

Design and validation of the Quantum Mechanics Conceptual Survey

S. B. McKagan,^{1,*} K. K. Perkins,² and C. E. Wieman^{3,1,2}

¹*JILA, University of Colorado and NIST, Boulder, Colorado 80309, USA*

²*Department of Physics, University of Colorado, Boulder, Colorado 80309, USA*

³*Department of Physics, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1*

(Received 12 July 2010; published 10 November 2010)

The Quantum Mechanics Conceptual Survey (QMCS) is a 12-question survey of students' conceptual understanding of quantum mechanics. It is intended to be used to measure the relative effectiveness of different instructional methods in modern physics courses. In this paper, we describe the design and validation of the survey, a process that included observations of students, a review of previous literature and textbooks and syllabi, faculty and student interviews, and statistical analysis. We also discuss issues in the development of specific questions, which may be useful both for instructors who wish to use the QMCS in their classes and for researchers who wish to conduct further research of student understanding of quantum mechanics. The QMCS has been most thoroughly tested in, and is most appropriate for assessment of (as a posttest only), sophomore-level modern physics courses. We also describe testing with students in junior quantum courses and graduate quantum courses, from which we conclude that the QMCS may be appropriate for assessing junior quantum courses, but is not appropriate for assessing graduate courses. One surprising result of our faculty interviews is a lack of faculty consensus on what topics should be taught in modern physics, which has made designing a test that is valued by a majority of physics faculty more difficult than expected.

DOI: [10.1103/PhysRevSTPER.6.020121](https://doi.org/10.1103/PhysRevSTPER.6.020121)

PACS number(s): 01.40.Fk, 01.40.gf, 03.65.-w

I. GOALS OF THE SURVEY

The Quantum Mechanics Conceptual Survey (QMCS) was designed to be used as a general survey of students' conceptual understanding of quantum mechanics. Our primary goal was to create a survey that could be used to measure the relative effectiveness of different curricula and teaching techniques, and thus to serve as a formative assessment tool for faculty engaging in reform of their courses. To achieve this goal it is important to focus on standard topics that are covered in most courses, both traditional and reformed. Therefore, the QMCS does not cover many topics that we believe are important but are not universally taught, including measurement, applications of quantum mechanics, and historical experiments. Further, to have an impact on faculty practice, the test questions must be written such that the average faculty member believes that they are measuring something important. A secondary goal in creating the survey was to provide a way to measure the prevalence of specific student difficulties. As we will discuss, we have found that this secondary goal is often in conflict with the primary goal.

The QMCS was originally intended to be applicable to all levels of quantum mechanics, from sophomore-level modern physics to graduate quantum mechanics. This intention was based on the assumption that there is a set of universally agreed-upon basic concepts that each level of quantum mechanics instruction builds upon at a deeper level. However, we have found that this is not a particularly good assumption, and we were not able to design a test that is truly applicable at all levels. The QMCS is most relevant to, and has been most thoroughly tested in, sophomore-level modern

physics courses. In these courses, we recommend giving the QMCS as a posttest at the end of a course only, and not as a pretest at the beginning of the course. As we will discuss, most students have little or no knowledge of any of the concepts measured by the QMCS at the beginning of a modern physics course, so it is both meaningless and demoralizing as a pretest.

The QMCS has also been used in junior-level and graduate-level quantum courses. However, faculty generally regard the test as too basic for these courses. In graduate courses, there is no difference between pretest and posttest scores, even for students who score low on the pretest, and there is no correlation between QMCS scores and final exam scores. These results are consistent with faculty assertions that they do not teach the concepts measured by the QMCS in graduate courses. However, graduate students in selective programs do score higher on the QMCS than do physics graduate students in less selective programs.

In keeping with standard practice for research-based conceptual tests in physics [1], we have protected the security of the test by keeping it from becoming available to students. After administering it in class we do not post solutions or allow students to take the tests home. We have administered it online to graduate students and advanced undergraduates, but always on a password-protected site with a warning not to distribute it. To further protect the security of the test, we do not include it in this article. It can be accessed from our website [2] by requesting a password from the authors.

II. REVIEW OF PREVIOUS RESEARCH

There is a long history of research-based conceptual tests of specific topics in physics, starting with the Force Concept Inventory (FCI) [3]. Widespread use of the FCI has played a particularly important historical role in spreading awareness

*Present address: American Association of Physics Teachers and McKagan Enterprises, Seattle, WA, 98144, USA.

of the deficiencies in traditional instruction in physics [4]. While many of the finer details of its validation have been criticized [5,6], defended [7–9], and studied [10–13], this test has several important strengths:

(1) It is written in everyday language without jargon or mathematics, so that students do not need to be familiar with a particular notational convention and can easily understand the questions even before taking a physics class.

(2) It focuses on conceptual understanding rather than standard problems so students cannot rely on memorized algorithms to answer the questions.

(3) Many of the distracters are very effective at eliciting students' preconceived ideas about the world and thus reveal the extent to which student learning in class has been superficial.

(4) Most faculty initially believe that the questions are "too easy," so they are startled when their students do badly (as often happens) and therefore motivated to examine their teaching.

The FCI has been used to measure the relative effectiveness of curricula in many physics courses throughout the world and has had a major impact on educational practices in introductory physics.

Many other research-based tests have been developed for introductory mechanics and other topics in physics. Some of these tests, like the FCI, are "concept inventories" that attempt to measure student understanding of a single narrowly defined concept [3,14–16]. Other tests are "surveys" that attempt to measure student understanding of a broad range of topics within a subject [17–19]. The QMCS falls into the latter category, but its range is still only a fraction of what is covered in a modern physics course.

Several research-based tests in quantum mechanics have already been developed, including the Quantum Mechanics Visualization Instrument (QMVI) [20,21], the Quantum Mechanics Concept Inventory (QMCI) [22], and the Quantum Physics Conceptual Survey (QPCS) [23]. However, all these tests serve different functions than the QMCS. While the authors of the QMVI claim that it is a general test of quantum mechanics, most of the questions focus on a single concept: the relationship between the shape of the wave function and the shape of the potential. While this is an important concept, it is not all there is to quantum mechanics, and it is possible to understand this concept quite well without having any conceptual understanding of how quantum mechanics describes reality. Further, the QMVI is extremely difficult, with many questions on topics that are not typically covered until graduate school, if at all, and we have known PhDs in physics to score less than 80% on it. The QMCI is designed to be a concept inventory rather than a survey, focusing on the concepts of "one-dimensional potential barriers, tunneling, and probability distribution," according to its author. For these particular concepts, it is much more thorough than the QMCS. The QPCS also has a narrower focus, on the concepts of wave-particle duality and the photoelectric effect.

In addition to these multiple-choice tests, Singh has developed surveys with open-ended questions in junior-level [24] and graduate-level [25] quantum mechanics. However, the purpose of these surveys appears to be mainly for research in determining student difficulties in quantum me-

chanics, rather than for comparing instructional techniques among different courses.

While some guidelines for developing conceptual tests in physics have been published [26,27], there is no consensus in the physics education research community about how conceptual tests should be validated. Many different validation techniques have been used, criticized, and defended. Some authors focus on validation through student interviews [15,22], others focus on statistical techniques [17,19,20], and others use some combination [16,18,26,27].

One issue that has been largely ignored in validation of conceptual tests is faculty buy-in. This turns out to be very important for a conceptual test in quantum mechanics. For a test to impact faculty practice in the broader physics community, rather than just being an internal tool of the physics education research community, average faculty members must believe that it is testing something they value for their students. Thus, tests that focus on identifying obscure misconceptions that faculty members do not recognize may be convincing to researchers in physics education research (PER), but will not have a significant impact on faculty practice. This is widely recognized in the development of surveys of student beliefs, where test designers interview physics faculty members in order to validate the "expert" view [28,29], but tends to be ignored in the development of conceptual tests, where test designers are more likely to assume that they know the expert view. As discussed below, we found there is large variation in faculty views on many topics in quantum mechanics.

In addition to research on the development of conceptual tests, we have drawn on research on student understanding of quantum mechanics. Previous research has explored many student difficulties in understanding quantum mechanics, including difficulty with understanding the relationship of amplitude and wavelength to potential [30–32] and with drawing correct wave functions for tunneling plane waves [30,31,33,34]. Research has also uncovered many specific incorrect beliefs common among students, such as that particles move along sinusoidal paths [22,35–38], that energy is lost in tunneling [22,30,31,33,34,39,40], and that reflection and transmission are due to particles having a range of energies [30,31,39]. QMCS questions designed to address these issues are discussed in Sec. V B.

III. DESIGN OF THE QMCS

A preliminary version of the QMCS was developed with questions based on initial ideas gathered from several sources: faculty interviews, a review of textbooks and syllabi, observations of students in lectures and study sessions, and a literature review of known student difficulties.

A. Faculty interviews

Before writing any questions for the QMCS, we interviewed eight faculty members (seven physics and one chemistry), most of whom had recently taught a course in modern physics, and asked them to list "the three most important concepts you would like your students to get out of a course

in modern physics.” The goal of these interviews was to develop a list of learning goals based on faculty consensus about what should be taught in modern physics, and then to develop QMCS questions to address each of these goals. However, the most important result of our faculty interviews is that there is *not* a faculty consensus about what should be taught in modern physics. Some faculty members responded that their primary goal in this class was not to teach concepts at all, but to teach calculational skills, history, or applications. Some faculty members listed concepts, but these concepts often did not overlap.

We grouped the concepts that did show some overlap into common themes, listed below, with the number of faculty members who gave a response related to each theme in parentheses:

- (i) wave function and probability (5),
- (ii) wave-particle duality (4),
- (iii) Schrödinger equation (4),
- (iv) quantization of states (4),
- (v) uncertainty principle (3),
- (vi) superposition (3),
- (vii) operators and observables (3),
- (viii) tunneling (2),
- (ix) measurement (2).

We note that only the first item on the list was given by more than half the faculty interviewed. While this list does not reflect the most important goals of many faculty, it is the closest we could come to a consensus on the most important topics to be covered in a modern physics. This list was used as a general guide in developing the QMCS, and we attempted to write questions to address each item in the list.

The last item in the list, measurement, was controversial. Some faculty listed measurement as one of the most important concepts in modern physics, and the authors consider it to be critical to understanding the connection between quantum mechanics and reality. However, several other faculty specifically said that measurement is *not* important and should not be taught in modern physics courses. Because we wanted the QMCS to focus on topics that are universally taught, this result, along with the results discussed below from our review of textbooks and syllabi, led us to exclude this topic from later versions of the QMCS. While some questions do indirectly address measurement, there are no questions in Version 2.0 with measurement as a primary learning goal.

B. Review of textbooks and syllabi

We reviewed textbooks and syllabi in modern physics and quantum mechanics to see what topics in modern physics are most important to experts. We found that textbooks had a great deal of overlap in terms of the types of examples and calculations they included, a result that makes sense given the limited number of solvable problems in quantum mechanics. There was much less overlap in terms of the concepts included and the degree to which conceptual understanding was emphasized at all.

Further, most textbooks included little or no discussion of the issues most frequently raised in the physics education

research literature on quantum mechanics. For example, most textbooks do not discuss whether particles move in sinusoidal paths, whether energy is lost in tunneling, or how to determine the qualitative shape of a wave function from a potential.

One fact we found particularly surprising is that not a single modern physics textbook we reviewed included any discussion of measurement, and only about half of introductory quantum mechanics discussed measurement at all. This is consistent with our results from faculty interviews that many faculty do not consider measurement to be an important concept.

Our results on measurement are consistent with results reported by Dubson *et al.* [41]. In their review of junior quantum textbooks, they found that 30% did not mention the wave function collapse and 40% made “only brief, passing mention” of it. They also found, in faculty interviews, that 11% of faculty thought that measurement should not be taught in junior quantum (they did not ask about modern physics), and an additional 26% “were deeply troubled by the idea of wave function collapse... and consequently teach it with some unease.”

C. Observations of students

Before implementing any course reforms or developing a survey, the first author observed lectures and ran an informal study session for students in an introductory modern physics course for engineering majors. The optional study session was advertised to students as an opportunity to work together and get help on their homework. The first author helped students with questions about any topic in the course, including homework and studying for exams, interacting with them directly and observing them talking to each other. She took notes about student questions and comments in lectures and study sessions, and shared these notes with the other authors. These notes were an important source of information about student ideas and difficulties with the course material, and informed both the development of the QMCS and the reforms that we implemented in the course the following semester. We wrote questions to address specific difficulties observed in lectures and study sessions, and based the wording of answers, both correct and incorrect, on quotes from students.

Study sessions and observations were continued in reformed courses in subsequent years, both by the authors and by undergraduate learning assistants who wrote weekly field notes about their interactions with students. Continuing observations were used to inform the refinement of the QMCS. We note that observations in lecture were much more informative in reformed classes in which students engaged in frequent small group discussions in response to clicker questions than in more traditional lecture courses because there were more opportunities to hear students thinking out loud.

D. Addressing known student difficulties

In addition to writing questions to address the topics that faculty believe are more important, we wrote questions to address known student difficulties, both from our own obser-

vations discussed in the previous section, and from the literature discussed in Sec. II. Many questions on the QMCS are either adapted, or taken directly, from questions in existing literature on student difficulties. For example, question 10 is adapted from the QMVI, and was included based on extensive research showing that students have difficulty relating the variance in local wavelength of a wave function to the shape of the potential [30,31,42]. Question 11 is taken directly from the QMVI. Question 7 and an earlier version of it are based on extensive research showing that students often believe that energy is lost in tunneling [22,30,31,33,34,39,40]. Question 5 is also taken from a survey used in previous research [43].

E. Design challenges specific to quantum mechanics

There are several issues that make it significantly more difficult to design a conceptual test in advanced topics such as quantum mechanics than in introductory mechanics. Thus, while we attempted to design a survey that has the strengths of the FCI, some of these strengths are not attainable.

One strength of the FCI is that it does not contain jargon and is comprehensible even to a student with no physics background. While it is possible to include more or less jargon in a test of quantum mechanics, one cannot avoid all reference to wave functions, energy levels, probability densities, and potential energy diagrams, concepts which have little meaning to students unfamiliar with quantum mechanics. This is true not only in quantum mechanics, but in many other areas of physics, such as electromagnetism, where students' low scores on pretests are often due more to a lack of knowledge of the subject than to incorrect beliefs.

Students also have far fewer preconceived ideas about quantum mechanics than about the topics covered in introductory mechanics. While students certainly have preconceived ideas about the world that quantum mechanics contradict (e.g., that a particle can have a well-defined position and momentum at the same time), they do not have preconceived ideas about wave functions and many of the entirely new concepts that are introduced in quantum mechanics. While one might imagine that students have learned about quantum mechanics through popular culture and movies, we were surprised by how infrequently students referred to any prior knowledge of quantum mechanics in interviews on the QMCS. In fact, one student, a 35-year-old engineering major, told us he had never heard that electrons could behave as waves before taking our course. The physics majors we interviewed usually had more awareness of quantum mechanics in popular culture than the engineering majors.

One result of students' lack of prior knowledge of quantum mechanics is that, at least in a modern physics course, giving the QMCS as pretest is both meaningless and demoralizing. We began giving the QMCS as both a pretest and a posttest and calculating normalized gains mainly because this is standard practice for research-based conceptual tests. However, no matter how much we emphasized to students that it was OK if they did not know the answers and they would not be graded, they often told us after taking the pretest that it was discouraging and scary. Due to this observa-

tion, and the low pretest scores discussed in the next section, we do not recommend giving the QMCS as a pretest in modern physics courses.

A more serious problem in developing a conceptual survey of quantum mechanics is that the very notion of a conceptual understanding of the this subject is controversial. By "conceptual understanding," we mean an understanding of the relationship between the mathematical description and the physical world, an understanding of what is actually happening, rather than just the ability to calculate a result. We emphasize conceptual understanding because we believe it is the basis of a true understanding of any subject. However, we recognize that this view of quantum mechanics is not universally accepted, with prominent physicists urging their students to "shut up and calculate," and even arguing that the best one can hope for in quantum mechanics is to be able to do a calculation and predict an experimental result and that understanding what is actually happening is not possible.

Much of what is typically taught in this subject involves algorithmic problem-solving skills that do not have an underlying conceptual base. For example, many faculty have the goal that their students should be able to solve the Schrödinger equation for specific abstract potentials such as a harmonic oscillator or an infinite square well. However, it is easy to memorize the method or solution for such specific problems, and difficult to apply this method to an entirely new problem. One standard technique for testing conceptual understanding in introductory physics is to put concepts in new contexts, so that students cannot rely on memorized patterns from contexts they have already studied. However, because there are so few solvable problems in quantum mechanics, it is difficult to find contexts that are likely to be new for all students. Further, most problems involve sufficiently complex mathematics that a conceptual understanding is not sufficient to solve most new problems.

A great deal of quantum mechanics involves specific "tricks" and techniques that are easy if one has seen them before, but nearly impossible to think of on one's own. For example, question 16 on the QMVI requires using symmetry to solve for the energies of a "half-well," a technique that is simple enough if one has learned it, but is not always taught. We were careful to avoid questions that tested whether students had learned a particular technique or method.

A further problem, discussed above, is that there is a lack of faculty agreement on what should be taught, especially at the modern physics level. Even when faculty used the same words to describe their goals in interviews, what they meant by those words was not always clear, and they did not necessarily value questions designed to address their stated goals. We also found that faculty did not value questions designed to elicit known student difficulties, such as the belief that particles move along sinusoidal paths, because they did not believe that such conceptual issues were relevant to solving problems in quantum mechanics. Thus, our two goals of comparing instruction that would be convincing to faculty and determining the prevalence of known difficulties were sometimes at odds.

Because of the problems discussed in this section, the final version of the QMCS contains only 12 questions, significantly fewer than we believe is ideal. Previous versions

TABLE I. Courses from which students were drawn for interviews. N is the number of students interviewed from each course. The numbers for the first three courses include students from several different semesters of the same course.

Course	N
Modern physics for engineering majors (Reformed)	24
Modern physics for engineering majors (Traditional)	8
Modern physics for physics majors (Traditional)	15
Junior quantum mechanics (Traditional)	1
Graduate quantum mechanics (Traditional)	1
Total	49

of the QMCS have contained up to 25 questions, but many of the original questions have been eliminated due to problems uncovered in student interviews, graduate student surveys, and discussions with faculty. Examples of problems with eliminated questions are discussed in Sec. V. Even the remaining 12 questions still have some problems (also discussed in Sec. V), but they are the questions that we have found to provide the most useful insight into student thinking.

The current version of the QMCS has too few questions to adequately probe student understanding of quantum mechanics. Further, the small number of questions leads to a discretization of possible scores and other problems with statistical analysis. Developing new questions is a project for future research, beyond the scope of this paper. Louis Deslauriers and Carl Wieman have developed and validated four additional questions for the QMCS, which will be described in a future publication.

IV. REFINEMENT AND VALIDATION OF THE QMCS

After developing an initial version of the QMCS as discussed in the previous section, we refined and validated it using student interviews, expert review, and finally, statistical analysis. This was an iterative process extending over approximately three years, in which we developed many different versions [44], which were administered in interviews to dozens of students and in class to hundreds of students in many different courses.

A. Student interviews

We have conducted interviews on various versions of the QMCS with 49 students. With the exception of one graduate student and four engineering majors who had completed modern physics the previous semester, all interviews were conducted either during the last week of a course in modern physics or quantum mechanics (see Table I for a breakdown of the courses), or shortly after the students completed the course. In most cases, students had already completed the survey in class before the interview (in the remainder, they saw it for the first time in the interview). In the interviews,

students were asked to talk through each question, explaining out loud which answer they had picked and why. The interviewer prompted the students to speak out loud if they were quiet and occasionally asked clarifying questions. More in-depth questions about students' thinking were saved for the end of the interview so that they would not affect students' answers to later survey questions. Each interview lasted about an hour, although a few enthusiastic students stayed up to 2 h. Students were paid \$20 for their time. This allowed us to recruit students with a wider range of abilities, rather than only the "good" students who would be willing to volunteer their time for an interview for free.

These interviews helped us to refine the wording of questions, understand student thinking, and eliminate some questions that were not serving their intended purpose. Some specific results of interviews have been discussed elsewhere [40,45]. Further results about student thinking about quantum mechanics will be discussed in Sec. V. Examples of general issues about survey design uncovered in interviews include:

(i) Students tend to avoid any response that includes words like "exactly," "definitely," or "certainly," because "nothing is ever definite in quantum mechanics." We have eliminated such words from the final version of the QMCS to ensure that students' responses are based on their understanding of the underlying concepts rather than on the wording.

(ii) Students often skim and miss the word "not" in negatively worded questions or answers. We have changed such questions to use positive wording where possible, and italicized the word "not" where it is unavoidable.

(iii) In early versions, many questions contained the option "I have no idea how to answer this question," to remove the noise of students who were guessing. However, most students were unwilling to select this option even when they frankly admitted in interviews that they had no idea how to answer the question, explaining that they preferred to guess than to admit that they did not know. When the survey was given as a posttest in class, fewer than 10% of students chose this option for any question. The few students who did pick this option in interviews often did so because they were not completely sure of their answer, even when they were able to make a reasonable educated guess. We have eliminated this option from all questions because whether students select it appears to be a function of personality rather than of whether they actually understand the question.

B. Faculty validation

After writing questions, we asked faculty members teaching courses in which we wished to give the survey to review it and give us feedback about the survey and how they expected their students to do on it. One important result from these discussions is that we did not fully succeed in our goal of making a survey like the FCI in that all faculty believe that it covers concepts that their students should (and do) know. The questions were designed to address concepts that many faculty listed as important, and most students are able to interpret the questions on the current version correctly. However, many faculty still do not necessarily value the

TABLE II. Average pretest scores in different modern physics courses for common questions asked all semesters and for Version 2.0 questions. V is the version number given in each course. We list the total number of students enrolled at the beginning of the semester, along with the number of students who took the QMCS pretest in each course. Uncertainties are standard errors on the mean. Courses in bold were given a version that included all questions in Version 2.0 and are thus included in the analysis of Version 2.0 in Sec. IV C 3.

Course	Version	No. students		Pretest scores		
		Total	QMCS	Common	V 2.0	
Modern Physics for Engineering Majors	Trad 1	1.0	No Pretest Given			
	Trad 2	1.9	69	60	$35\% \pm 2\%$	$34\% \pm 2\%$
	Ref 1	1.6	191	205	$31\% \pm 1\%$	NA
	Ref 2	1.7	184	189	$31\% \pm 1\%$	NA
	Ref 3	1.8	94	97	$34\% \pm 1\%$	$33\% \pm 1\%$
	Ref 4	1.9	156	149	$37\% \pm 1\%$	$36\% \pm 1\%$
Modern Physics for Physics Majors	Trad 1	1.2	No Pretest Given			
	Trad 2	1.6	80	60	$40\% \pm 2\%$	NA
	Trad 3	1.7	38	46	$38\% \pm 3\%$	NA
	Trad 4	1.8	69	63	$39\% \pm 2\%$	$41\% \pm 2\%$
	Ref 1	1.9	No Pretest Given			

questions, either because they do not think the concepts addressed by the questions are important, or because they misinterpret them even though students do not. Further, most faculty recognize that there are enough subtle issues in quantum mechanics that they do not expect their students to do particularly well on the QMCS, and thus are not shocked by the results in the way that they are by the results of the FCI.

We also attempted to get faculty feedback by sending out an online version of the survey asking for general feedback and ranking of the importance of questions to all physics faculty members at a large research university and to a listserv for faculty members in PER. This yielded only three responses, all of which were fairly atypical compared to the faculty we interviewed. We suspect that in addition to being very busy, faculty are unwilling to respond to such a survey for fear of getting the answer wrong. In fact, we have seen that a surprising number of faculty get some questions wrong because they do not have a correct understanding of the relevant concept. On the other hand, physics graduate students had no such qualms and were thus able to serve as an effective substitute for faculty for the “expert” response. We sent the online survey to all graduate students at the same university and listserv and received 27 responses from physics grad students, including detailed comments, within a month of sending the email. Almost half the responses came within 3 h. Many of these responses were from grad students at other universities that did not have PER groups, indicating that grad students forwarded the survey to their friends.

Grad students wrote extensively in the optional comment boxes on the online survey, which were very helpful in refining the wording of questions to make them more palatable to experts. The grad students often answered questions “incorrectly” for very different reasons than modern physics students. For example, in one question on measurement, which has since been eliminated because the topic is not

valued by most faculty, nearly half the grad students answered incorrectly because they were worrying about decoherence, a topic we had not intended to test. Questions which were too hard for most grad students to answer correctly were either modified or eliminated. For example, as will be discussed in Sec. V B 2, only 15% of grad students answered the original version of question 10 (see Fig. 7) correctly, and the question was modified to make it easier.

We have found that it is impossible to write questions to which experts cannot find obscure objections, and thus have focused on writing questions that are unambiguous to students and for which most experts can agree on the answer and importance. Even in question 1, one of the least controversial questions on the QMCS, one expert has pointed out that because the probability of emission is proportional to the energy spacing, increasing the spacing would lead to more photons being emitted from an ensemble, so E could be a correct answer. Such objections are difficult to address without adding a long list of caveats that tend to confuse students. Because we have never heard this reasoning from students, or even from more than one expert, we have left the question as it is.

C. Statistical analysis

Tables II and III and Figs. 1 and 2 show the average pretest and posttest scores, respectively, for two versions of the QMCS in different modern physics courses. The gray bars show average scores for the 12 common questions that were asked in all earlier versions of the test. The red bars show average scores for questions on the current version of the test, Version 2.0. The 12 common questions include nine questions that are included in Version 2.0 and three additional questions that have been eliminated [46]. Three of the questions in Version 2.0 (7, 10, and 12) were added relatively

TABLE III. Average posttest scores in different courses for common questions asked all semesters and for Version 2.0 questions. V is the version number given in each course. We list the total number of students enrolled at the beginning of the semester, along with the number of students who took the QMCS posttest in each course. Uncertainties are standard errors on the mean. Courses in bold were given a version that included all questions in Version 2.0 and are thus included in the analysis of Version 2.0 in Secs. IV C 3 and IV C 4.

Course	Version	No. students		Posttest scores		
		Total	QMCS	Common	V 2.0	
Modern Physics for Engineering Majors	Trad 1	1.0	80	68	$51\% \pm 2\%$	NA
	Trad 2	1.9	67	42	$53\% \pm 3\%$	$60\% \pm 3\%$
	Ref 1	1.6	189	169	$69\% \pm 1\%$	NA
	Ref 2	1.7	182	160	$65\% \pm 1\%$	NA
	Ref 3	1.8a	94	78	$63\% \pm 2\%$	$69\% \pm 2\%$
	Ref 4	1.9	153	124	$62\% \pm 1\%$	$65\% \pm 1\%$
Modern Physics for Physics Majors	Trad 1	1.2	77	66	$64\% \pm 3\%$	NA
	Trad 2	1.6	71	66	$53\% \pm 2\%$	NA
	Trad 3	1.7	43	29	$61\% \pm 4\%$	NA
	Trad 4	1.8a	65	57	$51\% \pm 2\%$	$50\% \pm 2\%$
	Ref 1	1.9	76	69	$66\% \pm 2\%$	$69\% \pm 2\%$
Junior QM	2.0 ^a	34	22	NA	$79\% \pm 4\%$	
Grad QM CU ^b	1.6	30	18	$88\% \pm 4\%$	NA	
Grad Mixed	1.5	NA	27	$93\% \pm 1\%$	NA	
Grad QM CSM 1	1.6	22	22	$74\% \pm 4\%$	NA	
Grad QM CSM 2	1.9	9	9	$88\% \pm 3\%$	$85\% \pm 4\%$	
Grad QM CSM 3	1.9	7	7	$76\% \pm 6\%$	$75\% \pm 6\%$	

^aThe test given in the Junior QM course was Version 2.0 plus three additional questions.

^bThe test was given as a pretest only in the Grad QM course at CU, but we have included it here rather than in Table II so that the data can be more easily compared to the other data for graduate students. Previous research has shown that there is no difference between pretest and posttest scores for graduate quantum courses.

late in the development of the test. Therefore, we have much more data for the common questions than for Version 2.0 questions.

We report the average scores for both the common questions and Version 2.0 questions to give a better sense of the variance of scores between semesters than one could get from the average scores for the Version 2.0 questions alone. Figures 1 and 2 show that, for courses in which we have data for both the common questions and Version 2.0 questions, the scores for the two versions are close, so that while one cannot make a rigorous quantitative comparison between the two versions, it is reasonable to compare scores from the two versions to get a rough sense of the kinds of scores that are typical in various courses.

We note that in previous publications where we have reported QMCS results [47,48], these results have been for the common questions shown here in gray. In these publications, because we were interested in looking at gains for individual courses, we reported scores for matched pre- and post-tests. In the current paper, because we are most interested in looking at pre- and post-tests individually, and because we did not give pretests in all courses, we report the average scores of all students who took the pre-test and the post-test, respectively. Because the data are not matched, the pre- and post-

test scores cannot be compared directly to each other or to the data reported in previous publications.

1. Pretest analysis

As shown in Fig. 1, modern physics students' pretest scores are only slightly higher than one would expect from random guessing. Students in the physics majors' course score slightly higher on the pretest than students in the engineering majors' course, a result that is consistent with the general belief among physics faculty that the physics majors are stronger students.

The slight variance from random guessing does not appear to be caused by a significant prevalence of prior knowledge or by attractive misconceptions in the distractors. Students do not score particularly well on any single question on the pretest, including those such as question 1 about which students might be expected to have some prior knowledge from a high school physics or chemistry course. They score slightly above random guessing on all questions, with the exception of question 9, on which the majority of students incorrectly answer that there are no allowed energy values so the overall score is worse than random guessing on this question. This question is not one on which students could be expected to have any prior knowledge or ideas, and the error

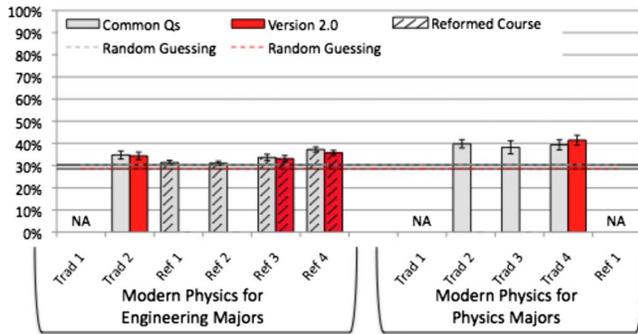


FIG. 1. (Color) Average pretest scores in different modern physics courses for common questions (gray/light) and Version 2.0 questions (red/dark). Solid colored bars indicate courses taught with traditional instruction and cross-hatched bars indicate reformed courses. Error bars show standard error on the mean. The two dashed horizontal lines show the average score that would be expected if students answered using random guessing for the common questions (gray/light) and Version 2.0 (red/dark). The chart includes some courses in which no pretest was given in order to compare more easily to Fig. 2.

can be explained by a literal interpretation of the graph: there are no allowed values inside the well because the total energy is above the well, rather than inside it. The question on which students score the highest above random guessing is question 11, on which students cannot be expected to have any prior knowledge, but on which the correct answer can be determined using test-taking skills.

We do not report pretest scores for junior quantum courses because we do not have them. We do not report pretest scores for graduate quantum courses because, as we have reported elsewhere [48], they are indistinguishable from posttest scores, even for students with low pretest scores.

2. Posttest analysis

More advanced students tend to score higher than less advanced students, with typical scores of 50–70 % for stu-

dents after modern physics courses, 80% for students after a junior-level quantum course, and 75–90 % for graduate students. When the QMCS is given pre and post, as we have reported elsewhere, students make significant gains in modern physics courses [47], and zero gains in graduate quantum courses [48]. More research is needed to determine whether the difference in average scores between modern physics students and graduate students is due to a selection effect or due to learning later in the undergraduate curriculum (presumably in junior quantum courses). In the one junior quantum course in which we have given the QMCS, there was an abbreviated pretest with only six questions. There was a significant gain (0.43) on these six questions, which suggests that students learned concepts tested by the QMCS in this course. However, this was a limited subset of questions and an atypical course including many uncommon teaching reforms [49], so more research is needed to determine whether this is a generalizable result.

The University of Colorado offers two modern physics courses, one for engineering majors and one for physics majors. The general belief among University of Colorado faculty is that the physics majors are stronger students. This belief is supported by slighter higher pretest scores on the QMCS common questions for the physics majors (38–40 %) than for the engineering majors (31–37 %). However, there is not much difference between the post-test scores for the two classes shown in Fig. 2.

Figure 2 shows scores from a variety of courses that were taught using traditional methods (solid-colored bars) and courses using reformed teaching methods (hatch-marked bars). The reformed teaching methods are described elsewhere for the modern physics courses [47], junior quantum course [49], and graduate quantum courses [48]. Early results suggested that reformed instruction in modern physics leads to higher scores on the QMCS [47]. After collecting more data, it is less clear that this is the case. As can be seen in Fig. 2, the scores for all reformed courses are higher than the scores for all traditional courses of the same type, but on average these differences are small. Students in reformed courses do score much higher than students in traditional

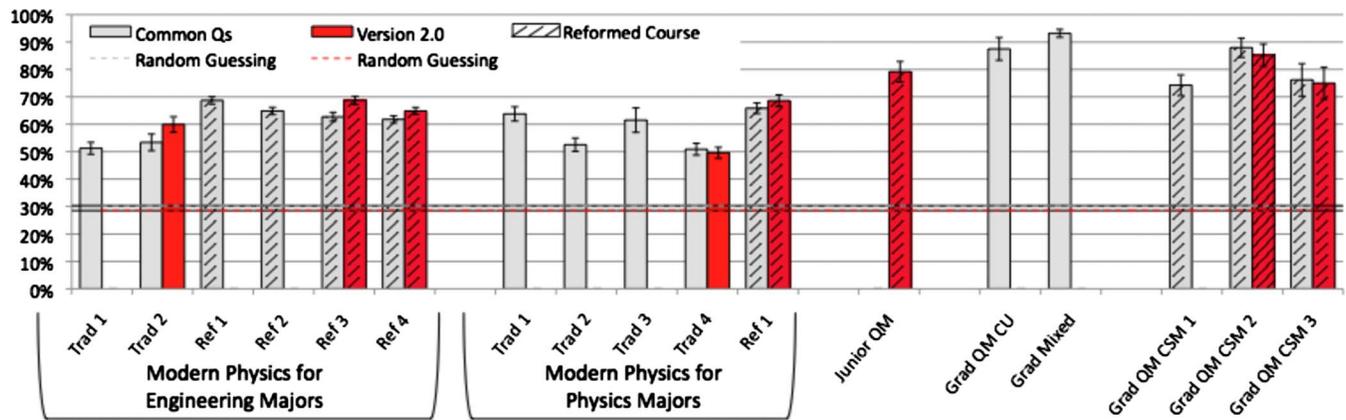


FIG. 2. (Color) Average posttest scores in different courses for common questions (gray/light) and Version 2.0 questions (red/dark). Solid colored bars indicate courses taught with traditional instruction and cross-hatched bars indicate reformed courses. Error bars show standard error on the mean. All scores are post-test scores except for Grad QM CU, which is from a pre-test for an introductory graduate quantum course at CU, and Grad mixed, which is from a survey given over the summer to graduate students from different universities who may or may not have taken a graduate quantum course.

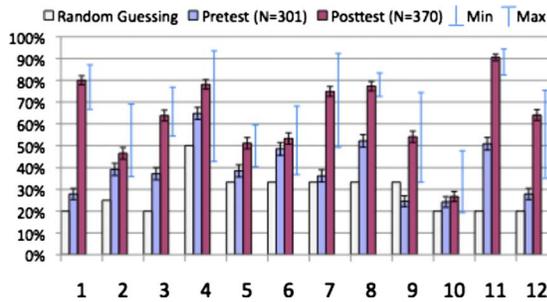


FIG. 3. (Color) Scores on individual questions in Version 2.0 in modern physics classes. This graph combines data from all five modern physics classes in which all Version 2.0 questions were given. The blue bars on the right show the minimum and maximum average posttest score for individual courses.

classes on some questions, with the largest differences on questions 4, 9, and 12. Reference [50] reports much more dramatic differences in overall QMCS scores between reformed and traditional instruction at the University of British Columbia.

3. Comparison of items on pre- and post-tests

Figure 3 shows the scores on individual questions in Version 2.0 in modern physics classes on the pretest and the posttest. We note that there is a large variation in gain across questions, with students making large gains on some questions, and little or no gains on others. In particular, we note that there is zero gain on question 6, about the uncertainty principle, and on question 10, about the qualitative relationship between wavelength and potential. These results suggest that students learn very little of what is measured by these questions in modern physics courses.

Some questions also have a large variation in posttest scores among courses, as shown by the blue bars. This range demonstrates that QMCS questions are measuring differences between courses and that different courses are achiev-

ing the learning goals measured by those questions to different extents. Examining the different teaching approaches used in the course which achieved the lowest and highest scores on each question would be informative, but is beyond the scope of this paper.

4. Item and test statistics for Version 2.0

We have computed item and test statistics that are commonly used as measures of validity and reliability of multiple-choice tests. For a more detailed discussion of the meaning and significance of these measures and their use in research-based tests in physics and other disciplines, see Refs. [27,19] and references therein.

For the remainder of the statistical analysis, we have combined posttest data from all modern physics courses in which we have given all questions in Version 2.0, that is, the five courses with red bars in the first two categories in Fig. 2. We average over all students, rather than over courses, because we are most interested in overall results.

Table IV shows the percentage of modern physics students who give each response for each question on Version 2.0. This table gives a sense of the difficulty of each question as well as the prevalence of student ideas represented by the distracters.

Table V shows selected item statistics for Version 2.0 questions for modern physics students. The item difficulty, or percent correct, gives a general sense of the difficulty of each item. (Item difficulty is also shown by the purple bars in Fig. 3 and the bold percentages in Table IV.) The QMCS contains questions with wide range of difficulties, from 0.27 for the most difficult question to 0.91 for the easiest question, with most questions falling in a range between 0.5 and 0.8.

Item discrimination is a measure of how well an item discriminates between strong students (those who do well on the test as a whole) and weak students (those who do not). It is calculated by taking the difference between the average scores of the top-scoring students and the bottom-scoring students. There are various percentages reported in the litera-

TABLE IV. Modern physics students' posttest responses to each question in Version 2.0. Correct answers are in bold.

Question	A (%)	B (%)	C (%)	D (%)	E (%)
1	2	3	6	80	9
2	39	1	46	14	
3	2	6	18	64	9
4	22	78			
5	30	51	19		
6	30	53	17		
7	2	75	23		
8	8	77	15		
9	14	31	54		
10	6	18	34	27	14
11	3	2	2	1	91
12	2	12	14	64	5

TABLE V. Item statistics for Version 2.0 questions given on posttest to modern physics students. $N=370$.

Question	Item difficulty	Discrimination	Pt. biserial coeff.
1	0.80	0.45	0.47
2	0.46	0.46	0.39
3	0.64	0.56	0.47
4	0.78	0.47	0.47
5	0.51	0.44	0.38
6	0.53	0.50	0.36
7	0.75	0.41	0.37
8	0.77	0.19	0.19
9	0.54	0.46	0.36
10	0.27	0.28	0.26
11	0.91	0.19	0.29
12	0.64	0.54	0.47
avg.	0.63	0.41	0.37

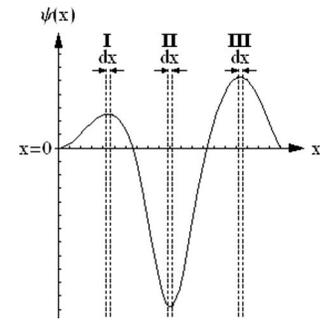
ture for how many students should be included in each group, ranging from 21% to 50%. We have chosen to use a common method of taking the difference in scores between the top and bottom 27% of students. There are various cut-offs reported in the literature for the minimum discrimination an item should have, with the most common being 0.3 [51]. However, Adams and Wieman point out that less discriminating items may serve useful purposes on a test designed to measure course effectiveness [27]. The majority of items on the QMCS have discrimination >0.4 , with the exception of three items with particularly low discrimination, questions 8, 10, and 11. The reasons for the low discrimination of questions 10 and 11 will be discussed in Sec. V.

The Pearson point biserial coefficient measures the consistency between an item and the test as a whole. A low point biserial coefficient indicates that student understanding of the concept measured by an item is not correlated with understanding of other concepts on the test. For tests designed to measure a single concept, low point biserial coefficients are undesirable and the literature says that they should be ≥ 0.2 [52]. However, as Adams and Wieman point out [27], for a test designed to measure multiple concepts, such as the QMCS, a low point biserial coefficient is not necessarily bad; rather it is an important indicator of a concept that is more difficult to learn than others. All questions on the QMCS have point biserial coefficients greater than the arbitrary cutoff of 0.2 with the exception of question 8.

Reliability statistics commonly calculated for an entire test include the Kuder-Richardson reliability index (which, if questions are graded as only correct or incorrect, is equivalent to the Cronbach Alpha coefficient), and Ferguson's delta.

The Kuder-Richardson reliability index, or Cronbach Alpha, is a measure of the overall correlation between items [53,54]. It is commonly stated that should be above the cutoff value of 0.7 for an instrument to be reliable [51]. However, Adams and Wieman argue that instruments designed to measure multiple concepts may have low Cronbach Alpha because these concepts may be independent [27]. This is the case for the QMCS, which has a Cronbach Alpha of 0.44.

11. The plot at right shows a snapshot of the spatial part of a one-dimensional wave function for a particle, $\psi(x)$, versus x . $\psi(x)$ is purely real. The labels, I, II, and III, indicate regions in which measurements of the position of the particle can be made. Order the probabilities, P , of finding the particle in regions I, II, and III, from biggest to smallest.



- A. $P(\text{III}) > P(\text{I}) > P(\text{II})$
 B. $P(\text{II}) > P(\text{I}) > P(\text{III})$
 C. $P(\text{III}) > P(\text{II}) > P(\text{I})$
 D. $P(\text{I}) > P(\text{II}) > P(\text{III})$
 E. $P(\text{II}) > P(\text{III}) > P(\text{I})$

FIG. 4. Question 11, taken from the QMVI, tests student understanding of the relationship between wave function and probability density.

Ferguson's delta is an indication of the discrimination of the test as a whole, as measured by the distribution of scores in the sample [55]. It is commonly stated that for a test to have good discrimination, Ferguson's delta should be ≥ 0.9 [52]. Ferguson's delta for the QMCS is 0.93.

V. VALIDITY ISSUES WITH SPECIFIC QUESTIONS

In discussions with faculty and graduate students who are interested in using the QMCS, both for teaching and for research, we have found that many of the issues that cause problems for students are not obvious to experts. In many cases, the questions that are most appealing to experts may not be the best questions for testing student understanding, either because they are not testing conceptual understanding, or because they are testing multiple concepts, or because students may interpret them in a way that experts do not expect. In this section, we discuss some of the issues that have come up in developing specific questions, including some questions that have been eliminated from Version 2.0. We hope that this discussion will aid instructors in interpreting their students' results and in modifying the QMCS appropriately for different instructional environments, and will aid researchers in using QMCS questions for research and in designing new questions to test students' conceptual understanding of quantum mechanics.

A. Questions designed to address important concepts

The QMCS contains many questions that were designed to address concepts that experts believe to be important. It has been difficult to design such questions for many reasons, including a lack of conceptual basis for the ideas that experts most value and a lack of faculty agreement on interpretation.

1. Relationship of probability density to wave function

The majority of faculty we interviewed stated that understanding wave functions and probability is one of the most important concepts for students in modern physics. Question 11, shown in Fig. 4, tests student understanding of the relationship between wave function and probability density. Most faculty who have reviewed the survey have commented

that this is a good question, and it is the question most like a typical homework or exam question.

However, this question is very easy, with 91% of students answering it correctly, and it shows little discrimination between strong and weak students. One can answer it simply by memorizing the statement, “the probability is the square of the wave function,” without any understanding of what is meant by probability or the wave function. In interviews with students currently enrolled in modern physics courses, nearly all recited something close to this phrase, circled the correct answer, and quickly moved on to the next question. The most time-consuming part of answering the question was simply searching through the responses to find the one that matched the answer they had already determined.

Out of four interviews conducted with students who had completed a modern physics course five months earlier, two struggled to remember whether the probability density was related directly to the wave function, or to the absolute value or square of the wave function. We never observed any such confusion among the remaining interview students who were either currently enrolled in modern physics or had completed it within the last two weeks. On the other questions on the survey, the four students who had completed the course much earlier had no more trouble than the rest of the students. While the sample size is small, these results suggest that students may memorize this relationship because it is repeated so often during the course, but do not understand its significance well enough to retain it in long-term memory.

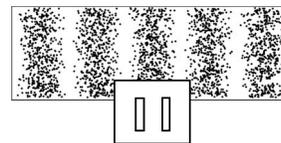
Question 11 is also the only question on which students score more than 20% higher than random guessing on the pretest; 51% of students answer this question correctly on the pretest, 31% more than would be expected if students were guessing randomly. Since it is highly unlikely that students would have learned the relevant concepts before starting a modern physics course, these results suggest that many students are able to use other reasoning skills to arrive at a correct answer without any understanding of the relevant physics. Since there are only two quantities mentioned in the problem, the most plausible guess is that a higher magnitude of one gives a higher magnitude of the other. The only other plausible guess is that it is the values, rather than the magnitudes, that are related, which would lead to answer A. This was the second most common answer on the pretest, selected by 21% of students. It is also possible to distinguish between these two plausible guesses without knowing any quantum mechanics simply by knowing that probability cannot be negative.

In spite of these issues, we have retained question 11 because it tests a concept that faculty consider very important and because students make significant gains on it between the pretest and posttest, which shows that they are learning something from the course. Adams and Wieman suggest that it is important for a test to contain items which satisfy these two criteria, so that faculty can see that the test measures student learning that they value.

2. Wave-particle duality

Wave-particle duality is another concept that many faculty in our interviews considered very important. However, be-

12. You shoot a beam of photons through a pair of slits at a screen. The beam is so weak that the photons arrive at the screen one at a time, but eventually they build up an interference pattern, as shown in the picture at right. What can you say about which slit any particular photon went through?



- Each photon went through either the left slit or the right slit. If we had a good enough detector, we could determine which one without changing the interference pattern.
- Each photon went through either the left slit or the right slit, but it is fundamentally impossible to determine which one.
- Each photon went through both slits. If we had a good enough detector, we could measure a photon in both places at once.
- Each photon went through both slits. If we had a good enough detector, we could measure a photon going through one slit or the other, but this would destroy the interference pattern.
- It is impossible to determine whether the photon went through one slit or both.

FIG. 5. Question 12 tests student understanding of the wave-particle duality.

cause this concept involves many subtle issues of interpretation, it is extremely difficult to test in a way in which experts can agree [41]. Question 12, shown in Fig. 5, is an attempt to address this important concept. Although we have attempted many versions of this question, we still have not been able to find any version for which all physics professors agree with us on the “correct” answer. The current version is an improvement over previous versions in that the majority of physics professors agree that our answer is the “best” answer of those given, even though many of them do not like it.

We believe that the correct answer is D because the wave function must go through both slits to get an interference pattern even though a detector can only measure a photon in one place. If one solves the Schrödinger equation with a wave function going through only one slit, one does not get an interference pattern, so it is not correct to say that the photon went through only one slit, or even that we do not know whether it went through one slit or both. In our wording, we have equated the photon with the wave function (or more precisely, with the electromagnetic wave, since a photon does not technically have a wave function), which we recognize is a particular choice of interpretation. However, without this choice, there does not appear to be any wave-particle duality, and thus no way of testing this concept.

In interviews with 27 physics faculty, Dubson *et al.* found wide disagreement on the correct interpretation of the wave function, with 48% of faculty interpreting the wave function as an information wave, 30% interpreting it as a matter wave, and the remainder holding some kind of mixed view [41]. Given these results, it is not surprising that we have had difficulty finding faculty consensus on any version of this question. Our wording favors a matter wave interpretation, as defined by Dubson *et al.* Faculty who prefer an information wave interpretation have argued that there is no correct answer, because while it is fair to say that the *wave function* goes through both slits, one should not say anything at all about the *photon*. Unfortunately, changing the wording of the question to satisfy these faculty would eliminate the fundamental issue of wave-particle duality from the question, making it merely about the mathematical details of the wave function.

Another problem with interpretation with this question is that the small minority of faculty who prefer Bohm’s inter-

pretation of quantum mechanics choose B as the correct answer. Further, a small minority of physics professors argue for answer E, either due to an agnosticism on interpretation or a lack of understanding of the implications of Bell's inequality.

In interviews, on the other hand, modern physics students never argue for B or E due to deep issues of interpretation, but only due to a clear misunderstanding of the implications of quantum mechanics. Further, although some previous versions of the question were difficult for students to interpret, the current version does not cause problems for students in terms of simply understanding what the question is asking. Many students state answer D almost verbatim even before reading the options, so it accurately reflects the way that students think about this question.

B. Questions designed to elicit student difficulties

Questions designed to elicit specific student difficulties cited in the literature have been problematic from the perspective of faculty buy-in. We have found that most faculty do not recognize the importance of such questions, either because they do not believe the underlying concept is important, or because they do not recognize the seriousness of the student misunderstanding. Below we discuss several examples of such questions and the issues that they have raised.

1. Belief that particles travel along sinusoidal paths

Question 4, which asks whether photons travel along sinusoidal paths, began as a distracter for a question on the meaning of the wave-particle duality. After seeing how often students in interviews responded that both photons and electrons do travel along sinusoidal paths, we discovered that many other researchers have reported on this difficulty [22,35–38]. This particular student difficulty is unlike many of the difficulties reported in introductory physics in that it is not due to prior intuition, but due to instruction. This is the only question where students sometimes do worse (often much worse) on the post-test than on the pretest. An example that illustrates how students might get this idea from instruction can be seen from an interview with a student who had just completed a modern physics course for physics majors. He adamantly agreed with the statement that photons travel in sinusoidal paths, saying that his professor was always saying, “photons move like this,” and moving his index finger in a sinusoidal path.

This question has been problematic from the perspective of faculty buy-in. Most faculty members who have reviewed the survey have not understood the purpose of this question because it does not occur to them that students might actually be thinking of photons and electrons as moving along sinusoidal paths. Furthermore, faculty members, graduate students, and advanced undergraduates often pick the “wrong” answer for this question because they find the literal statement so nonsensical that they automatically translate it in their heads to something like, “the wave function of the photon is sinusoidal,” which is a true statement. (We added the qualifying phrase “in the absence of external forces” because several graduate students pointed out that electrons

500 nm wavelength light is emitted from point 1. A photon is detected 2500 nm away at point 2. Which of the following drawings most accurately represents the trajectory that the photon followed from point 1 to point 2?

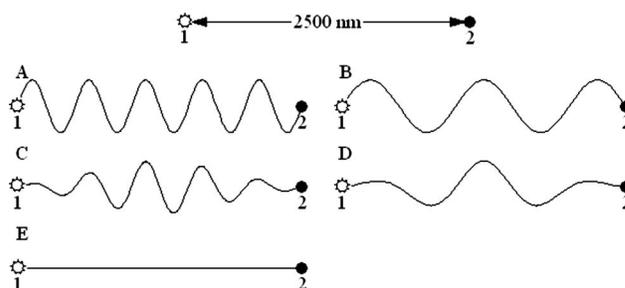


FIG. 6. An alternative version of question 4, taken from Knight. This version does not work well because too many students do not notice the word “trajectory.”

could travel in sinusoidal paths if the right forces were applied, for example in an undulator.)

However, the modern physics students we have interviewed almost never misinterpret the question. Of the 46 modern students we interviewed on this question, only two misinterpreted the statement to mean that the wave function or electromagnetic field is sinusoidal and therefore gave the incorrect answer even though they understood the concept. The remainder of the students, regardless of whether they thought the statement was true or false, interpreted it as meaning that electrons literally travel in sinusoidal paths. Many students moved their fingers or hands in a sinusoidal motion to illustrate the path. Thus, while the question is not valid for advanced students and experts due to interpretation problems, it does appear to be valid for modern physics students. We have attempted many different wordings of the question in order to avoid interpretation problems, but have not found any that work better than the current wording. In some versions of the QMCS, we tested out the question shown in Fig. 6, taken from Knight [38], which gets at the same concept with pictures rather than words. We found that this question was not valid even for modern physics students, as many students who correctly answered question 4 answered this question incorrectly, reporting that they missed the word “trajectory” or that they saw numbers and automatically went into calculation mode without reading the rest of the question.

2. Difficulty understanding relationship of amplitude and wavelength to potential

There is an extensive literature of physics education research demonstrating that students have difficulty understanding the qualitative relationships between the potential energy function and both the local wavelength and the amplitude of the wave function [30–32]. However, after reviewing textbooks and syllabi, we suspected that this difficulty might be due simply to a lack of coverage in standard courses, rather than to any deep underlying conceptual difficulty. With notable exceptions in the textbooks of French and Taylor [56] and of Robinett [57], these concepts do not appear to be covered in any depth in standard textbooks or syllabi of either modern physics or quantum mechanics courses.

The figure on the left below shows a potential energy function $U(x)$, where the potential energy is infinite if x is less than 0 or greater than L , and has a slanted bottom in between 0 and L , so that the potential well is deeper on the right than on the left. Which of the plots of $|\psi(x)|^2$ vs. x is most likely to correspond to a stationary state of this potential well?

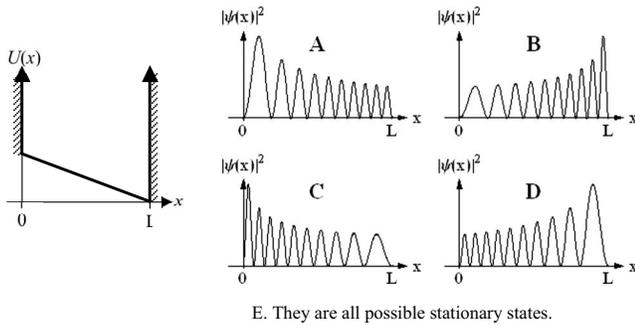


FIG. 7. The original version of question 10, taken from the QMVI. In the current version this question has been adapted in order to make it easier, by including only answers with larger magnitude on the left side. The correct answer is A.

The original version of question 10, taken from the QMVI, shown in Fig. 7, was included in early versions of the QMCS in order to test these difficulties. The original question required understanding two separate concepts: that the local deBroglie wavelength is shorter in regions of lower potential energy (because $KE = E - PE$ is larger), and that the amplitude is larger in regions of lower potential energy (because larger KE means the particle spends less time there). Each concept can be evaluated separately by looking at the percentage of students who answered either A or B (wavelength correct) and the percentage of students who answered either A or C (amplitude correct).

Our results suggest that these concepts are not well-understood by students or by experts. Figure 8 shows that modern physics students perform extremely poorly on this question, as one might expect from the results of previous research. An unexpected result was the poor performance of experts. Only 15% of graduate students were able to answer this question correctly, on par with modern physics students, and many faculty, including some very well-known and respected physicists, have answered it incorrectly or said that they do not know how to answer it.

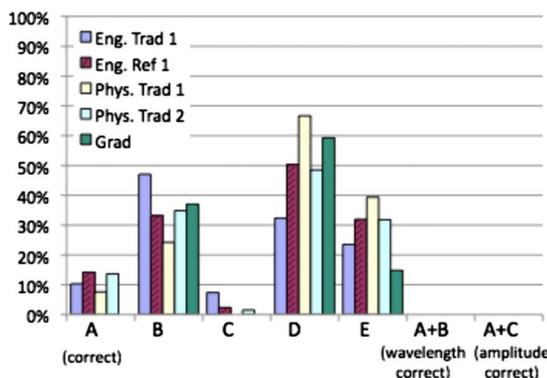


FIG. 8. (Color) Responses to the original slanted well question shown in Fig. 7 for students in four modern physics classes and for graduate students.

In spite of this poor performance, interview results suggest that the underlying concepts are not particularly difficult. When the interviewer reviewed the answers with students at the end of interviews, students who had initially gotten it wrong were almost always able to figure out the correct answer after a few leading questions from the interviewer (e.g., Is the kinetic energy larger where the potential energy is low or where it is high? If you have a pendulum swinging back and forth, does it spend more time in the regions where it moves quickly or in the regions where it moves slowly?). For most other questions on the QMCS, if students failed to answer them correctly on their own, it required much more extensive guidance to help them understand the correct answer.

Figure 8 shows that students in one modern physics course scored significantly higher on this question than both other modern physics students and graduate students. In interviews with nine students from this course, we learned that the professor had emphasized the concepts in this question in lecture. Eight out of nine students mentioned specific details from his lectures, such as that he frequently said that “the curvature encodes the kinetic energy” and that he used a pendulum example to explain why the amplitude is higher when the potential energy is higher. Even students who answered the question incorrectly remembered these lectures, but misremembered or misunderstood some aspect of them. Further, the students who answered correctly were transferring knowledge to a new situation: One student said in an interview that the professor had shown an example with a step in the well, not a slanted well, and that he was not sure that what he had said applied to a continuously varying potential but he thought it probably did. Students from other courses did not mention lectures when answering this question, although one of these students got the wavelength correct because he clearly remembered a homework problem designed to address this concept, which had been the only such instruction in the entire course. Thus, our results indicate that students’ inability to answer this question is due mainly to a lack of exposure, and it may be possible to remedy this with minimal instruction.

The amplitude concept in this question appears to be more difficult than the wavelength concept. This can be seen from the responses shown in Fig. 8, and from our student interviews and discussions with faculty, in which both groups had trouble with both concepts, but much more difficulty with the amplitude. While grad students performed as poorly as modern physics students on the question as a whole, they were more likely than most groups of modern physics students to give answers that included the correct wavelength. In our review of textbooks, we found that most modern physics textbooks have at least a homework question or two that address the qualitative relationship between wavelength and potential, but do not address the qualitative relationship between amplitude and potential at all.

Because the concept of amplitude addressed in this question is so poorly understood even by experts and is not covered in most textbooks, we modified the question to the current question 10, which tests only student understanding of wavelength, and not amplitude. In interviews with the current version, students are often bothered by the fact that the

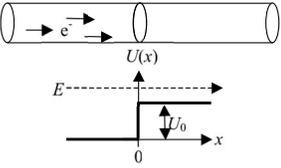
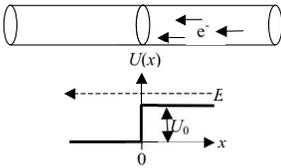
1. A beam of electrons that all have the same energy E are traveling through a conducting wire. At $x = 0$, the wire becomes a different kind of metal so that the potential energy of the electrons increases from zero to U_0 . If $E > U_0$, which statement most accurately describes the transmission and reflection of electrons?
- 
- A. All the electrons are transmitted because they all have $E > U_0$.
 B. Some of the electrons are transmitted and some are reflected because they actually have a range of energies.
 C. Some of the electrons are transmitted and some are reflected because they behave as waves.
 D. All of the electrons are reflected because they prefer to be in the region with lower potential energy.
2. Suppose that the physical situation is the same as in the previous question, but the electrons are traveling from the right to the left. If $E > U_0$, which statement most accurately describes the transmission and reflection of electrons?
- 
- A. All the electrons are transmitted because they all have $E > U_0$.
 B. Some of the electrons are transmitted and some are reflected because they actually have a range of energies.
 C. Some of the electrons are transmitted and some are reflected because they behave as waves.
 D. All of the electrons are transmitted because the potential energy drops, so there is no reason for them to be reflected.

FIG. 9. Two questions designed to elicit the belief that reflection and transmission are due to a range of energies, adapted from Ref. [30]. The correct answer to both questions is C.

amplitude is larger on the left, but they answer the question based on their understanding of the wavelength.

The modified question is still the most difficult question on the QMCS, shows little discrimination between weak and strong students, and shows no improvement between pretest and posttest. In spite of these issues, we have retained question 10 because it successfully illustrates that the relationship between wavelength and potential is not well understood by most students. Furthermore, as demonstrated by the one course in which we saw improvement on the original version of this question, it appears that it is possible to make significant gains in understanding of this concept with minimal but targeted instruction, and that this question can be used to test such instruction. We encourage others to use this question to test the effectiveness of different types of instruction for teaching this concept.

3. Belief that reflection and transmission are due to a range of energies

Figure 9 shows two questions designed to elicit the belief that reflection and transmission are caused by particles having a range of energies rather than due to the wave nature of matter, another student difficulty previously discussed in the physics education research literature [30,31,39]. The results of interviews with the first question have been discussed elsewhere [40], and will be summarized only briefly here: Students often believe that reflection and transmission are caused by a range of energies. Students with this belief may select answer B, which characterizes this belief, or A, stating that in this case there will not be reflection because there is not a range of energies. This is consistent with results that have been reported by other researchers.

However, while other researchers have attributed this error to “treating electrons as particles rather than waves,” our results are not consistent with this interpretation. Rather, we have consistently seen students make this error while clearly describing electrons as waves. As a result of this observation, we started asking students who had answered this question incorrectly in interviews to answer additional questions about what would happen to a wave pulse traveling along a rope that is heavier on one side than the other, and/or a beam of light traveling from air to glass and vice versa. We found that many students answered these questions incorrectly as well, indicating that the problem is not specific to quantum mechanics. Many students do not appear to view the behavior of transmitting only when there is enough energy as a property of particles, but of common sense. Thus, the problem appears to be that students do not understand wave behavior, rather than that they view electrons as particles. This distinction has important pedagogical consequences because it means that simply emphasizing the wave nature of electrons is not sufficient to address this error.

A further problem with these questions is that we have found that the “correct” answer (C) is somewhat unsatisfying to both faculty and students who do *not* have the incorrect belief that partial reflection is caused by a range of energies. Students sometimes reported that they chose C because all the other answers were clearly wrong, but it didn’t really seem like an explanation. For example, one student gave the following explanation for his answer: “I was pretty sure that some of the electrons were reflected and some were transmitted, and... somehow that has to do with them behaving as waves. And since that was all the question asked, I was like well, that’s about the level I know.” Another student said, “Because they behave as waves? That doesn’t really do a good job of explaining anything!” Several faculty members also reported discomfort with the answer, arguing that while it was true, it was not really an explanation. However, neither faculty nor students were able to offer a more satisfactory explanation. The original wording, in an early version that was tested with only seven students, was “because of the interaction with the potential step” rather than “because they behave as waves.” This wording was even more confusing to students, who were unable to interpret the meaning of the statement at all.

We have removed these questions from Version 2.0 for two reasons. First, we found in discussions with faculty that a substantial fraction of them did not cover the concepts addressed in these questions in their modern physics classes. Because this topic is not universally covered, the questions are not useful for comparing different courses because they measure whether the topic was covered rather than the quality of instruction. Second, our goal in the QMCS is to measure *conceptual* understanding of quantum mechanics, and we found in interviews that answering these questions does not require conceptual understanding. Students who answered correctly usually did so from memory, and did not display significantly deeper conceptual understanding than those who did not.

4. Difficulty drawing correct wave functions for tunneling plane waves

Another difficulty commonly cited in the literature is the inability to draw correct wave functions for tunneling plane

waves [30,31,33,34]. Commonly cited mistakes are drawing the transmitted part of the wave function with an offset or a smaller frequency, and drawing the tunneling part of the wave function as missing or sinusoidal. As discussed in detail in Ref. [45], early versions of the QMCS included a question designed to address this difficulty, but this question was eliminated because we saw in interviews that answering it correctly did not require any conceptual understanding, only memory.

5. *Belief that energy is lost in tunneling*

Students often believe that energy is lost in tunneling [22,30,31,33,34,39,40]. This difficulty, like the previous two difficulties, is quite difficult to assess in a conceptual way. The original question designed to address this difficulty, discussed in great detail in Refs. [40,45], does require conceptual understanding to answer, but we found that it is measuring too many things at once and, especially for students in reformed courses, answering it incorrectly does not necessarily indicate a belief that energy is lost in tunneling. To ensure that the question is measuring what we intend, we have replaced it with question 7, which asks more directly about energy loss in tunneling. The new question requires much less conceptual understanding, but it is easier to interpret the results. We have retained it because this is an important concept that students often have trouble understanding.

VI. SUGGESTIONS FOR USING THE QMCS

The QMCS is intended to be used as a formative assessment tool for faculty to measure the effectiveness of different teaching methods at improving students' conceptual understanding of quantum mechanics, and to use such measurements to improve their teaching. We believe that it is useful for this purpose in modern physics courses, and it may be useful for this purpose in junior quantum courses.

Version 2.0 of the QMCS should take about 20 min for students to complete.

We recommend giving the QMCS in class as a paper test with scantron answer sheets. To protect the security of the test, it is important to collect all tests and answer sheets, and not to post results, so that future students cannot use the test to study.

We have also administered the QMCS online on a password-protected site that is available for a limited time with security measures to prevent copying and pasting. We have found that this method works well for graduate students, who are typically motivated by the material and find the test intrinsically interesting. For modern physics students, administering the test online gives unacceptably low response rates. Online administration might work for students in a junior quantum course, but we have not tried it.

To reduce students' motivation to keep the test to study, we do not recommend basing any part of their course grades on their scores on the QMCS. We do recommend offering a small amount of participation credit to encourage them to take it.

A. Modern physics courses

The QMCS has been most thoroughly tested at this level, with extensive interviews and statistical analysis of results with students in a wide variety of courses at the University of Colorado, including courses for engineering majors and physics majors, and traditional and reformed courses.

We recommend giving the QMCS as a posttest only in modern physics classes. Students cannot be expected to know any of the content before such a class, and we have found that giving the QMCS as a pretest to modern physics students is extremely demoralizing.

We have found that a good way to motivate students to take the QMCS seriously is to give it on the second-to-last day of class as a "final exam prep." We then spend the last day of class going over the answers to questions on which students did particularly poorly. Students report that they find it useful when administered in this way. We encourage others to administer the QMCS in modern physics courses and to use the results to inform their teaching, as well as to publish them to inform the broader community.

B. Junior quantum courses

Very little testing of the QMCS has been done at the junior quantum course level. Our preliminary results suggest that it is a useful test at this level, and should be given as a pretest and a posttest in order to examine learning gains. However, further research is needed to determine typical scores that should be expected and to test the validity of the questions for students at this level. We encourage others to administer the QMCS in junior quantum courses, to use the results to inform their teaching, and to report their results.

C. Graduate quantum courses

The QMCS has been administered to dozens of graduate students in an online format allowing for optional comments on each question. We have used these comments to refine the test.

The QMCS does distinguish between students in more or less competitive graduate programs, and it may be useful for assessing graduate students' understanding of the most basic conceptual ideas in quantum mechanics. However, when we have administered it as a pretest and posttest in graduate courses, we find zero gains, even for students with very low pretest scores. The QMCS does not test any of the more advanced ideas that students are expected to learn in graduate quantum mechanics, and thus is not useful for assessing such courses.

VII. SUGGESTIONS FOR FURTHER RESEARCH

We have studied the validity and application of the QMCS extensively in modern physics and graduate quantum courses. However, because of the particular difficulties associated with conceptual assessment of quantum mechanics discussed in Sec. III E, further research is needed to develop ideal assessment tools for this subject. We encourage others to test the QMCS at different institutions and levels, as well as to develop new questions [50].

ACKNOWLEDGMENTS

We thank the Physics Education Research Group at the University of Colorado for extensive feedback on all aspects of the development of this survey. This work was supported by the NSF and the University of Colorado Science Education Initiative.

APPENDIX: LEARNING GOALS FOR QMCS

Below are the specific learning goals for each question on Version 2.0 of the QMCS, with the general concepts tested by each question in parentheses:

(1) Apply the concept of quantized energy levels to determine the relationship between energy and wavelength in an atomic transition (wave-particle duality, quantization of states).

(2) Interpret the solutions of the Schrödinger equation to recognize that electron energy levels are spread out in space (wave-particle duality, wave function and probability, uncertainty principle, operators and observables, measurement).

(3) Apply the concept of quantized energy levels to calculate the energies of photons emitted in transitions (quantization of states).

(4) Recognize that the wave-particle duality does NOT imply that particles move in wavelike trajectories. (wave function and probability, wave-particle duality).

(5) Apply the deBroglie relationship to determine the wavelength of a particle from its properties of motion (wave function and probability, wave-particle duality).

(6) Apply the uncertainty principle to determine the possible outcomes of a measurement (uncertainty principle, superposition, operators and observables, measurement).

(7) Recognize that energy is not lost in tunneling (tunneling).

(8) Use the Schrödinger equation to determine when energy is quantized (Schrödinger equation, quantization of states).

(9) Use the Schrödinger equation to determine when energy is not quantized (Schrödinger equation, quantization of states).

(10) Determine qualitatively the shapes of solutions to the Schrödinger equation (Schrödinger equation)

(11) Given a wave function, determine the probability distribution (wave function and probability).

(12) Describe the model that quantum mechanics gives to explain the double slit experiment (wave function and probability, measurement).

-
- [1] Edward F. Redish, *Teaching Physics with the Physics Suite* (John Wiley & Sons Inc., Somerset, 2003).
- [2] <http://per.colorado.edu/QMCS>.
- [3] D. Hestenes, M. Wells, and G. Swackhammer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- [4] R. R. Hake, Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
- [5] D. Huffman and P. Heller, What does the force concept inventory actually measure? *Phys. Teach.* **33**, 138 (1995).
- [6] P. Heller and D. Huffman, Interpreting the force concept inventory: A reply to Hestenes and Halloun, *Phys. Teach.* **33**, 503 (1995).
- [7] D. Hestenes and I. Halloun, Interpreting the force concept inventory: A response to March 1995 critique by Huffman and Heller, *Phys. Teach.* **33**, 502 (1995).
- [8] Ibrahim Halloun and David Hestenes, "The search for conceptual coherence in FCI data," (1995), <http://modeling.asu.edu/R&E/CoherFCI.pdf>
- [9] C. Henderson, Common concerns about the Force Concept Inventory, *Phys. Teach.* **40**, 542 (2002).
- [10] R. Steinberg and M. Sabella, Performance on multiple-choice diagnostics and complementary exam problems, *Phys. Teach.* **35**, 150 (1997).
- [11] Laura McCullough and David Meltzer, *2001 Physics Education Research Conference Proceedings* (PERC Publishing, Rochester, New York, 2001).
- [12] S. Rebello and D. Zollman, The effect of distracters on student performance on the Force Concept Inventory, *Am. J. Phys.* **72**, 116 (2004).
- [13] M. Dancy and R. Beichner, Impact of animation on assessment of conceptual understanding in physics, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010104 (2006).
- [14] R. Beichner, Testing student interpretation of kinematics graphs, *Am. J. Phys.* **62**, 750 (1994).
- [15] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula, *Am. J. Phys.* **66**, 338 (1998).
- [16] P. V. Engelhardt and R. J. Beichner, Students' understanding of direct current resistive electrical circuits, *Am. J. Phys.* **72**, 98 (2004).
- [17] D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke, and A. Van Heuvelen, Surveying students' conceptual knowledge of electricity and magnetism, *Am. J. Phys.* **69**, S12 (2001).
- [18] C. Singh and D. Rosengrant, Multiple-choice test of energy and momentum concepts, *Am. J. Phys.* **71**, 607 (2003).
- [19] L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010105 (2006).
- [20] E. Cataloglu and R. W. Robinett, Testing the development of student conceptual and visualization understanding in quantum mechanics through the undergraduate career, *Am. J. Phys.* **70**, 238 (2002).
- [21] E. Cataloglu, Ph.D. thesis, Pennsylvania State University, 2002.
- [22] J. Falk, Master's thesis, Uppsala University, 2004.
- [23] S. Wutiprom, M. D. Sharma, I. Johnston, R. Chitaree, and C. Soankwan, Development and Use of a Conceptual Survey in Introductory Quantum Physics, *Int. J. Sci. Educ.* **31**, 631

- (2009).
- [24] C. Singh, Student understanding of quantum mechanics, *Am. J. Phys.* **69**, 885 (2001).
- [25] C. Singh, Student understanding of quantum mechanics at the beginning of graduate instruction, *Am. J. Phys.* **76**, 277 (2008).
- [26] Rebecca S. Lindell, Elizabeth Peak, and Thomas M. Foster, 2006 *Physics Education Research Conference Proceedings*, edited by L. McCullough, P. Heron, and L. Hsu (AIP Press, Melville, NY, 2007).
- [27] W. K. Adams and C. E. Wieman, "Development and validation of instruments to measure learning of expert-like thinking," *Int. J. Sci. Educ.* (to be published).
- [28] E. Redish, J. Saul, and R. Steinberg, Student expectations in introductory physics, *Am. J. Phys.* **66**, 212 (1998).
- [29] W. K. Adams, K. K. Perkins, N. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010101 (2006).
- [30] B. Ambrose, Ph.D. thesis, University of Washington, 1999.
- [31] L. Bao, Ph.D. thesis, University of Maryland, 1999.
- [32] Homeyra R. Sadaghiani, Ph.D. thesis, The Ohio State University, 2005.
- [33] Jeffrey T. Morgan, Michael C. Wittmann, and John R. Thompson, 2003 *Physics Education Research Conference Proceedings*, edited by S. Franklin, K. Cummings, and J. Marx (AIP Press, Melville, NY, 2004).
- [34] M. C. Wittmann, J. T. Morgan, and L. Bao, Addressing student models of energy loss in quantum tunneling, *Eur. J. Phys.* **26**, 939 (2005).
- [35] R. N. Steinberg, G. E. Oberem, and L. C. McDermott, Development of a computer-based tutorial on the photoelectric effect, *Am. J. Phys.* **64**, 1370 (1996).
- [36] R. Müller and H. Wiesner, Teaching quantum mechanics on an introductory level, *Am. J. Phys.* **70**, 200 (2002).
- [37] R. V. Olsen, Introducing quantum mechanics in the upper secondary school: a study in Norway, *Int. J. Sci. Educ.* **24**, 565 (2002).
- [38] R. Knight, *Five Easy Lessons: Strategies for Successful Physics Teaching* (Addison Wesley, San Francisco, 2004).
- [39] D. Domert, C. Linder, and A. Ingerman, Probability as a conceptual hurdle to understanding one-dimensional quantum scattering and tunneling, *Eur. J. Phys.* **26**, 47 (2005).
- [40] S. B. McKagan and C. E. Wieman, 2005 *Physics Education Research Conference Proceedings*, edited by P. Heron, L. McCullough, and J. Marx (AIP Press, Melville, NY, 2006).
- [41] M. Dubson, S. Goldhaber, S. Pollock, and K. Perkins, 2009 *Physics Education Research Conference Proceedings*, edited by C. Henderson, M. Sabella, and C. Singh (AIP Press, Melville, NY, 2009).
- [42] Bradley S. Ambrose, 2004 *Physics Education Research Conference Proceedings*, edited by J. Marx, P. Heron, and S. Franklin (AIP Press, Melville, NY, 2005).
- [43] Peter Robert Fletcher, Ph.D. thesis, University of Sydney, 2004.
- [44] Old versions are available at <http://per.colorado.edu/QMCS>
- [45] S. B. McKagan, K. K. Perkins, and C. E. Wieman, "Deeper look at student learning of quantum mechanics: The case of tunneling," *Phys. Rev. ST Phys. Educ. Res.* **4**, 020103 (2008).
- [46] The three eliminated questions include the two questions about reflection and transmission at a potential step shown in Fig. 9, which were eliminated because we found that many faculty do not cover this topic in modern physics, and one question asking whether photons travel in sinusoidal paths, which was eliminated because we did not want to have two questions about a single topic that many faculty found of questionable value. See Sec. V B for discussion of eliminated questions.
- [47] S. B. McKagan, K. K. Perkins, and C. E. Wieman, 2006 *Physics Education Research Conference Proceedings*, edited by L. McCullough, P. Heron, and L. Hsu (AIP Press, Melville, NY, 2007).
- [48] L. D. Carr and S. B. McKagan, Graduate quantum mechanics reform, *Am. J. Phys.* **77**, 308 (2009).
- [49] S. Goldhaber, S. Pollock, M. Dubson, P. Beale, and K. Perkins, 2009 *Physics Education Research Conference Proceedings*, edited by C. Henderson, M. Sabella, and C. Singh (AIP Press, Melville, NY, 2009).
- [50] Louis Deslauriers and Carl Wieman, "Learning and retention of quantum concepts with different teaching methods," *Phys. Rev. ST Phys. Educ. Res.* (to be published).
- [51] R. Doran, *Basic Measurement and Evaluation of Science Instruction* (NSTA, Washington, D.C., 1980).
- [52] P. Kline, *A Handbook of Test Construction: Introduction to psychometric design* (Methuen, London, 1986).
- [53] G. Kuder and M. Richardson, The theory of the estimation of test reliability, *Psychometrika* **2**, 151 (1937).
- [54] L. J. Cronbach, Coefficient alpha and the internal structure of tests, *Psychometrika* **16**, 297 (1951).
- [55] G. A. Ferguson, On the theory of test discrimination, *Psychometrika* **14**, 61 (1949).
- [56] A. P. French and Edwin F. Taylor, *An Introduction to Quantum Physics* (Norton, New York, 1978).
- [57] Richard W. Robinett, *Quantum Mechanics; Classical Results, Modern Systems, and Visualized Examples* (Oxford University Press, New York, 1997).