PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH **20,** 020136 (2024)

Design and evaluation of a questionnaire to assess learners' understanding of quantum measurement in different two-state contexts: The context matters

Philipp Bitzenbauer [*]
Universität Leipzig, Institut für Didaktik der Physik, Vor dem Hospitaltore 1, 04103 Leipzig, Germany

Sergej Faletič
University of Ljubljana Faculty of Mathematics and Physics, Jadranska ulica 19, 1000 Ljubljana, Slovenia

Marisa Michelini
Unità di ricerca in didattica della fisica, Dipartimento di scienze matematiche, informatiche e fisiche,
University of Udine, via delle Scienze 206, 33100 Udine, Italy

Kristóf Tóth
Institute of Physics and Astronomy, ELTE Eötvös Loránd University,
Pázmány Péter prom. 1/A, H-1117, Budapest, Hungary

Gesche Pospiech [5]
TU Dresden, Department of Physics, Research Group Physics Education,
Helmholtzstraße 10, 01069 Dresden, Germany

(Received 28 June 2024; accepted 3 October 2024; published 29 October 2024)

The teaching and learning of quantum physics has recently become a topic of increasing interest in physics education research. In particular, the study of two-state systems is gaining importance as a means of teaching quantum physics at various educational levels. Meanwhile, a number of approaches have been developed that are also suitable for high school students. It can be assumed that the different approaches have different degrees of effectiveness in teaching central quantum concepts. However, suitable evaluation instruments to test this are still lacking. Therefore, as a first step, a short questionnaire on quantum measurement, suitable for both research and classroom use, was developed in several steps. First, a questionnaire with open and closed items was created and piloted with a total of $N = 120$ learners. The responses were evaluated qualitatively using a comprehensive coding manual, which provided insights into learners' conceptions. These results led to the development of an eight-item questionnaire that could be adapted to different teaching approaches. This questionnaire was subjected to expert review and, finally, successfully tested for its psychometric properties with a sample of $N = 201$ learners. Overall, our results provide initial empirical evidence that context (i.e., which two-state approach is used) does matter for student learning, but in general, two-state approaches appear to be particularly conducive to learning quantum concepts (specified in this article for quantum measurement) compared to traditional instruction.

DOI: 10.1103/PhysRevPhysEducRes.20.020136

I. INTRODUCTION

The field of quantum physics (QP) has received considerable attention in recent years, largely due to its cultural value and its role in the development of new technologies, which have made it an increasingly central aspect of our society. For this reason, QP has been introduced into secondary school curricula in very different ways [1–7]. However, it requires a completely new way of thinking in contrast to classical physics [8]. In order to fully comprehend the profound theoretical and cultural framework [9], it is essential to analyze students' perceptions of QP [10,11], conduct studies on students' learning difficulties [12], develop a framework for understanding the patterns of students' difficulties [13,14], investigate the dynamic of ontological reasoning [15], and it is of paramount importance to create appropriate tools to measure students'

---

[*] Contact author: philipp.bitzenbauer@uni-leipzig.de

thinking in QP. Alongside this, it has become evident that a comprehensive research program is necessary, commencing with a reconstruction of the foundational concepts for educational purposes [10,14]. Studies on the fundamentals of QP have served as the basis for numerous contributions to teaching and different possible formulations in QP have been utilized as a source of inspiration for didactic proposals [16,17].

The most common approaches hinge precisely on the illustration and argumentation of the birth of QP from a historical point of view [4]. These approaches emphasize the historical and cultural dimensions of QP, underscoring the presentation of the nature of science and the phenomenon of paradigm shift. In accordance with the historical development, they contain a lot of semiclassical models, which may create misconceptions [18–20], and the mathematical description they use is often compatible with the wave function formalism, which makes it difficult to present in secondary schools. Yet the mathematical description plays a conceptual role in QP: mathematics is the origin of the new theory and the interpretation of some aspects is still open today [21–23]. Therefore, if the pedagogical goal is the development of the quantum mechanical way of thinking, the historical approaches are not the best choice.

As a result, several other approaches to teaching QP have been proposed. The new epistemic nature of the theory has prompted conceptual and philosophical proposals [8]. Experiments in optics have been a privileged context for conceptual discussion, orienting the search for specific angles of attack such as interference [24], even with interpretative approaches such as Feynman's paths integrals [25], Mach-Zehnder interferometers [26,27], and simulations of the same [28], as well as beam splitters [29]. At the same time, the need to clarify the main elements that characterize and differentiate QP from classical physics has oriented physics education research to the formulation of didactic proposals focused on single aspects such as the uncertainty principle [30], the concept of state (no longer tied to intrinsic properties of the system, as in classical physics, but rather to the process by which a system is prepared in that state) [31,32], the superposition principle, the corpuscular and wave description of the world, the wave function, the locality, the entanglement [33], and early applications such as cryptography [23].

Among the various approaches, those using two-state systems offer several advantages.

In a two-state quantum mechanical system, all possible states can be expressed by two orthogonal basic states, consequently, the observed quantities can take only two values [34–38]. Well-known examples are polarization and spin. If spin is measured along an arbitrary direction, the resulting value will be either $+\hbar/2$ or $-\hbar/2$. There is no other possible outcome. If polarization is measured with a polarizer, the result will be that a photon either passes through the polarizer or is absorbed by the polarizer. Again, there are only two possible outcomes.

Mathematics plays a conceptual role in QP. However, the mathematics of general QP can exceed the knowledge of secondary school students by far. As an example take quantum states described by abstract vectors in N-dimensional Hilbert space, where N can go to infinity for some physical quantities. A superposition of such vectors is a sum, which becomes an integral in case of $N \to \infty$. On the other hand, two-state systems only use a two-dimensional vector space. Operations with it are well known in secondary school. So if we want to focus on the central concepts of QP, and the development of a way of thinking, an introduction to the foundations of mathematical formalism is beneficial, if not necessary, and two-state systems offer this opportunity. Moreover, two-state learning paths fit well with the expectations of the second quantum revolution [39], since a two-state quantum system can be a quantum bit.

Two-state systems are very convenient for introducing the concept of state, the superposition principle, the genuinely stochastic nature of measurement results, the probabilistic nature of predictions, incompatibility of certain quantities (often called uncertainty relations), and the effect of measurement in QP. Some approaches add time evolution and/or entanglement to these concepts. Various applets and simulations have been developed, often specifically associated with a path [40–42]. There have been numerous implementations of didactic paths based on two states with didactic intervention modules in limited time, usually 10–13 h [6,20,25] and from a minimum of 6–8 h [19,41,43], to 30 h and more [44,45]. All of the literature on approaches include some kind of evaluation of the approach, but the evaluations use different instruments and focus on different aspects which will be discussed in more detail in Sec. III.

While prior contributions had a more general approach to basic quantum concepts (e.g., see Ref. [46]), we put special attention to the measurement process and the effect of measurement in QP in this article. According to the research by Merzel *et al.* [47], measurement is the only topic that is seen as very important by all professional groups dealing with teaching QP, so it is a natural choice as the focus of our research. However, questionnaires on students' perceptions of quantum measurement are usually specific to a narrow group and teaching approach, making it difficult to compare different learning pathways. In this paper, we present a questionnaire we developed to compare students' perceptions of quantum measurement at different two-state approaches. In the context of our research, measurement is understood as any process that results in the final state being an eigenstate of the operator which represents the measured quantity. The precipitation of a general (superposition) state into an eigenstate is a completely stochastic process, which cannot be predicted with certainty. The probability distribution between the possible resulting eigenstates is determined by the initial state of the system. If this state is written as a superposition of eigenstates of the operator which represents the measured quantity, the squared magnitudes of the coefficients in the superposition determine the probabilities for each resulting corresponding eigenstate.

In the following, we will first provide in Sec. II our research objectives and rationale. Then, in Sec. III, we will provide some background on the research into two-state approaches so far. Here we will also provide an overview of each of the approaches involved in our study, and some research done on these approaches so far. In Sec. IV, we discuss the design of the instrument and all considerations taken into account in the development. Section V provides a description of the first pilot implementation of the questionnaire from which we redesigned the questionnaire into a multiple-choice-only format. In Sec. VI, we provide an implementation of the questionnaire with experts to test the validity of the questionnaire. In Sec. VII, we discuss the results of the implementation of the final questionnaire in the three courses described in Sec. III. The discussion of the results follows in Sec. VIII. We end the article with some conclusions in Sec. X.

II. RESEARCH OBJECTIVES

The many various two-state approaches, some of which will be outlined in Sec. III, cover many or most of the basic topics of introductory two-state QP: nondeterminism, complementarity, incompatibility, superposition, and the effect of measurement. However, each covers the topics with different teaching methods, different logical progression, and in different contexts. As will be discussed in Sec. III B, each approach demonstrates gains in students' understanding, but they each use a different instrument to measure these gains [18–20,41,43,48–51]. It is reasonable to expect that context and teaching methods matter when learning physics. The question then arises, do these approaches provide the same gains? Do these approaches provide uniform gains across all topics or do some approaches achieve better gains on some topics and other approaches on other topics?

To investigate these questions, a transferable and versatile instrument, kind of a quantum concept inventory, is needed. The purpose of developing such an instrument is to enable comparison between various two-state approaches; specifically, comparison of conceptual understanding developed in different approaches. In this context, we consider conceptual understanding to be understood as "knowledge of concepts and linking relationships that are directly connected to (or logically necessitated by) the definition of a concept or meaning of a statement" [52] (p. 2) and in this paper, we focus on developing such an instrument focusing on students' understanding of quantum measurement. We envision the use of this instrument in any two-state setting to gauge the development of conceptual understanding even when the approach itself is more geared toward a technology-oriented application.

Hence, the research objectives of this paper are threefold:

1. We develop a new instrument that allows to assess secondary school students' conceptual understanding of introductory aspects of quantum measurement.
2. We provide a psychometric characterization of the instrument in the sense of classical test theory.
3. Finally, bringing together insights from qualitative and quantitative pilot studies, as well as an expert survey, leads to an evidence-based argument for a valid test score interpretation.

III. RESEARCH BACKGROUND

Especially at the university level, there have been many studies on learning the key concepts of quantum theory, both in chemistry and in physics [14,31,53,54]. Also at the secondary level, numerous two-state approaches have been developed [20,23,41,55–60]. Tutorials and pre-post testing accompanied the numerous implementations of the paths, mainly in the context of optical polarization [61–63]. These were useful for analyzing students' learning processes, also with regard to the distinction between mutually exclusive and incompatible properties, the uncertainty principle, the nonlocal nature, and the impossibility of a trajectory for quantum objects [6]. However, there is a lack of general tests or conceptual inventories, which analyze learning regardless of the path, as has been done at the university level [56]. Only sporadic studies have been conducted without reference to a specific teaching and learning approach [64]. The choice of fundamental aspects to deal with also guides learning studies to understand how students acquire some aspects or concepts and not others. For example, some proposals, while introducing the concept of state, do not stress the fundamental role of the principle of superposition; most teaching proposals do not insist on the distinction between states and properties, which is crucial to distinguish between the classical and the quantum vision of phenomena. Even the issue of measurement in QP and the associated indeterminism, although addressed by all, is treated in a differentiated way, e.g., focusing on the probabilistic nature; stressing the genuinely stochastic, i.e., nonepistemic, nature of quantum indeterminism [14]. An interesting study concerned how students developed quantum conceptions as opposed to classical ones using a phenomenographic approach [24].

More recently, especially at university level, inventories have been administered [65,66]. It should be noted that very often the questions asked to students are related to the context and the questions formulated without reference to a defined phenomenology are expressed with a specific formalism and specific terminologies (e.g., pure state, eigenstate, eigenvalue), not dealt with at secondary school level. The study of the role of formalism was central to many studies at university level [6]. The role of real experiments for learning fundamental quantum concepts is one of the explored aspects, given the obvious difficulties in carrying out actual experiments in teaching laboratories [3,43]. The perspectives taken to examine the learning processes are therefore very contextual. However, it emerges that in secondary school some concepts such as the superposition state, measurement, and entanglement are quite difficult. In addition, in research on learning quantum concepts at the secondary school level, little information has been gained about how students learn the concepts of entanglement, nonlocality, and measurement [5].

We expect that learning of these fundamental concepts could be influenced by the way in which they are taught: the context, the logical progression of the learning sequence, and the pedagogical methods used. Therefore, we describe here in more detail the three two-state approaches that have been used with students in this study.

A. A brief overview of two-state approaches for teaching and learning quantum mechanics

In this section, we provide an overview of the frequently used two-state approaches with particular emphasis on the three approaches used to evaluate the questionnaire presented in this article.

The five main two-state paths are designed in the context of which-path-encoded photonic states [43,48,67], optical polarization [6,19,20,55,56,68], double-well potentials [41], qubits [23,58], spin [57,65]; and abstract two-state approach, realized via games [59,60].

We tested the questionnaire on three approaches belonging to the first three of the above categories. The approaches have been selected according to the following criteria: (a) the authors have direct experience with the approaches: their design, rationale, implementation, and evaluation, and (b) the approaches have been investigated and evaluated in prior research. These approaches are as follows: which-path encoded photonic states, optical polarization approach, and double-well approach. These approaches are described in more detail in the following subsections. These descriptions are kept as short as possible while still presenting the important nuances of the approaches. These nuances are crucial to understand the differences in the results of different cohorts on the questionnaire.

1. Approach 1: Which-path encoded single-photon approach (abbreviated BS)

Single-photon experiments have been described to "provide the simplest method to date for demonstrating the essential mystery of QP" [69] (p. 471). In the single-photon approach (in more detail described in earlier works, e.g., Refs. [43,70–72]), we consider the scenario in which a photon is sent onto a beam splitter, and the question is posed as to whether the photon is transmitted or reflected. The input state $|S\rangle$ represents the single-photon state emitted from a single-photon source, which is entering one input port of the beam splitter and zero photons entering the other input port. The two possible output states correspond to either a single photon reflected and zero photons transmitted (state $|R\rangle$), or a single photon transmitted and zero photons reflected (state $|T\rangle$). These output states are eigenstates of the number operator $\hat{n}$. The crucial point is that these two output states can be used as an orthonormal basis, which allows us to consider the whole system as a two-state quantum system. The teaching-learning sequence follows the following steps:

Step 1: Supported by interactive screen experiments developed by Bronner *et al.* [73], students examine single-photon states interacting with a $50{:}50$ beam splitter. The absence of coincident events at the beam splitter's output ports [74] is incompatible with classical light theories, as "a single photon can only be detected once" [75] (p. 173).

Step 2: In the experiment described in step 1, the single-photon state is initially converted into a superposition state $\frac{1}{\sqrt{2}}(|T\rangle + |R\rangle)$ through the $50{:}50$ beam splitter and is then detected at one of the detectors with equal probabilities of 50%. Consequently, quantum randomness emerges from the measurement performed on a quantum superposition state [76]. A reduced quantum formalism using Dirac notation presented in Ref. [72] is introduced as a mathematical representation of the phenomenon.

Step 3: The subsequent phase of the teaching-learning sequence is centered on the 1986 experiment by Grangier *et al.* [75]. The students observe that a single-photon state leads to both anticorrelation at a $50{:}50$ beam splitter and single-photon interference in a separate interferometer, such as a Michelson interferometer, through the use of interactive screen experiments by Bronner *et al.* [73] and GeoGebra simulations by Hennig *et al.* [72]. This process enables students to identify that the "quantum interference phenomenon shown experimentally is a consequence of the interplay of superposition and nonlocality" [77] (p. 17), which challenges the notion of photons as localizable (classical) particles. Instead, students can gain an understanding of photons as elementary energy portions of light [78] in accordance with the principles of quantum electrodynamics (e.g., see Refs. [79,80]).

Step 4: The Michelson interferometer is extended to a quantum eraser setup. In the accompanying interactive experiment [73], students can operate two polarizers and set them to the desired setting. Students observe the following: If both polarization filters are vertical or both are horizontal, the paths are not marked and single-photon interference can be observed. However, as soon as the two polarization filters are set at 90° to each other, optical paths are marked and can be distinguished. The interference pattern disappears because the interaction of a single-photon state with a polarization filter represents a quantum measurement process that leads to a projection onto an eigenstate. The question that will guide the remainder of the lesson is as follows: Can the path information be erased again after the interferometer? The answer is yes, as evidenced by the students' own interactive screen experiment [73]. A polarization filter set to 45° after the interferometer represents another measurement process in the so-called ×-basis, which restores a superposition in the ± basis and so interference is also restored.

2. Approach 2: The optical polarization approach

In the polarization approach, students explore the basic principles of QP through the linear polarization states of photons. The reader is referred to the articles [19,49,68] for a more detailed description of this learning path that uses the inquiry-based learning method. In this approach, only linear polarization is used to avoid complex numbers. Before beginning the QP studies, students interpret light polarization in the context of classical theory of light where they observe the behavior of ideal polarizers and birefringent calcite crystals through a series of experiments. Students arrive at Malus' law which gives the normalized light intensity (I/I_0) as a function of the angle between the polarization of light and the polarizer (θ), $I/I_0 = \cos^2\theta$. After the phenomenology, students accept that light consists of indivisible and indistinguishable photons and try to interpret experiments in the single-photon case via the linear polarization of photons.

Step 1: Students see that the Malus law violates the indivisibility of photons when a single photon is emitted onto a polarizer. Students see that the scope of the Malus law does not extend to the level of a single photon and reinterpret it as the probability of a photon passing through a polarizer. Students then practice statistical predictions in a computer program JQM (Java Quantum Mechanics [81]), which is an open environment for hypothesis exploration in which it is possible to generate the desired number of polarized photons or a not polarized beam that encounters one or more polarizers with different permitted direction or birefringent crystals and is detected by a counter. JQM enables students to analyze the statistical behavior of photons without learning the deeper mathematics required for statistics.

Step 2: In the next step students are introduced to new concepts, "(polarization) property" which is the conceptual analog of the "values of a quantity" (for a deeper explanation, we refer the reader to the papers [19]). The polarization properties of photons are represented by icons distinguishing them from the polarization state vectors. Students can understand that the horizontal and vertical polarization properties are "mutually exclusive properties" because if a photon certainly has one property, it certainly does not have the other. An observation always determines "mutually exclusive properties." The students then discover the existence of "incompatible properties," which we call the "uncertainty principle": a diagonally polarized photon cannot be said with certainty to pass through a polarizer with a horizontal permitted direction, because this would violate the probabilistic behavior of photons (a more detailed explanation of the uncertainty principle in the polarization approach can be reached via Refs. [82,83]). Students will also discover that measurement plays an active role in QP because it can change the polarization of photons so that measurements do not commute.

Step 3: Students reflect on the meaning of probability, giving them an insight into how quantum probability is different from what they have learned before. Two classical hidden variable hypotheses are formulated and refuted by thought experiments (see Ref. [19]). In the ideal quantum measurement, if the state of the photons is known, then the teacher identifies the superposition as the cause of the probability, where the superposition is interpreted in a qualitative way. After this, students are confronted with the "lack of trajectory" using birefringent calcite crystals. This is because a calcite crystal creates an entanglement between the polarization and spatial position of a single photon, which can be uncertain until the moment of detection.

Step 4: Students are introduced to the quantum states via two-dimensional vectors of the plane which corresponds to the polarization directions. These states are represented on a Bloch circle [19,23,84]. Students then recognize that the probability of a photon's transition from a polarization state $\boldsymbol{a}$ to another state $\boldsymbol{b}$ is equal to the square of the scalar product of the state vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ [which, according to the Malus law, gives $(\boldsymbol{a}\cdot\boldsymbol{b}) = \cos^2\theta$]. It is then possible to formulate the superposition as follows: not only are the measurable states possible states but also any linear combination of them (e.g., the diagonal polarization state is written as $\boldsymbol{d} = (1/\sqrt{2})(\boldsymbol{h} + \boldsymbol{v})$. Distinction between state (vector) and properties (icons) that live in different spaces reinforces the meaning of the measurement in QP as a transition to a new state: the precipitation of the system in the measured one and its genuinely stochastic nature of measurement.

Step 5: The projectors of the measurements can be written in outer product form so that the operator representing the measurement can be constructed, whose eigenvalue equation can be used to describe the measurement. These are explained in detail in Ref. [68].

3. Approach 3: The double-well approach

In the double-well approach [41], the context is an idealized double well explored with a simulation. The quantities under investigation are position (x) and energy (E). These quantities are incompatible in a square double well. The possible properties are $x = \mathrm{L}$ and $x = \mathrm{R}$, denoting left and right well, respectively, and $E = \mathrm{E}_1$ and $E = \mathrm{E}_2$, denoting the ground and first excited energy state, respectively. Of course, this is a pseudo-two-state system as the position eigenstates are not really limited to the two wells and the energy eigenstates are not really limited to the first two eigenstates.

Step 1: Students start with observations of a particle prepared in eigenstate $|\mathrm{L}\rangle$. Measurements on subsequent particles prepared all in the same way give the property $x = \mathrm{L}$ in 100% of cases. This is consistent with classical physics. Students explore energy eigenstates in the same way.

Step 2: Students are tasked with finding out whether a position eigenstate is also an energy eigenstate. They find that it is not and superposition states are introduced. Through carefully chosen questions, students further observe that when measuring a superposition state, the appearance of each property is stochastic, leading them to conclude that only the probability for an outcome can be predicted.

Step 3: Students are presented with an observational experiment which shows that a measurement in between two measurements can change the outcome of the second measurement. This activity proceeds according to the Investigative Science Learning Environment (ISLE) framework [85]. Students propose explanations for this phenomenon. The explanation that the measurement changes the state of the system into an eigenstate is always among the proposed explanations. The teacher proposes testing experiments (to maintain the cognitive load on the findings), students predict the outcomes of the testing experiment based on their explanations, the testing experiment is performed (with the simulation), and students reject explanations whose predictions do not match with the experimental outcome. The only explanation that cannot be rejected is that the measurement changes the state of the system into an eigenstate of the measured quantity. Hence, students have to accept this explanation. Any other explanations that they might have had did not pass the testing experiment.

Step 4: Students are asked how they interpret a superposition state. The interpretation that superposition is a statistical mixture always emerges. Following especially the testing steps of the ISLE framework outlined above [85], students discover that the outcome of the testing experiment does not match any of their predictions based on any interpretation that equates superposition with a statistical mixture. Hence, students have to reject this idea. A vector interpretation of superposition is then provided as an alternative.

Step 5: Students do activities related to the use of quantum concepts, usually quantum cryptography in the form of a game.

B. Empirical evidence on the learning effectiveness of two-state approaches

1. Approach 1: Which-path encoded single-photon approach

In a mixed-methods field study [50], $N = 173$ high school students' learning of QP was evaluated using the single-photon approach as described in Ref. [70]. Through the collection of both qualitative and quantitative data, the authors found that the study participants achieved an adequate conceptual understanding of quantum optics and built up a mostly adequate understanding of the essential features of QP while at the same time finding that students struggled to detach from classical particle conceptions of the photon [48]. Similar results have been obtained from the evaluation of an extracurricular online course that also uses the single-photon approach with younger students [18]. In a cluster-randomized study with a pre-post-test design, students participating in the single-photon approach were found to outperform their peers who had participated in a traditional quantum course following the historical development of quantum theory, both on aspects related to (a) quantum objects' properties and behavior (Cohen's $d = 0.27$) and (b) the probability interpretation of QP (Cohen's $d = 0.38$) [43]. To the best of our knowledge, no published work to date has undertaken a comprehensive examination of student learning about quantum measurement using the single-photon approach.

2. Approach 2: The optical polarization approach

Several research units implemented from 1998 to 2024 involved about 1500 students in about 50 classes involving secondary school teachers and physics education researchers in 4 countries. During this period, a professional curriculum was created using the Design-Based research method [19,20,49,86] Research has found that students fall into three independent ways of thinking about QP: classical, hidden variable, and QP way of thinking [20]. The classical way of thinking indicates that the students believe that quantum objects have well-defined properties (e.g., trajectories) and their behavior is not probabilistic. We say students have a hidden variable way of thinking if they believe that microscopic systems preserve some properties of classical macroscopic systems in theory, even if they are not knowable or detectable in real experiments (e.g., students believe in the temporal and spatial continuous description, but the motion is not accessible due to experimental limitations; because the measurement instrument uncontrollable disturbs the system [19,20,51], it is found that the Dirac polarization approach is an appropriate way to develop students way of thinking [20] while a large proportion of students from the age of 15 are able to master a significant proportion of basic concepts and laws. Research has also shown that, as with other approaches, students find it very difficult to let go of the classical image of movement and try to describe phenomena in a continuous spatial and temporal way [19,49].

3. Approach 3: The double-well approach

The double-well approach has been studied in the context of the effectiveness of student engagement with the ISLE process [85], which consists of building knowledge mimicking the process used by scientists: observing phenomena, proposing explanations or models, testing these explanations or models with testing experiments by predicting the outcomes using hypothetico-deductive reasoning and then passing judgment about the proposed explanations or models based on the comparison between the predictions and the actual outcomes of the testing experiments. In the study [41], students in small groups engaged in ISLE-based activities, which should lead the students to a specific conclusion about the behavior of the quantum world. In the study, it was measured how many groups of students arrived at the desired conclusions by engaging only with the activities without any instructor's intervention. For activities on incompatibility (step 2), the success rate was above 90%. For activities on probability and indeterminism (also step 2), the success rate was above 75%. For activities on the difference between a superposition and a statistical mixture (step 4), the success rate was above 60%. Activities on the role of measurement proved the most difficult. More than 60% of the groups were able to conclude that a measurement changes the state of the system, and more than 70% of the groups were able to conclude that the new state of an ensemble prepared in a superposition state is a statistical mixture of eigenstates. Despite this, only about 30% were able to identify that the measurement transforms a superposition state into the eigenstate of the measured quantity corresponding to the measured value. Since the study, this activity has been continuously improved based on the ISLE process to arrive at the current step 3 of the double-well approach.

The results of the described study indicate that students are able to arrive at many crucial conclusions about the behavior of the quantum world on their own. However, the exact effect of measurement seems to be the most difficult conclusion to make.

IV. DESIGN OF INSTRUMENT

We set out to design an instrument for the quantum measurement process. This topic seemed to be the most crucial one in the presented approaches. In designing the instrument, we followed the steps presented in Fig. 1. To make the instrument easily implementable, versatile, and have a valid test score interpretation, several considerations had to be taken into account.

A. Validity considerations

To arrive at an evidence-based argument for a valid test score interpretation, our study is based on the validity concept by Messick [87] as we formulate an intended

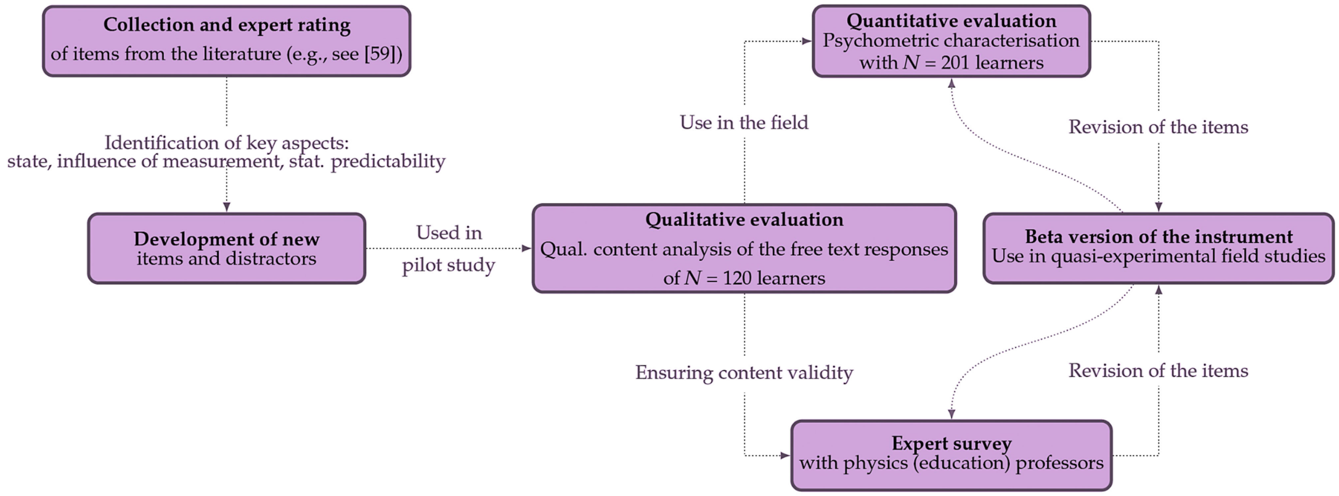

FIG. 1. The diagram illustrates the cyclical nature of the revision process through the use of curved gray arrows: The instrument was subjected to three different evaluation studies during the development process, including a qualitative evaluation to gain insight into students' reasoning about quantum measurement (see Sec. V), an expert survey to gain insights into the item quality from the perspective of physics (education) professors (see Sec. VI) and a quantitative evaluation aimed at a psychometric characterization of the final instrument (see Sec. VII). Herewith, it is important to note that the revision of a test instrument is an ongoing and iterative process.

test score interpretation as well as assumptions this interpretation is based on in the following (see, for example, Refs. [88,89]): We intend to interpret the test score as a measure of secondary school students' conceptual understanding of introductory aspects of quantum measurement. The development of the instrument presented in this paper was guided by the definition of conceptual understanding by Melhuish as "knowledge of concepts and linking relationships that are directly connected to (or logically necessitated by) the definition of a concept or meaning of a statement" [52] (p. 2). The assumptions this interpretation is based on are as follows:

A1: The items adequately represent the construct conceptual understanding of introductory aspects of quantum measurement (empirically substantiated in this study via expert discussion and expert survey, see Secs. IV D and VI).

A2: The items are unambiguous and the instructions are clear from physical and didactical points of view (empirically substantiated in this study via expert discussion and expert survey, see Secs. IV D and VI, as well as a qualitative pilot study, see Sec. V).

A3: The items and distractors are authentic and useful for the use in the target sample (empirically substantiated in this study via a psychometric characterization, see Sec. VII).

A4: The items are instructionally sensitive (empirically substantiated in this study via analysis of the item response distributions, see Sec. VII).

B. Target group and usability

The principal objective is to develop an assessment instrument that may be utilized in a variety of settings and contexts. It is therefore desirable to address as many different target groups as possible, with the widest possible range of knowledge and skills. Accordingly, the questionnaire must be suitable for use with both high school and university students. In order to achieve this objective, a set of criteria was established for the design process. First and foremost, the instrument must be concise (D1), with the potential for incorporation into a comprehensive quantum concept inventory. Nevertheless, it should encompass all the essential elements of quantum measurement (D2). These are outlined in Sec. IV C. Furthermore, the instrument should be equally applicable in the various contexts of two-state system approaches (D3). However, it is important to consider that high school students may not yet possess the same mathematical knowledge base as university students. Consequently, the mathematical structures and elements must be described verbally (D4, D5). Additionally, if the objective is to assess the efficacy of a teaching unit, the instrument allows for administration as a pretest and a post-test, which is also supported by its brevity. The following design principles D1 to D5 were identified:

D1: The instrument must be implementable within one class session.

D2: The items should cover all important aspects of quantum measurement.

D3: The items are either independent of different two-state contexts or can be adapted. Herewith, it is important to acknowledge that there are instances where it is not feasible to achieve a complete isomorphism between question formulations and possible answer options in different contexts. However, it is essential to strive for as much isomorphism as possible between the items in different contexts, with the overarching goal of ensuring that the same concepts are covered in the different distractors (see Table I).

TABLE I. The table shows an example of adaptation of an item between different contexts, in this case, the optical polarization approach and the double-well approach. The words and phrases that maintain the same meaning between the contexts are marked in bold.

Approach	Item	Answer options
Optical polarization	A measurement apparatus consists of **two devices arranged in series** for the measurement of orthogonally polarized states ($\|H\rangle$ (horizontal) and $\|V\rangle$ (vertical)). A single photon in **the equally weighted superposition state** of $\|H\rangle$ and $\|V\rangle$ enters the apparatus. What can be predicted?	(a) When the single photon **passes through the first device**, it will **certainly be absorbed by the second device.** (b) **The probability** that the photon passes through **both devices is 25%**. (c) **The probability** that the photon passes through **both devices is 50%**.
Double well	A measurement apparatus measures the position of an electron (L or R) **twice in immediate succession**. A single electron in the **equally weighted superposition state** of $\|L\rangle$ and $\|R\rangle$ enters the apparatus. What can be predicted?	(a) If the **first measurement shows the value L**, the **second one will certainly show value R** and vice versa. (b) **The probability** that we measure value L for the electron **both times is 25%**. (c) **The probability** that we measure value L for the electron **both times is 50%**.

D4: The formulation of the questions should take into account that different courses use different mathematical representations or none at all. So all questions should be phrased independently of mathematical representations.

D5: The items should be limited to time-independent phenomena because most preuniversity courses do not address time evolution.

C. Description of knowledge domain

This section delineates the fundamental elements of the content domain encompassed by the questionnaire, namely the quantum physical measurement process, from a scientific perspective. The questionnaire itself will be designed to be adaptable to the different target groups. First a few basic remarks about measurement in QP: The primary distinction between classical and QP is rooted in the mathematical-physical property of superposition, which is associated with the linearity of the QP world description. Consequently, the difficulties in interpreting QP arise from the transition from quantum superposition to the unambiguousness of the classical world. It appears that the pivotal point is as follows: the linear structure of QP allows for the existence of superposition and entanglement states. These develop deterministically in accordance with the Schrödinger equation, with no ambiguities. However, upon extracting information from the quantum system in question, an inconsistency arises. In a measurement, the quantum superposition must yield unambiguous results that are compatible with the well definedness of classical physics. This fundamental incompatibility thus constitutes the core of the interpretational debate (see, for example, Ref. [90]). The nature of these discrepancies is contingent upon the interpretation of QP. For instance, it manifests differently in the many-worlds interpretation (which posits the absence of measurements) or in the ensemble interpretation (assuming there is no problem with measurements). The resulting disparate solution approaches are not inherently true or false, as the transition between the quantum and classical regimes is not amenable to mathematical formalism. In this study, we adopt a viewpoint that is compatible with the ensemble interpretation (e.g, see Ref. [91]). However, this does not necessarily imply that the questionnaire can only be applied in the light of that interpretation.

Starting from the superposition principle, the mathematical description entails the following central properties P1 to P4 of the quantum measurement process in particular:

P1: Nonpredictability (in general) of single measurement results, whose nature has clearly to be distinguished from the case of nonpredictability in deterministic chaos in classical physics.

P2: Predictability and calculability of probability, as expressed through the scalar product in Hilbert space. In the case of polarization as a model system, the law of Malus can be used.

P3: Predictability of outcomes in the case of repeated measurements of the same observable (neglecting time evolution). From the mathematical description, this reflects the fact that the effect of measurement can be described by a projection on a basis state of the system in relation to the measurement device. From the physics standpoint, this implies that repeating the identical measurement on the same quantum object (in the state resulting from the initial measurement) must yield the same (and therefore predictable) outcome. This concept is relatively straightforward to comprehend in the context of stationary states in two-state systems. However, it can be challenging to reconcile in the case of continuous variables or time evolution, particularly in the context of measurements of position or momentum.

P4: Active role of the measurement, which generally changes the state of the quantum object, is in contrast to the role of measuring in classical physics, where the state of the system is only recorded but not changed.

Besides these central properties of the quantum measurement, fundamental differences in the role of probability in QP compared to that in classical physics, related to the general characteristics of QP, have to be considered. These properties constitute the core of the quantum measurement process, which will be reflected in the various components of the instrument (see Design Consideration D2 in Sec. II). With regard to the instrument under investigation, we have opted to group the items pertaining to the aforementioned content aspects into three groups, which we have also designated as content domains or content areas in the following:

1. “Statistical predictability” addresses the central properties P1, P2, P3.
2. “Knowledge of state” encompasses items that address the central property P3.
3. “Effect of measurement” includes items that address both P3 and P4.

D. Development of items

The development of the items was informed by a review of the relevant literature, the selection of items from existing questionnaires (https://www.physport.org/assessments/), and the author’s own teaching experiences, which provided insights into the typical questions or misunderstandings of students. A review of the literature revealed that the majority of assessments are designed for upper-division QP courses (see Refs. [65,92]). Some questionnaires are designed for high school students, yet do not explicitly address the measurement process [24]. The majority of questionnaires focus on the role of wave-particle dualism and lack specificity regarding two-state systems. From the outset, it was hypothesized that the various contexts described above could be associated with characteristic learning difficulties of students. This hypothesis gave rise to the intention to evaluate approaches to two-state systems. Despite the realization that the items pertaining to the measurement process could not be entirely separated from an understanding of uncertainty and superposition, the characteristics of the measurement process remained the primary focus and guided the selection of items. The collected items were graded and discussed in an expert discussion. Given the wide variety of the intended target groups defined above, the items and answer options were formulated in a way that they should be adaptable to different contexts and avoid mathematical terms such as “eigenstates” or “non-commutative operators.” Instead, a verbal description was chosen. As with this questionnaire, new ground was broken, and an exploratory approach seemed sensible. Therefore, a two-tier questionnaire was to be used. First, single-choice items with closed answer options were drafted. The distractors were formulated on the basis of a literature review, experiences with other questionnaires, and the author’s own teaching experiences. In order to gain a deeper insight into the students’ views, they were encouraged to provide a justification for their choice of answer option in an additional “explain” section of the items. Moreover, the pilot version of the questionnaire included open-ended items that covered all aspects described in the knowledge domain, thus providing additional insights that complemented those obtained from the closed items.

V. STUDY I: QUALITATIVE EVALUATION

The instrument of the pilot study (study I) consisted of nine closed and four open items. The primary objective of this stage was to enhance the instrument’s efficacy. Insights into students’ cognitive processes were gleaned from two sources: the responses to the open-ended items and the justifications provided for the selected answer options on the closed items (i.e., the justifications). Both response types were subjected to qualitative content analysis. The justifications offered by the students subsequently served as a reference for the refinement of the distractors in the final iteration of the questionnaire. This process is described in Sec. V D. The final version is discussed in Sec. VII.

A. Test administration and sample

In the initial phase of the study, the pilot questionnaire was tested by advanced students to gain an initial understanding of its functionality. Based on their feedback, a few minor adjustments were made. The pilot study involved prospective physics teachers and high school students. A total of 120 students from Germany, Hungary, Italy, and Slovenia were included in the sample, representing diverse backgrounds which allows to gain insight into a broad range of potential reasoning and ideas:

- 67 high school students (16- to 18-year-old) who had participated in a polarization approach before. All students attended standard physics class in high school, presenting a typical sample of students.
- 23 prospective teachers who have completed approximately half of their study program and have attended a traditional quantum theory lecture.
- 30 high school students (17-year-old) who attended compulsory physics classes, but also one voluntary physics lesson per week in grade 12, used to implement the double-well approach.

The high school students had the usual knowledge in mathematics, without calculus.

B. Data analysis

The data analysis was conducted in a series of stages. In the initial phase of the analysis, each item, the corresponding answer option, and the justification were co-coded, thereby inductively generating categories. This extensive and comprehensive category system was subsequently analyzed, and overarching categories were formed in several steps. During this process, a coding manual was developed that served for the coding of both the justifications and the answers to the open items.

The most pivotal stage of the process involved a 2-day intensive discussion workshop between experts, during which the developed coding manual was subjected to rigorous scrutiny and tested on randomly selected questionnaires. In the subsequent discussion, the coding manual was revised. To ensure the reliability of the categories and the coding process, again the responses from randomly selected questionnaires were independently coded by multiple intercoders. The codings were then compared, after which an intensive and detailed discussion was held between the intercoders. This led to adjustments to the categories and a clarification of the coding manual.

From this point on, the descriptive categories were condensed into a category system, consisting of 7 main categories and a total of 32 subcategories. These were supplemented by nine additional subcategories allowing for a detailed description of the students’ reasoning. After coding, it was observed that two of the main categories (with overall eight subcategories) only occurred in the responses to the open-ended questions. This can be explained by the fact that in the open-ended items, the respondents were explicitly asked, e.g., to indicate differences between classical physics and QP. Consequently, these specific categories are not considered in the evaluation of the justifications. In total, the justifications were coded with 5 main categories and the corresponding 24 subcategories. The results of this coding procedure are presented in the next section.

C. Results

For analyzing the justifications, 5 main categories with 24 subcategories were utilized. The main categories are:

- Predictability (PM),
- Role of probability for determining measurement results (MR),
- Possibilities for knowledge of quantum states (KS),
- Effects of measurement on quantum objects (EM), and
- General characteristics of quantum physics (GQ).

The subcategories included codings for justifications that could be attributed to quantum thinking and those that were deemed to be representative of classical thinking. To illustrate, we present an example from the main category, "Effects of Measurement." The category EM2, "So that only one result (eigenvalue/property) is possible," was deemed to be a quantum phenomenon, whereas the category EM6, "Disturbs the quantum object," was identified as classical thinking. In this manner, the subcategories could be characterized as representing either a quantum, a partly quantum, or a mostly classical justification. In some cases, it was not possible to determine with certainty whether the students held a classical or a quantum view. This was particularly evident in statements such as "Measurement changes the state," which were frequently encountered. In such instances, the statement was assigned to a category based on the context. In some cases, the context indicated that the statement was intended to convey the meaning of "disturbing," which would imply a classical perspective. Conversely, in other instances, the statement appeared to be aligned with the quantum perspective. However, in some instances, the underlying perspective remained ambiguous. These cases prompted a revision of the coding manual and a more detailed explanation of the guidelines. Consequently, they include hints regarding the appropriate category to use in a given situation, or alternatively, the category that should be avoided. Following this significant restructuring of the coding manual, an independent coding of randomly selected cases was conducted by several intercoders to ensure reliability.

1. Coding manual

In Table II, to illustrate the structure of the category "Predictability: Prediction of Results of Single Measurements," we cite an example from the coding manual. The final version of the coding manual can be found in the Supplemental Material [93].

2. Measurement—Quantum to classical opinions

The quantum measurement is at the heart of the difference between quantum and classical physics. The difficulties associated with understanding these differences are reflected in the students' statements. From these, it can be inferred that they have difficulties accepting the quantum behavior of states, as evidenced by the following quote: "To be able to determine the quantity very exactly, I must measure it infinitely exactly." (LA10l) or they think of the precision of measuring devices: "Our inability to predict certain outcomes is given first of all by the limitations of measuring devices" (UdA). This notion, which is more or less classical, can also be observed in the perception of

TABLE II. This excerpt from the coding manual for the category *Predictability* (PM) provides an example for a subcategory indicating quantum thinking (PM-3), a subcategory indicating classical thinking (PM-5), and a subcategory indicating mixed thinking (PM-1).

Sub category	Name of subcategory	Anchor example	Description
PM-1	Prediction not possible	"At first, I cannot predict the result for a single measurement" (TA(M)05H)	This category is coded when students indicate that the outcomes of single measurements cannot be predicted in general, not even in special cases. This is different from PM-3.
PM-3	Prediction possible (only) for eigenstates	"To get know the outcome of a measurement P (polarization), then we have to know the state of system which has to be a special state." (R 17)	This category is coded if the students state that in case of a measurement being performed on a quantum system in an eigenstate or if preparation and measurement are done in the same way, the outcome can be predicted. If probability is mentioned, then PM-4 is coded.
PM-5	Single measurement outcomes can be predicted with probability 1.	"However, an infinitely accurate measuring equipment could set this probability equal to 1." (LA10I)	This coding is used if the students state that (a)—in general—the outcomes of a measurement may only be predicted by means of probability statements due to insufficiently precise devices (differs from PM-4) or (b) if they judge whether (and how) the result might be changed by using exact measuring devices or measuring in immediate sequence or knowing exactly the initial conditions.

measurement as disturbance: “A measured value on small size scales in QP cannot be predicted exactly.” (SK3M), corresponding to the coding EM6: (disturbance of quantum state) with the anchor example: “A measurement process disturbs a quantum system.” This disturbance can be of the same or a different nature. The measurement process is said to influence the result state of the quantum system. This influence is not distinguished from the state of the object itself. The act of measurement itself influences the object being measured (ND08M). Alternatively, the influence may be conceptualized as a change, such as a projection onto an eigenstate or as a disturbance. A central point is the conception of the measurement as influencing the state of the quantum system in one way or another: “Each measurement process influences the result state of the quantum system.” (IR08O). In this context, the distinction between the object itself and its state is not made: “The measurement itself influences the object to be measured” (ND08M) or between a change as projection on an eigenstate or a disturbance.

If the measurement process was regarded as a projection, the coding (EM1) was assigned, as given by the anchor example: “Before the measurement process, the system has a free choice between the possible states regarding the measurement. During the measurement, it must assume one of these states to be measured, which is why it is in this state after the process.” However, some students have already developed an understanding of the distinctive characteristics of quantum measurement: “An infinitely accurate measurement equipment is quantum mechanically impossible” (NB03O) or “Infinitely accurate measuring equipment still does not allow a QP certain statement.” (Erl6) The students learn the role of predictability and probability which they can clearly indicate: “When a physical quantity is measured on a quantum object I can predict an outcome only with a probability” (UdA). On the other hand, it becomes clear that some important exceptions or characteristics are not clear to them: “Two measurements of the same quantum object never give the same result” (LA10l) or even: “The state of the system [...] after the measurement will be given by a distribution of states, each corresponding to one of the possible outcomes of the measurement,” giving examples that the measurement process could result in a superposition, corresponding to coding EM 5. The assertion that the result can be predicted in the case of eigenstates is mentioned only by students who have experienced the polarization approach. In this context, the students directly argue with the properties of polarization: “in case the filter was inclined by 0 or 90 with respect to the polarization of the photon then you will have certain outcomes” (UdE).

D. Implications for instrument development

The analysis of the results led to several adjustments to the wording of the items and an optimized construction of the answer options, particularly the distractors. These were constructed by taking student justifications. This resulted in the final questionnaire with eight items and closed single-choice answer options. This process is illustrated here with the example of the measurement process. A closed item was formulated based on students’ responses to the open item “Please explain, how a measurement influences a quantum system” (see Table III). The answer options (distractors) were derived directly from students’ answers.

In a similar manner, all the items were revised, reformulated, and the distractors were based on students’ answers. Table IV provides an overview of the items on the final questionnaire, along with the corresponding categories of answer options and the scores. The item formulations can be found in Table IX.

In sum, the findings from this qualitative study ensure that all items in the final version of the instrument are unambiguous and that the item instructions are clear, thereby contributing to the verification of validity assumption A2 (see Sec. II).

TABLE III. An example of a closed item with answer options corresponding to the codings (categories) found from students’ answers to open items or in justifications of answers. Also, the codings and their scores indicating the degree of quantum thinking are given.

Item: A quantum system that is in a certain state is subjected to a measurement. Indicate the statement that best describes the result.

Answer option	Corresponding category from coding manual	Score
(a) If the quantum system was in one of the possible result states at the beginning, its state becomes a superposition of possible result states after the measurement.	Category EM5 (measurement causes superposition)	1
(b) If the quantum system was in a superposition of states corresponding to possible measurement results, it is in one of the possible result states.	Category EM 1 (projection on eigenspace)	2
(c) If the measurement is very accurate, then the measurement itself influences the possible results.	Category EM 6 (disturbance of quantum state)	0

TABLE IV. Item categories, their corresponding options with references to the coding manual, and scores (indicator for 0-classical thinking, 1-mixed thinking, 2-quantumlike thinking). The coding manual can be found in the Supplemental Material [93].

Item	Option	Reference to category of coding manual	Score
1	A	PM1	1
	B	PM3	2
	C	PM5	0
2	A	PM5	0
	B	PM1	1
	C	PM4	2
3	A	EM6	0
	B	EM2	2
	C	EM7	1
4	A	KS2	0
	B	KS4	1
	C	KS1	2
5	A	GQ4	2
	B	EM6	0
	C	GQ5	1
6	A	EM5	1
	B	EM1	2
	C	EM6	0
7	A		2
	B	MR2	0
	C	KS2	1
8	A	PM1	1
	B	PM5	0
	C	PM1	1
	D	PM3	2

VI. STUDY II: EXPERT SURVEY

As mentioned, the questionnaire avoids mathematical questions and seeks to explore the understanding of quantum measurement conceptually, due to the involvement of secondary school students. As a result, the questions and answers are presented in a simplified form, which carries with it the risk of compromising deep physical interpretation. Therefore, in addition to the authors and the teachers who assisted in the research, the questionnaire created was filled in by experts in order to keep the questionnaire as physically correct as possible.

A. Sample

We conducted an anonymized expert survey to gain insight into the quality of the items from the perspective of physics professors in the area of quantum theory. A total of $N = 3$ responses were received. The experts are from the authors' universities (one from Germany, one from Slovenia, and one from Hungary). More precisely, one of the experts is a quantum education researcher, one who has been responsible for undergraduate QP courses for a long time, and the last one is a quantum theorist.

B. Methodology

The study was conducted by having professors complete the test in the same way as high school students and college students. We examined whether professors gave the same answers to the questions and, if so, where the differences were observed. Additionally, we asked the experts for feedback on the items in terms of both content and language.

C. Results and implications

With the exception of a few questions, the experts' answers were all the same as those we had considered correct during the questionnaire (see the questionnaire in the Appendix). Exceptions were questions 4 and 7, in which two professors marked what we thought to be correct and only one other answer was on the false options. In the 4th question, the false option "b" is as follows: If a quantum object is first prepared in a well-defined state, and then, immediately afterward, a physical quantity is measured on it; the state in which the quantum object will be after the measurement is given by a distribution of states, each corresponding to one of the possible outcomes of the measurement. One expert who answered correctly added to his answer that

> "I chose 'c', neglecting time evolution (which is totally acceptable in high school), and interpreting the term 'well-defined' state as either a state which is not a superposition, either well-defined superposition. I think 'b' is not 'totally correct' because if the initial state is not a superposition (say, spin + along the x axis) and the measurement that follows is of the same quantity (spin along the x axis), then the state will not be 'a distribution of states', but the same state. However, if you consider 0 probability as part of the distribution, then you have two correct answers."

This answer indicates that if students complete the questionnaire, it is important that students do not include probability 0 events as part of the distribution in their previous studies, which is the standard way. In question 7, the false option "b" is if a measurement apparatus consists of two devices arranged in series for the measurement of orthogonally polarized states [$|H\rangle$ (horizontal) and $|V\rangle$ (vertical)]; and a single photon in the equally weighted superposition state of $|H\rangle$ and $|V\rangle$ enters the apparatus; the probability that the photon passes through both devices is 25%. One respondent, who answered question 7 correctly, wrote:

> "I am not sure that I understand this question. What I understand: two polarizers, one after the other, one in '—' orientation and the other in '-' orientation. A photon which is polarized 'x'

> enters the apparatus. After the first polarizers it either pass (50%) either absorbed (50%). If it passes, then it is '—' oriented, therefore will certainly be absorbed in the second polarizer."

Another, very similar response also appeared from an expert:

> "I think [the questions] are clear (but not necessarily easy, every question demands serious thinking, but knowledge of the basic principles of QP suffices). Only the question one before last in relation to V and H polarization appears unclear to me, because it is not specified how the device works. Between the lines I understood that each of the devices transmits one polarization and absorbs the other. This is not self-evident, it would be good to write it down explicitly."

We responded that such devices are the only ones that students have seen (polarizers), but this might not be true, as polarizing beam splitters could be another device that distinguishes between polarizations and is probably used in Mach-Zehnder approaches. These respondents are right, but it is important that these tasks are designed to be adaptable to a specific context; any difficulty in understanding is due to the general wording used, which is because of the different learning materials. It also happens that in a given T/L approach, we have to replace the example of polarization with another context.

We believe that the responses are reasonably consistent, and the minimal variation in some responses is likely to be due to the fact that time evolution of states and measurement uncertainties due to nonideal observation were ignored due to the target audience. Overall, the results of this expert survey contribute to verify the validity assumptions A1 and A2 (see Sec. II).

VII. STUDY III: QUANTITATIVE EVALUATION

In this section, we describe the quantitative evaluation of the so far final version of the questionnaire on the measurement process (see also caption of Fig. 1)

A. Sample

The final version of the instrument comprising a total of eight single-choice items was completed by $N = 201$ upper-level secondary school students (aged 16 to 18 years) from Germany, Hungary, Italy, and Slovenia after instruction. None of the students had received instruction on QP prior to participation in this study. Among the main study learners, there were 11 who participated in the which-path-encoded single-photon approach (see Sec. III A 1, approach BS), 84 who participated in the optical polarization approach (see Sec. III A 2, approach POL), 42 who participated in the double-well approach (see Sec. III A 3, approach DW), and 64 who attended a traditional quantum theory course following the historical development (for a description of the traditional approach (TR) see Ref. [43]). All of the interventions took place in a real classroom setting and covered the key aspects of quantum measurement as part of the instrument to be evaluated. Our instrument was administered as a post-test after instruction.

B. Data analysis

The collected data were first coded based on the scoring rubric outlined in Table IV: each student response (i.e., selection of an answer option on the items) was rated using an ordinal scale ranging from 0 (indicator for classical-like thinking) to 2 (indicator for quantumlike thinking). Summing up the scores for the eight items of the instrument, each student was assigned a score between 0 and 16. The score allows an interpretation of students' understanding of quantum measurement: The higher the score, the more the students' thinking aligns with the current scientific understanding of the topic. After this data preparation, the data analysis consisted of two consecutive steps:

1. In a first step, we evaluated the psychometric descriptives based on classical test theory as outlined in Ref. [94]. Therefore, we analyzed the items difficulty where we refer to the widely used tolerance range of 0.2 to 0.8 (see Ref. [95]) as well as the discriminatory power with values above 0.2 being considered acceptable (see Ref. [96]). Additionally, Cronbach's alpha [97] was calculated as a measure of internal consistency.
2. In a second step, we—more locally—analyzed the distribution of responses to the different answer options and checked that each answer option was chosen by at least 5% of the learners as suggested in Ref. [67]. Moreover, we compared the response distributions between the different cohorts comprised in our main study sample whose students attended the different two-state approaches described above. This procedure contributes to ensure instruction sensitivity of the items included in the questionnaire under investigation, since
   (a) "an instructionally sensitive test should be able to detect differences in instruction received by students" ([98], p. 3), and
   (b) such items are assumed to be particularly instructionally sensitive if the item parameters vary across learning groups (for example see [99]).

To further substantiate the results, an analysis is carried out within the content areas covered by the instrument by calculating the relative scores achieved by the students in the three content areas and comparing them across the four cohorts (i.e., between the students who had participated in the different two-state approaches under investigation).

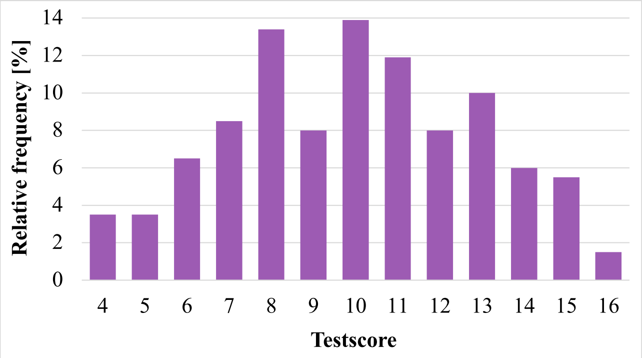

FIG. 2. Distribution of the students' test scores.

C. Results of quantitative evaluation

The results reported in this subsection refer to the final version of the instrument comprising eight single-choice items as emerged from the qualitative evaluation. The complete items can be found in the Appendix.

1. Psychometric characterization

In the eight single-choice items of the instrument, the students could score a maximum of 16 points (for the coding scheme, see Table IV). The students reached an average score of $m = 9.94$ points (median $\mathrm{Mdn} = 10.00$ points) with a standard deviation of $\mathrm{SD} = 2.96$ points, ranging from 4 points (scored by 7 participants) to 16 points (scored by 3 participants). The test score distribution is shown in Fig. 2.

The psychometric properties of the instrument, i.e., item difficulties and their discriminatory power, are shown in Table V. Cronbach's alpha as an estimator of the instrument's internal consistency was found to be $\alpha = 0.57$. Although this value indicates a rather low internal consistency, it can be considered acceptable due to it being composed of only eight items (e.g., see Ref. [100]). It can be seen from Table V that the psychometric characterization suggests further evaluation of item 2 in the future. Both item difficulty and discriminatory power of this item lay outside the suggested ranges with values of 0.84 and 0.12, respectively. The exclusion of item 2 would result in a slight increase of Cronbach's alpha, while at the same time being relevant for the instrument's content coverage. Acceptable psychometric properties can be confirmed for the other items of the instrument leading to a verification of the validity assumption A3 (see Sec. II).

TABLE V. Psychometric properties of each item: Item difficulty, discriminatory power, and adjusted Cronbach's alpha $\overline{\alpha_n}$, i.e., the Cronbach's alpha of the scale if the respective item n is omitted.

Item no.	Item difficulty	Discriminatory power	$\overline{\alpha_n}$
1	0.64	0.27	0.53
2	0.84	0.12	0.57
3	0.58	0.27	0.53
4	0.60	0.43	0.47
5	0.55	0.21	0.55
6	0.70	0.40	0.48
7	0.48	0.19	0.56
8	0.60	0.33	0.51

2. Instruction sensitivity

The response distribution for all items of the instrument, for the total sample as well as divided by cohorts, is presented in Table VI. It becomes apparent that all answer options are chosen by a significant share of the sample students, further supporting validity assumption A3. Only answer option 3 of item 1 which states that "Almost infinitely accurate measuring equipment must be used" in order to be able to predict with certainty the result of a measurement on a quantum state seems borderline in all cohorts since it is chosen by only slightly more or less than 5% of the students from all cohorts. A refinement of the formulation "almost infinitely accurate" seems a sensible approach in order to increase the attractiveness of this answer option.

As can be seen from the response distributions shown in Table VI, there are differences in the response patterns between the students from the different cohorts. There is a global tendency for the proportion of students who chose the answer options indicating quantumlike thinking to be higher for the approach focusing on single photons incident on beam splitters (BS group) and the double well approach (DW group) than for students taking a traditional quantum theory course focusing on historical development (TR group) or the light polarization approach (POL group).

However, a more holistic picture emerges when comparing the differences in students' responses to items within each content area. Therefore, we summed the students' scores on the items within the three content domains, which is justified by the sufficient internal consistencies of the resulting subscales as assessed by Cronbach's alpha, taking into account the small number of items for each domain (>0.50 for each subscale). We then divided the scores by the total possible score within the domains to compare relative domain scores across cohorts. An overview of the relative domain scores for the different cohorts can be found in Table VII.

It can be seen that the mean relative domain scores (m) vary between the different instructional approaches, supporting the validity assumption A4 regarding the instructional sensitivity of the instrument. More specifically, students who had participated in an approach focused on single-photon incident on beam splitters (BS) or the double-well (DW) approach prior to test administration outperformed their peers, indicating their effectiveness in promoting conceptual learning of quantum measurement. Overall, students who had been introduced to QP via one of the two-state approaches (POL, BS, % DW groups) outperformed students who had taken a traditional QP course (TR group) following

TABLE VI. Response distribution for all items of the instrument, divided by the cohorts comprised in the sample: students who participated in (a) a traditional quantum theory course following the historical development (abbreviated TR), in (b) the light polarization approach (abbreviated POL), in (c) an approach focusing on single-photons incident on beam splitters (abbreviated BS), or in (d) the double-well approach (abbreviated DW). The relative frequencies (in %) are given and are written in bold italics for answer options which serve as indicators for quantumlike thinking, in italics for answer options which serve as indicators for mixed thinking, and in normal font for answer options which serve as indicators for classical thinking.

	Answer option 1				Answer option 2				Answer option 3				Answer option 4			
Item	TR	POL	BS	DW	TR	POL	BS	DW	TR	POL	BS	DW	TR	POL	BS	DW
1	*79.7*	*70.2*	*36.4*	*21.4*	***12.5***	***25.0***	***54.5***	***73.8***	7.8	4.8	9.1	4.8	···	···	···	···
2	6.3	15.5	0.0	11.9	*3.1*	*23.8*	*9.1*	*2.4*	***90.6***	***60.7***	***90.9***	***85.7***	···	···	···	···
3	50.0	22.6	36.4	16.7	***31.3***	***50.0***	***54.5***	***59.5***	*18.8*	*27.4*	*9.1*	*23.8*	···	···	···	···
4	28.1	23.8	0.0	7.1	*60.9*	*27.4*	*45.5*	*33.3*	***10.9***	***48.8***	***54.5***	***59.5***	···	···	···	···
5	***40.6***	***40.5***	***54.5***	***54.8***	48.4	32.1	36.4	16.7	*10.9*	*27.4*	*9.1*	*28.6*	···	···	···	···
6	*21.9*	*34.5*	*27.3*	*11.9*	***46.9***	***53.6***	***72.7***	***76.2***	31.3	11.9	0.0	11.9	···	···	···	···
7	***21.9***	***52.4***	***45.5***	***0.0***	62.5	8.3	45.5	45.2	*15.6*	*39.3*	*9.1*	*54.6*	···	···	···	···
8	*29.7*	*45.2*	*36.4*	*2.4*	3.1	15.5	0.0	31.0	*42.2*	*21.4*	*27.3*	*57.1*	***25.0***	***17.9***	***36.4***	***9.5***

TABLE VII. Mean values m and standard deviations SD of relative domain scores (in %) for the different cohorts comprised in the sample: students who participated in (a) a traditional quantum theory course following the historical development (abbreviated TR), in (b) the light polarization approach (abbreviated POL), in (c) an approach focusing on single-photons incident on beam splitters (abbreviated BS), or in (d) the double-well approach (abbreviated DW).

			Relative domain score m (SD)			
Content domain	Items	Possible domain score	TR	POL	BS	DW
Knowledge of state	4 and 6	4	49.6 (30.7)	66.7 (27.3)	81.8 (22.6)	79.2 (29.2)
Effect of measurement	1, 3, and 5	6	46.4 (24.1)	59.3 (20.8)	63.6 (26.7)	75.0 (23.6)
Statistical predictability	2, 7, and 8	6	60.1 (20.2)	65.3 (21.8)	72.7 (23.4)	62.7 (19.4)

the historical development in each of the content areas covered by the instrument. In other words, the context (i.e., which two-state approach is used) does indeed matter for student learning, but in general, two-state approaches appear to be particularly conducive to learning quantum concepts (specified in this article for quantum measurement) compared to traditional instruction. While in the study presented in this paper, we refrain from more in-depth statistical analyses of group differences due to the different sizes of the cohorts and the development of a new instrument that is the focus of this study, the data presented here suggest that the newly developed instrument is useful for use in cluster-randomized field studies in the future.

VIII. DISCUSSION

Our aim was to create an instrument that measure the students' conceptual understanding of the quantum measurement in two-state systems: the total scores of students appropriately represent their knowledge, and the scores in the different domains of quantum measurement also ensure us how developed their conceptual understanding in that aspect. Following classical test theory [67,94], Table V included item difficulty and item discriminatory values for each item. Often the accepted item difficulty tolerance value is between 0.2 and 0.8; and a discriminatory value above 0.2 is considered good, while others suggest 0.3 [101]. The reliability of the instrument was calculated using Cronbach's alpha as an estimate of internal consistency [97]; thus yielding $\alpha = 0.56$. However, given the small number of questions in our instrument ($n = 8$), it is natural that we did not obtain a large α. According to Ref. [102], for concept inventories values of $\alpha \geq 0.55$ are considered sufficient, and our instrument has been found above this limit (see Table VIII). Combined with the adjustment $\alpha^* = \alpha \cdot (n - 1)/n$ for scales of small length n by Bauer [100], this yields different thresholds. This indicates that our instrument has sufficient reliability, and the low score is mainly due to the small numbers of tasks in the questionnaire alpha scales with the number of items. Furthermore, the reliability of the questionnaire may also be due to the fact that we used a very wide range of teaching approaches and very different students (from 16-year-old secondary school students to university students) since the aim was to create an instrument that would allow comparisons between different teaching materials. Taking these into account will increase the quality of the questionnaire.

We therefore discuss our four expected assumptions (A1–A4 in Sec. II) on the interpretation of the instrument score according to our research results.

TABLE VIII. *Categorical Judgment Scheme* adopted from Ref. [96]. Values in parentheses specify the number of items that are allowed to lie outside this suggestion [96]. In the last row, $\alpha^* = \alpha \cdot (n-1)/n$ is the modified threshold for Cronbach's alpha suggested for scales of small length n in Ref. [100].

Criterion	Excellent	Good	Average	Poor	Our instrument on quantum measurement
Item statistics					
difficulty	0.2–0.8	0.2–0.8 (3)	0.1–0.9	0.1–0.9 (3)	Good
Discrimination	>0.2	>0.1	>0.0	>−0.2	Good
Total score reliability					
α of total score	>0.9	>0.8	>0.65	>0.5	Poor
α-with-item-deleted	All items less than overall α	(3)	(6)	(9)	Good
α^* ($n = 8$)	>0.79	>0.7	>0.57	>0.48	Average

A1: In constructing the instrument, we used test questions from the literature, supplemented this with experts' reflections and experiences (see Sec. IV D), and finally surveyed experts (Sec. VI). The result is that the items adequately represent the conceptual understanding of the introductory aspects of quantum measurement in two-state systems.

A2: The use of the literature and experts' experience and reflections (Sec. IV D) and feedback from the survey of experts (Sec. VI), together with the qualitative analysis (see Sec. V), in the construction of the instrument, justify that the items are clear and the instructions are physically and didactically straightforward. The only problem that arose in the experts' survey was caused by items 4 and 7, which were due to the circumstances of secondary school: we neglected the time evolution of states; we did not include states of impossible events in the distribution of outcomes; and we formulated the polarization question (item 7) in a very general way due to the general language. However, time evolution is not usually taught in schools; in standard education, impossible events are not included in the distribution; and the general formulation is justified since the aim was to make our instrument applicable to as wide a sample as possible and to translate it to an arbitrary two-state quantum system T/L approach.

A3: The result of the quantitative psychometric categorization (Sec. VII) shows that items and distractors are valid and useful for use in the target sample. Figure 2 clearly shows that students scored very selectively and the distribution of scores fits well with the normal distribution. As Table V shows, only question 2 was found to be worthy of future consideration, since it was found to be too difficult and therefore not the perfect choice for testing students' understanding, while its low discriminatory power makes it less able to separate students by knowledge. The lower reliability was due to the low number of questions and the very wide target audience. However, the reason for this is self-evident, as the instrument was designed to fit comfortably into a 45-min lesson and could be used in any school or university classroom presenting a two-state system.

A4: The large-sample empirical study (Sec. VII) shows that the items are educationally sensitive because students' responses show a high degree of dispersion, with at least 5% of students marking each response option. Table VI summarizes well that students in different learning pathways show completely different response patterns, which is also shown by the conclusions drawn from Table V.

Our new instrument has been found to be suitable for testing the conceptual understanding of measurements in two-state quantum systems by students at different levels of QP. A comparison of educational approaches using two-state quantum systems with traditional approaches (see Table VI) concludes that a better conceptual understanding of quantum measurement in two-state systems can be achieved with educational approaches using two-state systems than with traditional approaches. However, the questionnaire is also useful in practice, by making the conceptual understanding of measurement an educational objective, the instructor can choose a more appropriate teaching approach for this purpose. In addition, the created test helps the instructor to recognize the most important aspects of quantum measurement in the teaching and to incorporate misconceptions corresponding to the questions of the instrument into the lessons. We also think that the questionnaire could be well integrated into a high school or undergraduate university QP course, e.g., in the sense of peer instruction.

IX. LIMITATIONS

During the design and evaluation of a new instrument, decisions have to be made that influence the findings. Hence, it is necessary to be transparent about the choices made and the limitations that go along with them. Regarding the study presented in this paper, the following limitations are to be considered:

1. Content-related limitation: Not all possible two-state QP teaching approaches have been included in our sample, e.g., the electron spin approach has not been considered in this project [103]. Instead, we focused on teaching-learning approaches methods [traditional (TR), which-path encoded single-photon approach (BS), optical polarization (POL), and double-well

approach (DW)] which (a) the authors have direct experience with their design, rationale, implementation, and evaluation and (b) have been investigated and evaluated in prior research. Also, one needs to consider that the objective of this project was to develop an instrument rather than empirically investigating the different two-state approaches.

2. Sampling and study design-related limitations: In addition, the number of students participating in the four teaching approaches was not the same for all approaches, and one of the approaches included in this study (the BS approach) was particularly underrepresented in the total sample. Therefore, conclusions drawn from statistical comparisons between groups are limited and should be considered primarily for the evaluation of the instrument. Future empirical research aimed at comparing the different teaching approaches in terms of their impact on student learning will require more robust sampling. In addition, the physics and mathematics backgrounds of the participating students may have been different; for example, the polarization approach involved an average of 16-year-old students, and many of them received only qualitative instruction, leading to a limited interpretation of the results in Table VI. As outlined above, research comparing instructional approaches should take care to control for students' prior knowledge and skills through pretesting. However, for our purpose of evaluating a newly developed instrument, the inclusion of a group of learners with a wide range of abilities may even be considered advantageous in order to assess the functioning of the items.
3. Instrument-related limitation: The instrument developed in this study includes a total of eight items which limits both statistical power and breadth of content coverage. However, we describe a complex design process, including literature review, qualitative evaluation, expert feedback, and quantitative evaluation, which resulted in items that (a) can be easily translated to different two-state contexts and which (b) are tailored to the secondary school level in terms of mathematics and language used—this is what distinguishes this instrument from other ones published in the literature (see Sec. III).

X. CONCLUSION

In this paper, we describe the design and evaluation of a short eight-item instrument to assess students' conceptual understanding of the core aspects of quantum measurement relevant to secondary school students. The innovative aspect of this instrument lies in the design of its items: We describe a complex design process, including literature review, qualitative evaluation, and expert discussions, which resulted in items that can be easily translated into different two-state contexts. Thus, the items allow for an empirical comparison of students' learning of quantum measurement using different two-state approaches. Such a comparison has not been possible before due to the lack of appropriate instruments. In this respect, this paper paves the way for further test development using the idea of isomorphic items, also with respect to quantum topics other than quantum measurement. In the future, we will use the instrument developed in this project for cluster-randomized field trials comparing the effects of different two-state approaches on student learning, in order to substantiate the first empirical evidence presented here with larger samples.

Data supporting the findings presented in this paper are available upon request from the corresponding author.

ACKNOWLEDGMENTS

One of us (K. T.) is grateful for receiving an excellence-based stipend provided to young Hungarian science teachers ("FITT" Creator Award) by one of Gedeon Richter Plc's public benefit foundations (Richter Gedeon Nyrt. Centenáriumi Alapítvány), because this support contributed financially to his work. The work of K. T. is also supported by the Research Program for Public Education Development of the Hungarian Academy of Sciences (Project No. SZKF-7/2022).

APPENDIX

TABLE IX. The final version of the questionnaire. The correct answer is indicated in bold.

Item	Choices
1. Which of the following statements describes a method for predicting with certainty the result of a measurement of a quantity P for a single quantum object?	(a) There is no such method, because in quantum physics, it is not possible to predict the result of a single measurement. **(b) The quantum object must be prepared so that it attains a certain value of the quantity P.** (c) Almost infinitely accurate measuring equipment must be used.

(Table continued)

TABLE IX. *(Continued)*

Item	Choices
2. Indicate which statement is correct about the result of a measurement of a quantity P on a single quantum object?	(a) We could predict the outcome of the measurement if we could fully obtain the necessary information about the quantum object. (b) We cannot predict anything about a single measurement because the value of the result is not precisely defined. **(c) We can predict the probability of obtaining a certain value if the state of the quantum object is known.**
3. Indicate the reason why the value of a given quantity may be indeterminate.	(a) A measurement in quantum physics disturbs the quantum object because quantum objects are microscopic. **(b) The laws of nature state that the results of individual measurements in quantum physics are not determined, except in a few special cases.** (c) In general, a quantum system cannot be brought into exactly the same state twice.
4. A quantum object is first prepared in a well-defined state. Immediately afterward, a physical quantity is measured on it. Choose the statement that best describes the final state.	(a) The measurement process in quantum physics changes the state in an unpredictable way. (b) The state in which the quantum object will be after the measurement is given by a distribution of states, each corresponding to one of the possible outcomes of the measurement. **(c) If preparation and measurement process concern the same physical quantity, the state of the quantum object remains the same.**
5. A quantity A is first measured on a set of identical quantum objects. Then, a different quantity B is immediately measured on those quantum objects from this set which are in the state $\lvert a\rangle$. Some of these quantum objects are then in a state $\lvert b\rangle$. What can be said about their state?	**(a) When B is measured, the objects in the state $\lvert a\rangle$ are transferred to another state $\lvert b\rangle$, which is completely different from the state $\lvert a\rangle$.** (b) The process of measuring B disturbs the state $\lvert a\rangle$, which is equal (or very close) to the state $\lvert b\rangle$ only if the disturbance caused by the measurement process is minimal. (c) The process of measuring B will progressively evolve the state $\lvert a\rangle$ toward the state $\lvert b\rangle$ according to the laws of quantum physics.
6. A quantum system that is in a certain state is subjected to a measurement. Indicate the statement that best describes the result.	(a) If the quantum system was in one of the possible result states at the beginning, its state becomes a superposition of possible result states after the measurement. **(b) If the quantum system was in a superposition of states corresponding to possible measurement results, it is in one of the possible result states.** (c) If the measurement is very accurate, then the measurement itself influences the possible results.
7. A measurement apparatus consists of two devices arranged in series for the measurement of orthogonally polarized states ($\lvert H\rangle$ (horizontal) and $\lvert V\rangle$ (vertical)). A single photon in the equally weighted superposition state of $\lvert H\rangle$ and $\lvert V\rangle$ enters the apparatus. What can be predicted?	**(a) When the single photon passes through the first device, it will certainly be absorbed by the second device.** (b) The probability that the photon passes through both devices is 25%. (c) The probability that the photon passes through both devices is 50%.
8. A given quantum object can be described using two states (two-state system). Imagine that this quantum object is in a superposition state and Alice performs a measurement on it. For this purpose, she chooses a suitable measuring device. Now Bob measures the same quantum object immediately afterward with an identical measuring device. Alice expects Bob to get the same result as she does.	(a) No, because the results of measurements in quantum physics are unpredictable. (b) Yes, if she and Bob measure accurately enough. (c) No, because the result of a second measurement cannot be predicted precisely. **(d) Yes, because if the same measuring device is used twice in a row for the same object, the result must be the same.**

[1] B. Akarsu, Instructional designs in quantum physics: A critical review of research, Asian J. Appl. Sci. **4**, 112 (2011).

[2] E. K. Henriksen, B. Bungum, C. Angell, C. W. Tellefsen, T. Frågåt, and M. V. Bøe, Relativity, quantum physics and philosophy in the upper secondary curriculum: Challenges, opportunities and proposed approaches, Phys. Educ. **49**, 678 (2014).

[3] D. Zollman, Oersted lecture 2014: Physics education research and teaching modern modern physics, Am. J. Phys. **84**, 573 (2016).

[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] K. Krijtenburg-Lewerissa, H. J. Pol, A. Brinkman, and W. R. van Joolingen, Key topics for quantum mechanics at secondary schools: A Delphi study into expert opinions, Int. J. Sci. Educ. **41**, 349 (2019).

[6] M. Michelini and A. Stefanel, Approaches on T/L quantum physics from PER literature, in *Teaching-Learning Contemporary Physics: From Research to Practice*, edited by B. Jarosievitz and C. Sükösd (Springer International Publishing, Cham, 2021), pp. 3–17.

[7] M. Michelini and A. Stefanel, Research based educational paths on quantum mechanics for high school students, in *Connecting Research in Physics Education with Teacher Education 3* (The International Commission on Physics Education, University of the Basque Country and Dublin City University, 2022), pp. 40–75.

[8] G. Pospiech, Philosophy and quantum mechanics in science teaching, Sci. Educ. **12**, 559 (2003).

[9] G. Ghirardi, *Sneaking a Look at God's Cards* (Princeton University Press, Princeton, NJ, 2004).

[10] G. Pospiech, M. Michelini, A. Stefanel, and L. Santi, Central features of quantum theory in physics education, Front. Phys. Educ. 85 (2008), https://www.fisica.uniud.it/urdf/laurea/idifo1/materiali/g6/O-G6+C1_MQUDWSGirepOpatiaTot.pdf.

[11] Z. Patterson and L. Ding, Students' pre-instructional perspectives of quantum physics, presented at PER Conf. 2020, virtual conference, 10.1119/perc.2020.pr.Patterson.

[12] H. K. E. Stadermann and M. J. Goedhart, Secondary school students' views of nature of science in quantum physics, Int. J. Sci. Educ. **42**, 997 (2020).

[13] 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).

[14] M. Michelini, L. Santi, A. Stefanel, and M. L. Chiofalo, Entangled research methods for the building of coherent conceptual thematic learning paths and connecting research with praxis, Nuovo Cimento C **46**, 196 (2023).

[15] J. R. Hoehn, J. D. Gifford, and N. D. Finkelstein, Investigating the dynamics of ontological reasoning across contexts in quantum physics, Phys. Rev. Phys. Educ. Res. **15**, 010124 (2019).

[16] D. F. Styer, M. S. Balkin, K. M. Becker, M. R. Burns, C. E. Dudley, S. T. Forth, J. S. Gaumer, M. A. Kramer, D. C. Oertel, L. H. Park, M. T. Rinkoski, C. T. Smith, and T. D. Wotherspoon, Nine formulations of quantum mechanics, Am. J. Phys. **70**, 288 (2002).

[17] C. Baily and N. D. Finkelstein, Teaching quantum interpretations: Revisiting the goals and practices of introductory quantum physics courses, Phys. Rev. ST Phys. Educ. Res. **11**, 020124 (2015).

[18] M. S. Ubben, J. M. Veith, A. Merzel, and P. Bitzenbauer, Quantum science in a nutshell: Fostering students' functional understanding of models, Front. Educ. **8**, 1192708 (2023).

[19] K. Tóth, M. Michelini, and P. Bitzenbauer, From light polarization to quantum physics: Supporting lower secondary school students' transition from gestalt to functional thinking, Eurasia J. Math. Sci. Technol. Educ. **20**, em2449 (2024).

[20] M. Michelini and A. Stefanel, A path to build basic quantum mechanics ideas in the context of light polarization and learning outcomes of secondary students, J. Phys. Conf. Ser. **1929**, 012052 (2021).

[21] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, England, 1987).

[22] *Mathematics in Physics Education*, 1st ed., edited by G. Pospiech, M. Michelini, and B.-S. Eylon (Springer, Cham, 2019), pp. X, 385.

[23] G. Pospiech, A. Merzel, G. Zuccarini, E. Weissman, N. Katz, I. Galili, L. Santi, and M. Michelini, The role of mathematics in teaching quantum physics at high school, in Teaching-Learning Contemporary Physics: From Research to Practice, edited by B. Jarosievitz and C. Sükösd (Springer International Publishing, Cham, 2021), pp. 7–70.

[24] R. Müller and H. Wiesner, Teaching quantum mechanics on an introductory level, Am. J. Phys. **70**, 200 (2002).

[25] M. d. l. Á. Fanaro, M. R. Otero, and M. Arlego, Teaching basic quantum mechanics in secondary school using concepts of Feynman Path Integrals Method, Phys. Teach. **50**, 156 (2012).

[26] M. B. Schneider and I. A. LaPuma, A simple experiment for discussion of quantum interference and which-way measurement, Am. J. Phys. **70**, 266 (2002).

[27] A. Pereira, F. Ostermann, and C. Cavalcanti, On the use of a virtual Mach–Zehnder interferometer in the teaching of quantum mechanics, Phys. Educ. **44**, 281 (2009).

[28] 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).

[29] C. H. Holbrow, E. Galvez, and M. E. Parks, Photon quantum mechanics and beam splitters, Am. J. Phys. **70**, 260 (2002).

[30] E. M. G., Introducing the uncertainty principle, in *Seminar on the Teaching of Physics in Schools 2*, edited by A. Loria and P. Thomsen (Gyldendal, Oslo, 1975), pp. 220–256.

[31] I. M. Greca and O. Jr. Freier, Does an emphasis on the concept of quantum states enhance students' understanding of quantum mechanics?, Sci. Educ. **12**, 541 (2003).

[32] R. G. Newton, What is a state in quantum mechanics?, Am. J. Phys. **72**, 348 (2004).

[33] S. Faletič, M. Bondani, P. J. Emigh, K. Krijtenburg-Lewerissa, G. Pospiech, and M. Michelini, Symposium on teaching and learning quantum physics, in Physics Education Today: Innovative Methodologies, Tools and Evaluation, edited by C. Fazio and P. Logman (Springer Nature Switzerland, Cham, 2024), pp. 55–72.

[34] D. Paul Adrien Maurice, *The Principles of Quantum Mechanics* (Clarendon, Oxford, 1958).

[35] R. P. Feynman, *The Feynman Lectures on Physics* (Addison-Wesley, Reading, MA, 1965).

[36] A. P. French and E. F. Taylor, *An Introduction to Quantum Physics* (Norton, New York, 1978), Vol. 3.

[37] J. J. Sakurai, *Modern Quantum Mechanics* (Clarendon, Oxford, 1985).

[38] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, New York, 2010).

[39] Quantum Flagship, https://qt.eu (2022).

[40] M. Michelini, L. Santi, and A. Stefanel, JQM per affrontare nella scuola secondaria i fondamenti di meccanica quantistica, in Proceedings of Didamatica (2016).

[41] S. Faletić, A double well on-line simulation and activities for active learning of introductory quantum mechanics, Eur. J. Phys. **41**, 045706 (2020).

[42] A. Kohnle, M. Douglass, T. J. Edwards, A. D. Gillies, C. A. Hooley, and B. D. Sinclair, Developing and evaluating animations for teaching quantum mechanics concepts, Eur. J. Phys. **31**, 1441 (2010).

[43] P. Bitzenbauer, Effect of an introductory quantum physics course using experiments with heralded photons on preuniversity students' conceptions about quantum physics, Phys. Rev. Phys. Educ. Res. **17**, 020103 (2021).

[44] Introduction to quantum physics—development and evaluation of a new course, https://www.fisica.uniud.it/~stefanel/SNFMI/Articoli/ArticoliMQ/FischlerNarst.pdf (1999) [accessed: June 27, 2024].

[45] E. Y. Weissman, A. Merzel, N. Katz, and I. Galili, Teaching quantum physics as a structured physics theory in high school, J. Phys. Conf. Ser. **1929**, 012051 (2021).

[46] A. Anshu and S. Arunachalam, A survey on the complexity of learning quantum states, Nat. Rev. Phys. **6**, 59 (2024).

[47] A. Merzel, P. Bitzenbauer, K. Krijtenburg-Lewerissa, K. Stadermann, E. Andreotti, D. Anttila, M. Bondani, M. L. M. Chiofalo, S. Faletič, R. Frans *et al.*, The core of secondary level quantum education: A multistakeholder perspective, Eur. Phys. J. Quantum Technol. **11**, 27 (2024).

[48] P. Bitzenbauer and J.-P. Meyn, Von Koinzidenzen zu Wesenszügen der Quantenphysik: Erste Ergebnisse einer summativen Evaluation des Erlanger Unterrichtskonzepts zur Quantenoptik, PhyDid B **1**, 157 (2020), https://ojs.dpg-physik.de/index.php/phydid-b/article/view/1024/1122.

[49] M. Michelini, A. Stefanel, and K. Tóth, Implementing Dirac approach to quantum mechanics in a Hungarian secondary school, Educ. Sci. **12**, 606 (2022).

[50] P. Bitzenbauer, *Quantenoptik an Schulen: Studie im Mixed-Methods Design zur Evaluation des Erlanger Unterrichtskonzepts zur Quantenoptik* (Logos Verlag Berlin, Berlin, 2020), Vol. 303.

[51] M. Michelini, L. Santi, and A. Stefanel, Building quantum formalism in upper secondary school students, in *Teaching and Learning Physics Today: Challenges? Benefits?* (Université de Reims Champagne Ardenne, Reims, 2011).

[52] K. Melhuish, The group theory concept assessment: A tool for measuring conceptual understanding in introductory group theory, Int. J. Res. Undergrad. Math. Educ. **5**, 359 (2019).

[53] I. M. Greca and O. Jr. Freier, 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).

[54] P. Lautesse, A. Vila Valls, F. Ferlin, J. L. Héraud, and H. Chabot, Teaching quantum physics in upper secondary school in France: Quantons' versus 'wave–particle' duality, two approaches of the problem of reference, Sci. Educ. **24**, 937 (2015).

[55] G. C. Ghirardi, R. Grassi, and M. Michelini, *Thinking Physics for Teaching*, edited by Aster (Plenum Publishing Corporation, New York, 1996), p. 329.

[56] M. Michelini, Approaching the theory of quantum mechanics: The first steps towards a coherent synthesized interpretation with a supporting formalism, in *Frontiers of Physics Education*, edited by R. Jurdana-Sepic *et al.* (Zlatni rez, Rijeka, 2008), pp. 93–101.

[57] D. H. McIntyre, *Spin and Quantum Measurement* (PH 425 Paradigm 5, 2002), https://www.if.ufrj.br/~carlos/fismod/seminarios/SternGerlach/SternGerlach_programas/SpinBook02.pdf.

[58] C. Sutrini, G. Zuccarini, M. Malgieri, M. C. Ceruti, G. Pancani, and C. Macchiavello, Logic circuits and optical circuits: A teaching learning sequence to build quantum computation for high school students, Il Nuovo Cimento C **46**, 80 (2023).

[59] A. López-Incera and W. Dür, Entangle me! A game to demonstrate the principles of quantum mechanics, Am. J. Phys. **87**, 95 (2019).

[60] M. L. Chiofalo, C. Foti, M. Michelini, L. Santi, and A. Stefanel, Games for teaching/learning quantum mechanics: A pilot study with high-school students, Educ. Sci. **12**, 446 (2022).

[61] M. Michelini, R. Ragazzon, L. Santi, and A. Stefanel, Discussion of a didactical proposal on quantum mechanics with secondary school students, Il Nuovo Cimento **27C**, 555 (2004).

[62] K. Tóth, M. Michelini, and P. Bitzenbauer, From light polarization to quantum physics: Supporting lower secondary school students' transition from gestalt to functional thinking, Eurasia J. Math. Sci. Technol. Educ. **20**, em2449 (2024).

[63] M. Bondani, M. L. Chiofalo, E. Ercolessi, C. Macchiavello, M. Malgieri, M. Michelini, O. Mishina, P. Onorato, F. Pallotta, S. Satanassi, A. Stefanel, C. Sutrini, I. Testa, and G. Zuccarini, Introducing quantum

technologies at secondary school level: Challenges and potential impact of an online extracurricular course, Physics **4**, 1150 (2022).

[64] A. Mashhadi and B. Woolnough, Insights into students' understanding of quantum physics: Visualizing quantum entities, Eur. J. Phys. **20**, 511 (1999).

[65] H. R. Sadaghiani and S. J. Pollock, Quantum mechanics concept assessment: Development and validation study, Phys. Rev. ST Phys. Educ. Res. **11**, 010110 (2015).

[66] G. Zhu and C. Singh, Surveying students' understanding of quantum mechanics in one spatial dimension, Am. J. Phys. **80**, 252 (2012).

[67] P. Bitzenbauer, Development of a test instrument to investigate secondary school students' declarative knowledge of quantum optics, Eur. J. Sci. Math. Educ. **9**, 57 (2021).

[68] M. Michelini, R. Ragazzon, L. Santi, and A. Stefanel, Proposal for quantum physics in secondary school, Phys. Educ. **35**, 406 (2000).

[69] B. J. Pearson and D. P. Jackson, A hands-on introduction to single photons and quantum mechanics for undergraduates, Am. J. Phys. **78**, 471 (2010).

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

[71] P. Bitzenbauer, Practitioners' views on new teaching material for introducing quantum optics in secondary schools, Phys. Educ. **56**, 055008 (2021).

[72] 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).

[73] P. Bronner, A. Strunz, C. Silberhorn, and J.-P. Meyn, Interactive screen experiments with single photons, Eur. J. Phys. **30**, 345 (2009).

[74] H. J. Kimble, M. Dagenais, and L. Mandel, Photon antibunching in resonance fluorescence, Phys. Rev. Lett. **39**, 691 (1977).

[75] 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).

[76] P. Bronner, A. Strunz, C. Silberhorn, and J.-P. Meyn, Demonstrating quantum random with single photons, Eur. J. Phys. **30**, 1189 (2009).

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

[78] P. Bitzenbauer and J.-P. Meyn, Toward types of students' conceptions about photons: Results of an interview study, in *Physics Teacher Education: What Matters?* (Springer, Cham, 2022), pp. 175–187.

[79] D. G. Jones, Teaching modern physics-misconceptions of the photon that can damage understanding, Phys. Educ. **26**, 93 (1991).

[80] R. Kidd, J. Ardini, and A. Anton, Evolution of the modern photon, Am. J. Phys. **57**, 27 (1989).

[81] M. Michelini, L. Santi, and A. Stefanel, JQM per affrontare nella scuola secondaria i fondamenti di meccanica quantistica, Didamatica (2016), https://core.ac.uk/download/pdf/154285679.pdf.

[82] K. Tóth, Integrating dirac approach to quantum mechanics into physics teacher education, AIP Conf. Proc. **2843**, 050011 (2023).

[83] K. Tóth, Dirac's approach to quantum mechanics in physics teacher education: From linear to circular polarisation, J. Phys. Conf. Ser. **2750**, 012023 (2024).

[84] K. Tóth and T. Tél, Quantum uncertainty: What to teach?, Phys. Educ. **58**, 025019 (2023).

[85] E. Etkina, D. T. Brookes, and G. Planinsic, *Investigative Science Learning Environment* (Morgan & Claypool, San Rafael, CA, 2019).

[86] M. Michelini, R. Ragazzon, L. Santi, and A. Stefanel, Discussion of a didactic proposal on quantum mechanics with secondary school students, Il Nuovo Cimento **27C**, 555 (2005).

[87] S. Messick, Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning, Am. Psychol. **50**, 741 (1995).

[88] M. T. Kane, Current concerns in validity theory, J. Educ. Measure. **38**, 319 (2001).

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

[90] R. Omnes, *The Interpretation of Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1994).

[91] L. E. Ballentine, The statistical interpretation of quantum mechanics, Rev. Mod. Phys. **42**, 358 (1970).

[92] C. Singh and E. Marshman, Review of student difficulties in upper-level quantum mechanics, Phys. Rev. ST Phys. Educ. Res. **11**, 020117 (2015).

[93] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.20.020136 for coding manual used in study I (qualitative evaluation).

[94] P. V. Engelhardt, An introduction to classical test theory as applied to conceptual multiple-choice tests, in Getting Started in PER (2009), Vol. **2**, p. 1.

[95] T. J. Kline, *Psychological Testing: A Practical Approach to Design and Evaluation* (Sage Publications, New York, 2005).

[96] N. Jorion, B. D. Gane, K. James, L. Schroeder, L. V. DiBello, and J. W. Pellegrino, An analytic framework for evaluating the validity of concept inventory claims, J. Eng. Educ. **104**, 454 (2015).

[97] K. S. Taber, The use of Cronbach's alpha when developing and reporting research instruments in science education, Res. Sci. Educ. **48**, 1273 (2018).

[98] M. S. Polikoff, Instructional sensitivity as a psychometric property of assessments, Educ. Meas. **29**, 3 (2010).

[99] A. Robitzsch, Methodische herausforderungen bei der kalibrierung von leistungstests, Bildungsstandards Deutsch und Mathematik: Leistungsmessung in der Grundschule (2009), p. 42.

[100] A. Bauer, Cronbachs α im Kontext des Grundmodells der Klassischen Testtheorie und darüber Hinaus [Cronbach's

α in the context of the basic model of classical test theory and beyond], Master's thesis, Ludwig-Maximilians-Universität München, Munich, Germany, 2015.

[101] H. Fisseni, *Lehrbuch der psychologischen Diagnostik [Textbook of Psychological Diagnostics]* (Hogrefe, Göttingen, 1997).

[102] G. Lienert and U. Raatz, *Testaufbau und Testanalyse [Test Construction and Test Analysis]*, 6th ed. (Verlagsgruppe Beltz, Weinheim, Germany, 1998).

[103] C. Albert and G. Pospiech, Quantenphysik in Klasse 9: Ergebnisse einer Akzeptanzbefragung für ein Spin-First-Unterrichtskonzept, PhyDid B **1**, 369 (2023).