasoning and inviting students to explain phenomena. The answer options reflect classical, dualistic, and quantum physical reasoning patterns. A prototypical item for this category is *Item 27* (Fig. 9), which asks for an explanation of quantum interference in the double-slit experiment with electrons. In classical physics, electrons are perceived as hard tiny particles of mass m. They are therefore not able to interfere (option 3). From a quantum physical point of view, whether electrons are particles or not is irrelevant. One observes a probability pattern due to the quantum interference of the probability amplitudes (the state function, option 1). Dualistic reasoning resolves the contradiction, interference of classical particles, by assigning a wavelike property to the classical particle causing the interference as described in classical wave models (option 2). Using these items, it is possible to analyze the extent to which students use quantum reasoning and also to uncover the experimental context in which quantum reasoning is preferred.

4. Wave optics

To ensure students' preknowledge of interference in the context of wave optics, the questionnaire contains four pertinent items. The item VW 1 (Fig. 10) prototypically demonstrates this item category.

D. Final conclusions on research goal 1

The Rasch analysis and the EFA provided clear criteria for the item selection. The separation reliability (selectivity) of 0.830 is high [40], and the items thus work together sufficiently well defining a unidimensional measurement scale. The additional explorative factor analysis gave indications of a further underlying factor focusing on naive wave-particle dualism. However, following the principle of parsimony, a one-dimensional Rasch model and thus a unidimensional measurement scale should be employed.

[3]The complete item list pool26 is shown in the Supplemental Material [49].

TABLE VII. Categorization of the items.

	Content knowledge PSI	Observations of quantum phenomena	Explanations of quantum phenomena	Pre-knowledge wave optics
Item	3, 4, 5, 7, 11, 12, 19, 21	1, 2, 6, 13, 22	15, 16, 17, 18, 20, 24, 25, 26, 27	VW 1, VW 3, VW 4, VW 6

The assumption of the interpretation-use argument could be supported by five validation studies, particularly the think-aloud study, the expert rating, and the Rasch analysis:

- The think-aloud study confirms the enhancement of students' cognitive activities while solving the items.
- The expert rating provides fruitful indications for the improvement of item formulation and confirms the scientific correctness of the item pool.
- The fitted Rasch model provides a measurement scale that can be assumed to be unidimensional regarding the model fit indices.

The test score and its interpretation can thus be seen as valid.

Item 19

Two single photons were successively impinge on an optical 50/50 beam splitter. Detectors were installed in both beam splitter outputs. What will be observed?

1. Each detector will register a photon. Only the second photon will allow for a prediction which detector will register the photon.
2. Both detectors will register two photons.
3. Always only one detector will register a photon. However, we can only predict the probability of which detector will register a photon.

FIG. 7. Item 19 refers to the key idea of probability.

Item 2

When the double-slit experiment is done with single photons, one observes…

1. … a probability distribution with maxima and minima, looking like an interference pattern of classical waves on the screen behind the double-slit.
2. … two clearly defined detection spots on the screen behind the double-slit.
3. … two clearly defined detection spots and a maximum zero order on the screen behind the double-slit.

FIG. 8. Item 2 refers to three principally possible, but different observations.

Item 27

Evaluate the following statement:

"Electrons must be waves, due to their interference in the double-slit experiment."

1. The statement is wrong, because the probability amplitudes interfere.
2. The statement is right, because electrons are particle which demonstrate wave properties in some experimental setups.
3. The state is wrong, because electrons are particles, which cannot interfere.

FIG. 9. Item 27 asks for the explanation of quantum interference in the double-slit experiment with electrons.

VW 1

How would one describe the interference of waves?

1. In general, the superimposition of at least two electromagnetic waves.
2. In general, the superimposition of at least two waves.
3. In general, the superimposition of exactly two waves.

FIG. 10. A typical wave optics preknowledge item.

V. GOAL 2: TESTING QUANTUM REASONING

A. Motivation and research goal

The development of specific quantum reasoning is expected to be fruitful for introductory quantum physics in upper secondary school and for university undergraduates. Several projects can be found that aim to develop teaching sequences, real experiments, simulations, and games for teaching phenomenological quantum physics [2,6,9,21,22,53,54]. In 1986, Grangier *et al.* demonstrated a strong anticorrelation effect and single photon interferences in an analog experiment, providing compelling arguments against any dualistic interpretation [55]. In educational research, the experiment has been discussed since the early 2000s because of its potential to overcome dualistic concepts and to motivate quantum reasoning [3,9]. It is thus seen as a key experiment for quantum reasoning [9]. However, there is still no research on the assumed effect of the key experiment. The following will present the initial results of a study and will answer the research question: To what extent does engaging students with the key experiment lead to using the key ideas PSI in quantum reasoning?

B. Method

To answer the research question, a one-group pre-post study with 80 second term students in the undergraduate physics lab was conducted. Students answered the questionnaire one week before and one week after they performed the experiment.

Because one main point of interest lies in the change from pre to post in terms of the contents and more specific interpretation of possible changes, a Rasch analysis and a pre-post comparison using the Wright Map are used [56]. By using Rasch as a probabilistic model approach, the scores are determined by the student's ability and item difficulty. Scores generated using a probabilistic model account for the difficulty of items are linear. Subsequently, a Rasch analysis for pre- and post-test was carried out.

To ensure the model conformity of the items and the test takers, the MNSQ values for person and item IN- and OUTFIT were regarded again. Items with MNSQ values greater than 2.0 (underfitting the model) or with MNSQ values lower than 0.5 (overfitting the model) represent a misfit [42]. Fourteen people (17.5%) and one item ($VW4$) had to be excluded in order to ensure comparability of the pre- and post-test. The average ability in pre and post was compared first. To get detailed information about what students had learned when performing the key experiment, the quartiles of student abilities were compared with the corresponding items. This approach makes it possible to interpret the students' learning gains at the content level and to compare them with the learning objectives [56].

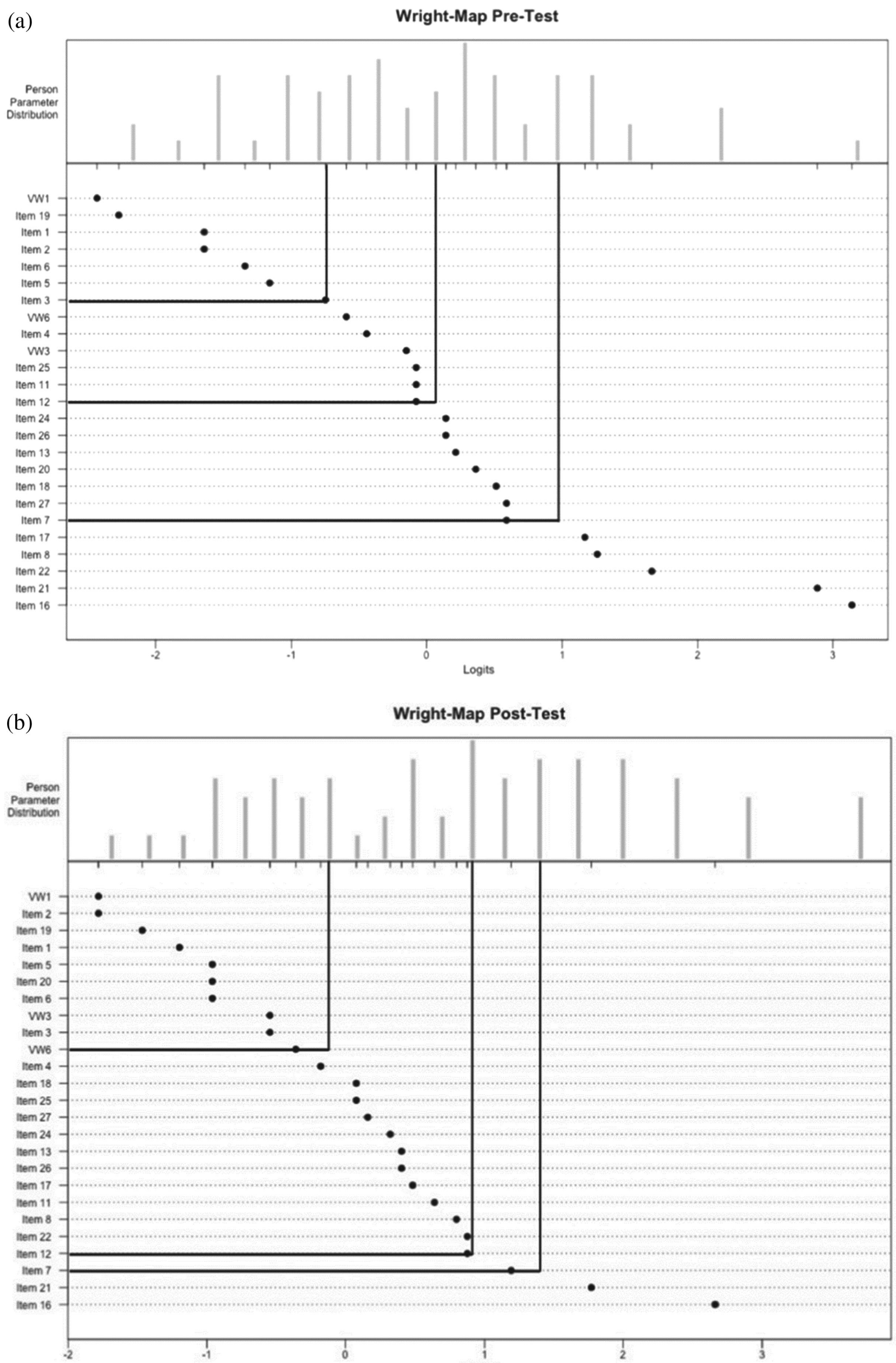

FIG. 11. (a) Quartiles of students' abilities in the pretest. (b) Quartiles of students' abilities in the post-test.

TABLE VIII. Description of the quartiles of student ability.

Quartile	The items' content	Example item
Pretest		
1	School knowledge about quantum physics	Item 2: *When the double-slit experiment is done with single photons, one observes…*
2	Classical interference and the photon	VW 6: *How a change of the interference fringes, from light to dark can be facilitated in the Michelson interferometer?*
3	Interpretation of the behavior of single quantum objects	Item 24: *Rate the following*: *"A single photon's interference cannot be explained in classical ways."*
4	Quantum interference/bomb test	Item 22: *Is it possible to observe a single photon's interference?*
Post-test		
1	School knowledge about quantum physics/ classical waves	Item 3: *Rate the following:* *"Due to the observation of interference in the single photon double-slit experiment, the photon must be divided into half by the double-slit."*
2	Interpretation of single quantum objects' behavior/quantum interference (superposition)	Item 12: *Consider the following: A single photon is inside the Mach-Zehnder interferometer in which the second beam splitter is missing. The detectors D1 and D2 haven't detected a photon yet. Indicate whether the following statements about the photon are true (t) or false (f).*
3	Superposition in general	Item 7: *Indicate whether the following statements are true (T) or false (F)* *If one measures a certain value of an electron's velocity, the measurement determines the velocity.*
4	Bomb test	Item 16: *Only by positioning a bomb into the Mach-Zehnder interferometer's arm, the click probability of the detectors D1 and D2 is changed. How would you explain this?*

C. Results

Figure 11 shows the Wright Maps for each test.

- When comparing the item difficulties, there is no significant change from pre to post (with the exception of *Item 16* and *Item 21*; the last two items in Fig. 11). Thus, the pre- and post-person abilities are reasonably comparable and are not conflated with changes in item difficulties.
- The histograms in the upper part of the Wright Maps represent a statistically significant increase in the person's ability on average, which was proven by a nondirectional t test ($t = -7.655$, d.o.f. $= 65$, $p = 5.896 \times 10^{-11}$; 95% confidence interval [-1.475, -0.180], $\Delta = -0.828$, $d = 0.942$).

The Rasch analysis provides information about the changes in student's ability and the item contents [56]. Therefore, this approach gives insight into the effect of the key experiment.

At first, the key experiment helps to gain access to the basics of quantum physics: from pre to post the number of items, which can potentially be answered correctly by the students in the first and second quartile increases, while the upper quartiles were characterized by a lower number of corresponding items. Furthermore, the quartiles can be characterized by its corresponding items' content, as shown in Table VIII. The following can be observed:

Although the first quartile can be characterized as school knowledge, items regarding classical interference and classical waves are in the pretest in the second quartile. From pre to post, the items $VW4$ and $VW6$ correspond to the first quartile: the students can solve them after they perform the key experiment in the post-test. From pre to post, the students are also able to explain single quantum object interference. In the pretest, only students in the third and fourth quartiles could answer these items correctly. In the post-test, seven students (10.6%) in the second quartile could explain the interference.

In terms of the principle of superposition, a different situation can be seen:

- *Item 8* regarding the Mach-Zehnder interferometer without the second beam splitter can be answered correctly by students in the fourth quartile in the pretest and by students in the second quartile in the post-test. This result can be seen as a strong indication that the key experiment helps students make sense of quantum physics phenomena.
- *Item 11* and *Item 12* remain in the second quartile. They ask for an interpretation of photon behavior, prepared by the first beam splitter of a Mach-Zehnder interferometer. Students had to think about the superimposed state. Can this be interpreted as photons taking both paths concurrently or not? In the pretest, these two items were easier to answer than items asking for the explanation for single quantum object interference. Perhaps, the interpretation of this state had already been discussed in school quantum physics and was then remembered by the students.

TABLE IX. Examples of the impact of the key experiment on students' reasoning.

	Pre	Post
Beam splitter (classical)	*"The single photon will not separate because it is a clear localized particle."*	*"Like little balls at a wall, single photons reach the detector for reflection and for transmission with always 50% probability (and perhaps with a very small probability both of them coincident)."*
Beam splitter (dualistic)	*"The setup specifies the character. Here photons show themselves as particles."*	*"Photons are reflected or transmitted with about the same probability, but not both (clear determined by the detectors). It follows, in this case they behave like particles."*
Beam splitter (quantum)	*"Photons are not always clearly localized, but only with a certain probability. So, they should take either route, reflection and transmission."*	*"Single photons are indivisible and are reflected or transmitted with 50% probability. This looks like the behavior of particles. However, both possibilities exist, only at the detector it is determined, reflection or transmission."*
Interferometer (classical)		
Interferometer (dualistic)	*"Now the photon reveals itself as wave. It will be split at the beam splitter. After reflection at the mirrors these parts are superimposed at the beam splitter leading to interference."*	
Interferometer (quantum)	*"We see maxima and minima of the counting rate due to interference. The particles interfere with themselves and the photon is no longer visible. The intensity of the light is the same as the probability."*	*"The photon is not localized and the probability amplitudes superimpose at the beam splitter after reflection at the mirrors. Depending on the position of the mirrors one gets maxima or minima. Particles are not able to produce a curve like this. We see a wavy character of the photon."*

- *Item 7*, *Item 16*, and *Item 21*, regarding the principle of superposition in general, remain in the third and fourth quartiles.

The identification of a student's reasoning is challenging because nobody knows what they are really thinking. On the other hand, the answers given in the interview open the door to some very basic mental patterns of the students. Table IX shows translated paraphrases of students' answers from the interview before and after they performed the experiment.[4]

As noted above, approaching quantum physics is clearly not a simple task. Without empirical evidence from a qualitative study, we get the impression from the interviews that much more must be done, to motivate students to enrich their reasoning with the concept of quantum states and their superposition to overcome dualistic boundaries (see conclusions and limitations in this article).

D. Conclusion goal 2, the effect of a quantum optical key experiment on students' quantum reasoning

The results show that the engagement with the key experiment seems to increase the person's ability on average and fosters quantum reasoning, allowing them to explain single quantum object interference. By looking at the concept of superposition, the engagement with the key experiment fosters a student's reasoning of the principle only in the context of interference or interferometer. Knowledge of the principle in general is still challenging and can be found only for students in the third and fourth quartiles. Most of the students deem the principle of superposition necessary for the occurrence of interference, but not as an exclusive quantum physical key idea. This example illustrates a possible application of the questionnaire: evaluating the experiments' effect on students' quantum reasoning. Furthermore, by comparing the ability quartiles and corresponding items in the pre- and post-test, it could be seen where students made progress in quantum reasoning and the content that is still difficult for them (in this study, the explanation of single photon interference versus the application of the principle of superposition).

VI. DISCUSSION OF LIMITATIONS

A. Limitation of the pre-post study

Due to the use of the same test in pre and post, the possibility of retest effects cannot be ignored [57]. Furthermore, the questionnaire is less selective in the third ability quartile of the post-test, because only one item corresponds to the students' ability. To overcome these limitations, additional interviews were conducted with a subgroup of 36 students, which allowed for an analysis of the student's reasoning for the explanation of simultaneous emergence of indivisibility and interference of single photons. See Ref. [21] for the initial results of the interviews. Furthermore, the study does not allow for a comparison of the quantum optical key experiment with

[4]An exhaustive compilation of students' answers (in German only) is available from the author M. W., moritz.waitzmann@dq-mat.uni-hannover.de.

other prototypical experiments in quantum teaching. Therefore, engaging students with a single photon quantum double-slit experiment combined with the use of quantum reasoning for its explanation might have the same effect. Further research is, therefore, necessary to compare the key experiment's effect on students with other prototypical experiments in quantum teaching.

B. Sample size limitation of the think-aloud study

The sample size is limited. In the think-aloud study, only two groups were interviewed, with differences in the duration of the interview and the depth of the discussion. However, both groups did not show great deviance. The expert rating confirms the results of the think-aloud study regarding problematic items. Furthermore, both groups were students of teaching and could have gained skills in quantum physics in their theoretical and experimental physics lectures. Research on quantum conceptions of students' teaching shows that they are familiar with typical quantum phenomena, such as the tunnel effect or the hydrogen atom. On the other hand, they have only a poor understanding of the fundamentals of quantum physics and are often attached to (semi-) classical conceptions [3]. It therefore seems reasonable to equate the students of teaching with undergraduate physics students at the beginning of quantum education. To use the questionnaire for upper secondary school students and to raise the number of participants a little, an additional think-aloud study should be conducted with this group.

C. The expert rating limitation

To check for scientific correctness in the construct, only physics scientists were asked, while experts on quantum education were not asked. The scientific point of view was thus proved by the physics experts, but an educational perspective on the questionnaire's theoretical framework is missing. On the other hand, the quantum reasoning approach was derived from fundamental characteristics and principles of quantum physics, meaning that a restriction on a physical expert rating seems reasonable.

VII. SUMMARY AND OUTLOOK

The work presented here provides an inventory that is able to test quantum reasoning on behalf of the key ideas of probability, superposition, and interference (PSI). The developed and validated instrument gives a deeper insight into students' usage of quantum reasoning for the explanation of quantum phenomena taught in upper secondary schools. The experimental setup seems suitable for overcoming dualistic concepts and motivating the usage of quantum reasoning tools, due to the demonstration of the dualism's internal inconsistencies and contradictions. Furthermore, it demonstrates the questionnaire's functionality and shows that students used the key ideas of probability and interference for quantum reasoning. However, an understanding and application of the principle of superposition would require a revision of student engagement with the experiment and likely further teaching materials that focus on the principle of superposition.

ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 274200144—SFB 1227.

---

[1] P. Dirac, *The Principles of Quantum Mechanics*, 3rd ed. (Oxford University Press, New York, 1947).

[2] K. Krijtenburg-Lewerissa, H. J. Pol, A. Brinkman, and W. R. van Joolingen, Insights into teaching quantum mechanics in secondary and lower undergraduate education, Phys. Rev. Phys. Educ. Res. **13**, 33 (2017).

[3] R. Müller, *Quantenphysik in der Schule [Quantum physics in School]* (Logos-Verl, 2003), Vol. 26.

[4] H. K. E. Stadermann, E. van den Berg, and M. J. Goedhart, Analysis of secondary school quantum physics curricula of 15 different countries: Different perspectives on a challenging topic, Phys. Rev. Phys. Educ. Res. **15**, 010130 (2019).

[5] R. P. Feynman, R. B. Leighton, and M. Sands, *Quantenmechanik [Quantum Mechanics]* (de Gruyter, Berlin, 2015).

[6] G. Pospiech and M. Schöne, Quantenphysik in Schule und Hochschule [Quantum physics at upper secondary and university level], in *Proceedings of the PhyDid B—Beiträge zur DPG Frühjahrstagung* (2012), http://phydid.physik.fu-berlin.de/index.php/%20phydid-b/article/view/392/509.

[7] A. Zeilinger, Quantenexperimente zwischen photon und fulleren [Experiments in quantum physics from photons to fullerenes], Phys. Unserer Zeit **31**, 199 (2000).

[8] J. Küblbeck and R. Müller, *Die Wesenszüge der Quantenphysik: Modelle, Bilder und Experimente [The Traits of Quantum Physics: Models, Pictures and Experiments]* (Aulis-Verl. Deubner, Cologne, 2002).

[9] R. Scholz, S. Wessnigk, and K. A. Weber, A classical to quantum transition via key experiments, Eur. J. Phys. **41** 055304 (2020).

[10] U. S. Di Uccio, A. Colantonio, S. Galano, I. Marzoli, F. Trani, and I. Testa, Design and validation of a two-tier questionnaire on basic aspects in quantum mechanics, Phys. Rev. Phys. Educ. Res. **15**, 010137 (2019).

[11] OECD, *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving* (OECD Publishing, 2017), 10.1787/9789264281820-en.

[12] P. Lautesse, A. Vila Valls, F. Ferlin, J.-L. Héraud, and H. Chabot, Teaching quantum physics in upper secondary school in France, Sci. Educ. **24**, 937 (2015).

[13] A. K. Mohan, Philosophical standpoints of textbooks in quantum mechanics, Sci. Educ. **29**, 549 (2020).

[14] C. Friebe, Physikalisch-mathematische Grundlagen. [Physical and Mathematical Foundations], *Philosophie der Quantenphysik [Philosophy of Quantum Physics]*, edited by C. Friebe, M. Kuhlmann, H. Lyre, P. M. Näger, O. Passon, and M. Stöckler (Springer Spektrum, 2018), pp. 1–40.

[15] W. Heisenberg, *Quantentheorie und Philosophie: Vorlesungen u. Aufsätze [Quantum Theory and Philosophy: Lectures and Essays]*, edited by J. Busche (Reclam, Stuttgart, 1979).

[16] V. S. Bhatta, Critique of wave-particle duality of single photons, J. Gen. Philos. Sci. **52**, 501 (2021).

[17] M. Bunge, Analogy in quantum theory: From insight to nonsense, Br. J. Philos. Sci. **18**, 265 (1968).

[18] E. Marshman and C. Singh, Framework for understanding the patterns of student difficulties in quantum mechanics, Phys. Rev. ST Phys. Educ. Res. **11**, 020119 (2015).

[19] E Marshman and Ch. Singh, Investigating and improving student understanding of quantum mechanics in the context of single photon interference, Phys. Rev. Phys. Educ. Res. **13**, 010117 (2017).

[20] I. M. Greca and O. Freire, Teaching introductory quantum physics and chemistry: Caveats from the history of science and science teaching to the training of modern chemists, Chem. Educ. Res. Pract. **15**, 286 (2014).

[21] M. Waitzmann, K.-A. Weber, S. Wessnigk, and R. Scholz, Key experiment and quantum reasoning, Physics **4**, 1202 (2022).

[22] P. Bitzenbauer and J.-P. Meyn, A new teaching concept on quantum physics in secondary schools, Phys. Educ. **55**, 055031 (2020).

[23] C. Singh, Interactive learning tutorials on quantum mechanics, Am. J. Phys. **76**, 400 (2008).

[24] K. L. McNeill and J. Krajcik, Inquiry and scientific explanations: Helping students use evidence and reasoning, in *Science as Inquiry in the Secondary Setting*, edited by J. Luft, R. Bell, and J. Gess-Newsome (National Science Teachers Association Press, Arlington, VA, 2008), pp. 121–134.

[25] A. A. diSessa, Toward an epistemology of physics, Cognit. Instr. **10**, 105 (1993).

[26] A. A. diSessa, Conceptual change in a microcosm: Comparative learning analysis of a learning event, Hum. Dev. **60**, 1 (2017).

[27] S. Vosniadou, X. Vamvakoussi, and I. Skopeliti, The Framwork Theory Approach to the Progress of conceptual change, in *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, London 2008), pp. 3–34.

[28] S. Goldhaber, S. Pollock, M. Dubson, P. Beale, K. Perkins, M. Sabella, C. Henderson, and C. Singh, Transforming upper-division quantum mechanics: Learning goals and assessment, *presented at PER Conf.* (Ann Arbor, MI, 2009), 10.1063/1.3266699.

[29] S. B. McKagan, K. K. Perkins, and C. E. Wieman, Design and validation of the Quantum Mechanics Conceptual Survey, Phys. Rev. ST Phys. Educ. Res. **6**, 020121 (2010).

[30] S. Wuttiprom, M. D. Sharma, I. D. Johnston, R. Chitaree, and C. Soankwan, Development and use of a conceptual survey in introductory quantum physics, Int. J. Sci. Educ. **31**, 631 (2009).

[31] R. P. Feynman, *QED: The Strange Theory of Light and Matter* (Princeton University Press, Princeton, NJ, 1986), ISBN 0691083886.

[32] V. Mešić, K. Neumann, I. Aviani, E. Hasović, W. J. Boone, N. Erceg, V. Grubelnik, A. Sušac, D. S. Glamočić, M. Karuza, A. Vidak, A. Alihodžić, and R. Repnik, Measuring students' conceptual understanding of wave optics: A Rasch modelling approach, Phys. Rev. Phys. Educ. Res. **15**, 010115 (2019).

[33] *Testtheorie und Fragebogenkonstruktion [Test Theory and Test Design]*, 2nd ed., edited by H. Moosbruger and A. Kelava (Springer, Berlin, Heidelberg, 2008).

[34] M. T. Kane, Validating the interpretations and uses of test scores, J. Educ. Measure. **50**, 1 (2013).

[35] R. Berger, C. Kulgemeyer, and P. Lensing, Ein multiple-choice-test zum konzeptuellen Verständnis der Kraftwirkung auf Ladungsträger in statischen elektrischen und magnetischen Feldern [A multiple-choice-test on the conceptual understanding of the effect of forces in static electric and magnetic fields], Zeitschrift für Didaktik der Naturwissenschaften **25**, 197 (2019).

[36] M. Planinic, W. J. Boone, A. Susac, and L. Ivanjek, Rasch analysis in physics education research: Why measurement matters, Phys. Rev. Phys. Educ. Res. **15**, 020111 (2019).

[37] V. Groth, Analyse von Schülerinterviews zur kognitiven Validierung von Multiple-Choice-Aufgaben sowie zur Identifizierung von Schülervorstellungen zu der Bewegung von Ladungsträgern in elektrischen und magnetischen Feldern. [Analysis of interviews for cognitive validation of multiple-choice items and for identification of student's concepts of the motion of charge carriers in electrical and magnetical fields], Universität Osnabrück Unpublished Masterarbeit, 2016.

[38] X. Fun and S. Sun, Item response theory, in *Handbook of Quantitative Methods for Educational Research*, edited by T. Teo (Sense, Rotterdam, 2013), pp. 45–67.

[39] H. Moosbrugger, Item-response-theories (IRT), in *Test Theory and Test Design*, edited by H. Moosbrugger and A. Kelava (Springer, Berlin, Heidelberg, 2008), pp. 227–293.

[40] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 2nd ed. (Routledge, London, 2012).

[41] K. Neumann, Rasch-Analyse naturwissenschaftsbezogener Leistungstests. [Rasch analysis of science related test instruments], in *Methoden in der naturwissenschaftsdidaktischen Forschung [Research Methods in Science*

Education], edited by D. Krüger, I. Parchmann, and H. Schecker (Springer, Berlin Heidelberg, 2014), pp. 355–370.

[42] W. J. Boone, J. R. Staver, and M. S. Yale, *Rasch Analysis in the Human Sciences* (Springer, New York, 2014).

[43] J. M. Linacre, Sample size and item calibration stability (Nummer 7:4) (1994), https://www.rasch.org/rmt/rmt74m.htm.

[44] P. Mair, R. Hatzinger, M. J. Maier, T. Rusch, and R. Debelack, Package "eRm" (2020), https://cran.r-project.org/web/packages/eRm/eRm.pdf.

[45] W. Revelle, Package "psych" (2020), https://cran.r-project.org/web/packages/psych/psych.pdf.

[46] K. Backhaus, B. Erichson, W. Plinke, and R. Weiber, in *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung [Methods of Multivariate Analysis: An Application-Oriented Introduction]*, 14th ed., edited by H. Moosbrugger and A. Kelava (Springer Gabler, Heidelberg, 2016), 10.1007/978-3-662-46076-4.

[47] H. Moosbrugger and K. Schermelleh-Engel, Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). [Explorative (EFA) and conformatory factor analysis (CFA)], in *Testtheorie und Fragebogenkonstruktion [Test Theory and Test Design]*, edited by H. Moosbrugger and A. Kelava (Springer, Berlin Heidelberg, 2008), pp. 325–343.

[48] I. Neumann, K. Neumann, and R. Nehm, Evaluating instrument quality in science education: Rasch–based analyses of a nature of science test, Int. J. Sci. Educ. **33**, 1373 (2011).

[49] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.20.010122 for the final questionnaire (pool 26) as well as the frequencies of the students' chosen answers in the field testing.

[50] J. Bortz and C. Schuster, *Statistik für Human- und Sozialwissenschaftler [Statistics for Human- and Socialscience] (7th revised and expanded Edition)* (Springer, New York, 2010).

[51] E. Sober, Parsimony arguments in science and philosophy—A test case for naturalism P, Proc. Addresses Am. Philos. Assoc. **83**, 117 (2009), https://www.jstor.org/stable/25656206.

[52] M. Hopf and H. Schecker, Schülervorstellungen zu fortgeschrittenen Themen der Schulphysik [Students' conceptions on advanced topics in school phyiscs] in *Schülervorstellungen und Physikunterricht [Students' Conceptions and Physics]*, edited by H. Schecker, T. Wilhelm, M. Hopf, and R. Duit (Springer, Berlin, Heidelberg, 2018), pp. 225–242.

[53] V. Borish and H. J. Lewandowski, Implementation and goals of quantum optics Experiments in undergraduate instructional labs, Phys. Rev. Phys. Educ. Res. **19**, 010117 (2023).

[54] V. Borish and H. J. Lewandowski, Seeing quantum effects in experiments, Phys. Rev. Phys. Educ. Res. **19**, 020144 (2023).

[55] P. Grangier, G. Roger, and A. Aspect, Experimental evidence for a photon anticorrelation effect on a beam splitter: A new light on single photon interferences, Europhys. Lett. **1**, 173 (1986).

[56] T. C. Pentecost and J. Barbera, Measuring learning gains in chemical education: A comparison of two methods, J. Chem. Educ. **90**, 839 (2013).

[57] *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften [Research Methods and Evaluation in Human- and Socialscience]*, 5th ed., edited by N. Döring and J. Bortz (Springer, Berlin, Heidelberg, 2016), 10.1007/978-3-642-41089-5.