Validation and administration of a conceptual survey on the formalism and postulates of quantum mechanics

Emily Marshman[1] and Chandralekha Singh[2]

[1]*Department of Physics, Community College of Allegheny County, Pittsburgh, 15212 Pennsylvania, USA*
[2]*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, 15260 Pennsylvania, USA*

We developed and validated a conceptual survey that focuses on the formalism and postulates of quantum mechanics covered in upper-level undergraduate quantum mechanics courses. The concepts included in the quantum mechanics formalism and postulate survey (QMFPS) focus on Dirac notation, the Hilbert space, state vectors, physical observables and their corresponding Hermitian operators, compatible and incompatible observables, quantum measurement, time dependence of quantum states and expectation values, and spin angular momenta. Here we describe the validation and administration of the survey, which has been administered to over 400 upper-level undergraduate and graduate students from six institutions. The QMFPS is valid and reliable for use as a low-stakes test to measure the effectiveness of instruction in an undergraduate quantum mechanics course that covers relevant content. The survey can also be used by instructors to identify students' understanding of the formalism and postulates of quantum mechanics at the beginning and end of a graduate quantum mechanics course since graduate students are expected to have taken an undergraduate quantum mechanics course that covers the content included in the survey. We found that undergraduate students who engaged with research-validated learning tools performed better than students who did not on the QMFPS after the first semester of a junior or senior level quantum mechanics course. In addition, the performance of graduate students on QMFPS after instruction in the first semester of a core graduate-level quantum mechanics course was significantly better than the performance of undergraduate students at the end of the first semester of an undergraduate quantum mechanics course. However, both undergraduate and graduate students struggled with many questions on the QMFPS. A comparison with the base line data on the validated QMFPS presented here can aid instructors in assessing the effectiveness of their instructional approaches and help them identify the difficulties their students have with quantum formalism and postulates in order to help students develop a solid grasp of the formalism and postulates of quantum mechanics.

I. INTRODUCTION

Learning quantum mechanics (QM) is challenging partly because it is abstract as well as nonintuitive and students often transfer ideas from classical mechanics to quantum mechanics inappropriately [1–4]. Several studies have focused on student difficulties with concepts [5–15] and formalism [16–24] in QM. We must help students develop a coherent knowledge structure of the foundational concepts related to the formalism and postulates of quantum mechanics before they can solve novel, complex problems. Furthermore, developing a robust understanding of quantum mechanics requires a solid grasp of linear algebra,

differential equations, and special functions [25]. Regardless of the mathematical complexity of quantum mechanics problems, students must develop a functional understanding of quantum mechanics. This entails developing a good knowledge structure of the underlying concepts and being able to reason systematically about relevant quantum mechanics concepts involved while also developing quantitative skills to solve the problems instead of using a plug-and-chug approach.

Research-based conceptual surveys (both free response and multiple-choice format) are useful tools for evaluating student understanding of various topics without focusing heavily on their mathematical skills. Furthermore, carefully developed and validated surveys can play an important role in measuring the effectiveness of a curriculum and instruction. If well-designed multiple-choice pretests and post-tests are administered before and after instruction in relevant concepts, they can provide one objective means to measure the effectiveness of a curriculum and instructional

approach in a particular course. When compared to free response, multiple choice is free of grader bias and such tests can be graded with great efficiency. Furthermore, the results are objective and amenable to statistical analysis so that different instructional methods or different student populations can be compared. Also, good instructional design requires taking into account the prior knowledge of the students. An effective way to assess the prior knowledge of students, i.e., what the students know before instruction in a particular course, is to administer conceptual surveys as pretests. When pretests are compared with post-tests, the comparison can give us one objective measure of the effectiveness of instruction.

Several multiple-choice conceptual surveys have been developed for use in physics and astronomy courses [26]. For example, in introductory physics, researchers have developed many multiple-choice surveys to determine the knowledge states of students at the beginning and end of instruction in a particular course and/or topic, e.g., the Force Concept Inventory, Conceptual Survey of Electricity and Magnetism, Rotational and Rolling Motion Survey, Energy and Momentum Survey, etc. [27–30]. In addition, conceptual surveys have been developed for use in QM (quantum mechanics) courses [31–33]. For example, the quantum mechanics conceptual survey (QMCS) [31] was developed for sophomore-level modern physics courses. It focuses on wave functions and probability, wave particle duality, the Schrodinger equation, quantization of states, uncertainty principle, superposition, operators and observables, and tunneling. It contains 12 questions. The quantum mechanics concept assessment (QMCA) [32] is a 31-item survey that focuses on the time-independent Schrodinger equation, time evolution, wave functions and boundary conditions, and probability and it can be used in an upper-division junior or senior level QM course. The quantum mechanics visualization instrument (QMVI) was developed to evaluate students' conceptual understanding of core topics in quantum mechanics in the undergraduate curriculum, especially their visualization skills [33]. Furthermore, the quantum mechanics assessment tool (QMAT) gauges student learning in a first semester junior-level quantum mechanics course and focuses on wave functions, measurement, time dependence, probability, infinite square well, one-dimensional tunneling, and energy levels [34]. The introductory quantum physics conceptual survey (IQPCS) focuses on basic quantum concepts related to quantization and uncertainty [35]. The quantum mechanics survey (QMS) [36] covers topics in nonrelativistic quantum mechanics in one spatial dimension typically covered in the first semester of an upper-level undergraduate course and involves concepts such as possible wave functions, bound or scattering states, measurement, expectation values, time dependence of wave function and expectation values, stationary and nonstationary states, role of the Hamiltonian, uncertainty principle, and Ehrenfest's theorem. It can be used in most junior and senior-level quantum mechanics courses if relevant concepts are covered.

However, previously developed conceptual surveys for use in QM courses do not focus explicitly on the postulates or formalism of quantum mechanics. For example, Dirac notation, the Hilbert space, state vectors, physical observables and their corresponding Hermitian operators, compatible and incompatible observables, projection operators and writing operators in terms of their eigenstates and eigenvalues are not covered in other QM conceptual surveys. In addition, other previously developed QM surveys do not include concepts related to spin angular momenta. Therefore, we developed and validated a QM conceptual survey that focuses on the formalism and postulates of quantum mechanics that includes these concepts. Here, we discuss the development and validation of the quantum mechanics formalism and postulates survey (QMFPS), which is a 34-item multiple-choice test appropriate for use in an upper-level undergraduate quantum mechanics course as a post-test (after instruction in relevant concepts) or graduate level quantum mechanics course as a pretest (at the beginning of the course) or post-test [19]. The survey can be used to identify upper-level undergraduate students' final and graduate students' initial and final knowledge states related to the formalism and postulates of quantum mechanics at the beginning and end of a course to assess the effectiveness of a quantum mechanics curriculum in which relevant concepts are covered. The results of the survey can also be used to guide the development of instructional strategies to help students learn these concepts better.

II. QMFPS SURVEY DEVELOPMENT AND ADMINISTRATION

According to the standards for multiple-choice test design, a high-quality test has five characteristics: reliability, validity, discrimination, good comparative data, and is tailored to the population [37–39]. Furthermore, the development of a well-designed multiple-choice test is an iterative process that involves recognizing the need for the test, formulating the test objectives, constructing test items, performing content validity and reliability check, and distribution [37–39]. Below, we describe the development of the QMFPS and how we ensured that the test was developed based upon the standards of multiple-choice test design.

A. Development of the survey

We recognized the need for a conceptual survey focused on the formalism and postulates of quantum mechanics in that previously developed QM conceptual surveys do not focus explicitly on the postulates or formalism of quantum mechanics. In particular, there are no QM surveys that focus explicitly on Dirac notation, the Hilbert space, state vectors, physical observables and their corresponding

Hermitian operators, compatible and incompatible observables, projection operators and writing operators in terms of their eigenstates and eigenvalues. Furthermore, other QM surveys do not include concepts related to spin angular momenta. Therefore, we developed the QMFPS, which focuses on assessing students' conceptual understanding of the formalism and postulates of QM, including Dirac notation, the Hilbert space, state vectors, physical observables and their corresponding Hermitian operators, compatible and incompatible observables, quantum measurement, time dependence of quantum states and expectation values, and spin angular momenta. The final version of the survey is included on PhysPort [40]. Table I shows one possible categorization of the questions on the survey based upon the concepts, although the categorization may be done in many other ways.

While designing the survey, we focused on making sure that it is valid and reliable [37–39]. Validity refers to the extent to which the test consistently measures whatever it is supposed to measure, and reliability refers to the extent to which the test measures what it measures consistently [37–39]. To ensure that the survey is valid for low-stakes group assessment of QM curriculum and instructional approaches that focus on relevant topics, we consulted with 6 faculty members regarding the goals of their QM courses and topics their students should have learned related to the formalism and postulates of quantum mechanics in upper-level undergraduate QM. In addition to carefully looking through the coverage of these topics in several upper-level undergraduate quantum mechanics classes, we also browsed over several homework, quiz, and exam problems that faculty in these courses at the University of Pittsburgh (Pitt) had given to their students in the past when we started the development of the survey. We also gave open-ended questions on relevant topics to students in upper-level QM and interviewed some students one on one to get an in-depth understanding of their reasoning behind their responses. These interactions with faculty members and students helped us formulate the test

objectives and construct the preliminary test items in initial versions of the survey. We note that the faculty members were not only consulted initially before the development of the survey questions, but we also iterated different versions of the survey with several instructors at Pitt at various stages of the development to ensure that the test content was valid, i.e., that the test items matched the objectives of the test and the test items were accurate, formatted correctly, and were grammatically correct. The faculty members reviewed different versions of the survey several times to examine its appropriateness and relevance for the upper-level quantum mechanics courses and to detect any possible ambiguity in item wording. These valuable comments and feedback from faculty members also helped to ensure that the test was designed with the target population (upper-level undergraduate and graduate students) in mind; i.e., the difficulty level of the questions was appropriate for this target population. In addition to the analysis of the responses to the open-ended questions and interviews, the alternative choices for the multiple-choice questions were informed by prior research on common student difficulties in QM on these topics [1–4,14–16,19–24].

The individual interviews were conducted with 23 students using a think-aloud protocol [41] at various phases of the test development to better understand students' reasoning processes while they answered the open-ended and multiple-choice questions. Within this interview protocol, students were asked to talk aloud while they answered the questions so that the interviewer could understand their thought processes. The interviews were invaluable and often revealed unnoticed difficulties (not necessarily clear from written responses), and these were incorporated into new versions of the survey. This allowed us to refine the survey further to ensure that the questions were relevant and clearly worded. The interviews also allowed us to further confirm that the difficulty level of the test was appropriate for upper-level undergraduate and graduate students (i.e., the test was designed for the target population in mind).

TABLE I. One possible categorization of the survey questions, the number of questions that fall in each category, and the question numbers belonging to each category. The number of questions in different categories do not add up to 34 because some questions fall into more than one category.

| Topic	Item number
Quantum states	1,4,7,11,12,13,14,15,17,19
Eigenstates of operators corresponding to physical observables	1,4,7,14,15,17,18, 20
Time development of quantum states	3,4,5,6,7,26,32,34
Measurement	2,3,4,5,7,8,9,13,19,21,23,24,25,27,28,31,32,33,34
Expectation value of observables	5,10,22
Time dependence of expectation value of observables	15,16,17,29,30
Commutators or compatibility	16,17,19,20,27,28
Spin angular momentum	20,21,22,23,24,25,26,27,28,29,30
Dirac notation	4,9,10,11,12,13,14,18
Dimensionality of the Hilbert space	1
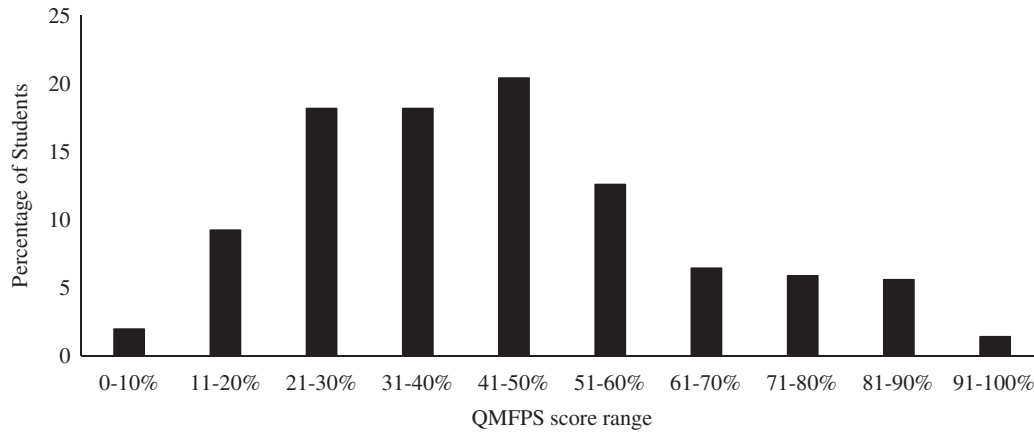
FIG. 1.   Histogram of student scores on the QMFPS.

The final version of the survey can be accessed via the link in Ref. [40]. Each question has one correct choice and four incorrect choices. We find that almost all of the students are able to complete the QMFPS in one 50 min class period after instruction in relevant concepts. Students can answer the QMFPS questions without performing complex calculations, although they do need to understand the basics of linear algebra since that is central to the formalism and postulates of QM. The survey can be used in a junior or senior level undergraduate quantum mechanics course (e.g., at the level of the first four chapters in D. Griffiths' QM textbook [42]) as a post-test, as long as students have learned Dirac notation. It can also be administered in a graduate-level QM course as both a pretest or post-test to determine students' initial and final knowledge states in regards to the formalism and postulates of QM. While the QMFPS should not be administered as a high-stakes test and the data should be interpreted for the class as a whole to gauge the effectiveness of instruction, it is suggested that students receive some credit for completing the survey in order for students to take it seriously. For example, if the survey is given as a post-test, it can count as a graded low-stakes quiz. If the survey is given as a pretest in a graduate course, it can count as a quiz for which students should be given full credit for trying their best.

B. Administration of the validated survey

The reliability check is performed during a large-scale administration of the final form of the test [37–39]. The validated QMFPS was administered to 464 students from 6 institutions over a period of four yr.[1] Of the 464 students, 350 were undergraduate students and 114 were graduate students. The undergraduate students were enrolled in the first semester of a QM course at the junior or senior level.

The graduate students were enrolled in a graduate-level core QM course. The undergraduate students completed the survey as a post-test at the end of their first semester in QM, and the graduate level students completed the survey after at least two months into the first semester of graduate level quantum mechanics. Both the undergraduate and graduate students worked through the survey during a 50 min class period. Some of the undergraduate students were enrolled in QM courses that used research-based learning tools such as concept tests and quantum interactive learning tutorials ($N = 43$). The survey was given to a subset of these students twice, once at the end of the first semester and then again at the beginning of the second semester after the winter break ($N = 15$). This large-scale administration allowed us to collect comparative data by administrating the test to various groups of students for whom it was designed.

C. General test statistics

The average score on the survey after instruction is 41% (including only the first score of the students who took the survey twice). The standard deviation is 20%, with the highest score being 100%. The average score of undergraduate students is 37% and the average score of graduate students is 52%. The fact that the graduate students' performance is better than undergraduate students' performance provides another measure of content validity since graduate students are supposed to know these concepts better overall. There is a significant difference between the graduate and undergraduate students' average scores (p value of t test < 0.001). Figure 1 shows a histogram of the students' scores on the QMFPS.

The average post-test score for the upper-level students who used concept tests and group discussion and Quantum Interactive Learning Tutorials (QuILTs) was 58% (S.D. $= 20$%). The average post-test score for other undergraduate students who did not use research-based learning tools was 32% (S.D. $= 16$%).

[1]Some instructors did not administer the last four questions of the QMFPS. There were 358 students who answered the first thirty questions on the QMFPS.

There is a significant difference between the scores of students who used research-based learning tools and those who did not (p value of t test < 0.0001).

D. Reliability analysis

We performed various statistical analyses to determine if the QMFPS is reliable. If a test is administered twice at different times to the same sample of students, then one would expect a highly significant correlation between the two test scores (test-retest reliability), assuming that the students' performance is stable and that the test environmental conditions are the same on each occasion [37–39]. Since testing students twice in a very short interval is not practical, one way to determine overall test reliability is via the Kuder-Richardson reliability index (KR-20), which is a measure of the self-consistency of the entire test. According to the standards of test design [37–39], the KR-20 should be higher than 0.7 to ensure that the test is reliable. The KR-20 for the QMFPS is 0.87, indicating that the survey is very reliable.

Performing item analysis can provide further insights into the survey's reliability. Table II shows that, on the QMFPS, the item difficulty (percentage correct) for undergraduate students ranges from 0.13 to 0.70 and the item difficulty for graduate students ranges from 0.16 to 0.90. Table III shows the distribution of student responses for each survey item. We calculated item discrimination for each item on the survey to ensure that the test is reliable. One way to measure item discrimination is by calculating the point-biserial coefficient. It is a measure of consistency of a single test item with the whole test—it reflects the correlation between a student's score on an individual item and their score on the entire test [37–39]. The point-biserial coefficient has a possible range of $-1 + 1$ to. If an item has a high point-biserial coefficient, then students with high total scores are more likely to answer the item correctly than students with low total scores. A negative point-biserial value indicates that students with low overall scores were more likely to get a particular item correct than those with a high overall score and is an indication that the particular test item is probably defective. Ideally, point-biserial coefficients should be above 0.2 [37–39]. Table II shows the point-biserial coefficients for each item on the QMFPS. The average point-biserial is 0.41 and ranges from 0.20 to 0.62. The standards of test design [34,35] indicate that the survey questions have reasonably good item discrimination. The question with the lowest point-biserial coefficient of 0.20 (Q 32) was also the most difficult question on the test for all students (item difficulty is 0.16).

Another aspect of survey reliability is construct-related validity, which is associated with understanding the nature of the characteristics being measured and the consequences of the uses and interpretations of the results [37–39]. A construct is an individual characteristic that is used to explain the performance on an assessment. For example,

TABLE II. Item difficulty (percentage of students answering the question correctly) for undergraduate students (UG), graduate students (G), and all students combined (ALL) and item discrimination (point-biserial coefficient) for each item on the QMFPS.

Item number	Item difficulty			Item discrimination (point-biserial coefficient)
	UG	G	ALL	
1	30%	53%	37%	0.49
2	65%	59%	63%	0.38
3	19%	43%	26%	0.59
4	62%	90%	70%	0.48
5	40%	47%	42%	0.31
6	37%	52%	42%	0.37
7	35%	47%	38%	0.33
8	22%	26%	23%	0.35
9	32%	50%	37%	0.53
10	55%	82%	63%	0.53
11	44%	77%	53%	0.49
12	51%	80%	59%	0.50
13	40%	80%	51%	0.55
14	37%	71%	47%	0.42
15	23%	59%	39%	0.42
16	32%	36%	33%	0.23
17	34%	61%	42%	0.42
18	37%	72%	48%	0.44
19	32%	49%	37%	0.26
20	29%	55%	37%	0.43
21	56%	74%	61%	0.48
22	23%	32%	26%	0.34
23	70%	82%	73%	0.38
24	42%	56%	46%	0.37
25	33%	38%	34%	0.42
26	47%	74%	55%	0.50
27	52%	53%	53%	0.27
28	41%	48%	43%	0.34
29	18%	24%	19%	0.39
30	18%	35%	22%	0.39
31	19%	45%	27%	0.54
32	16%	16%	16%	0.20
33	13%	49%	25%	0.62
34	19%	18%	19%	0.31

mathematical reasoning is a construct that can be used to explain students' performance on a mathematics assessment [37–39]. In our survey, understanding of the formalism and postulates of QM is a construct that can be used to explain performance on the QMFPS. One way to collect evidence of construct validity involves related measures studies. Related measures studies investigate correlations between different assessment measures. For example, one would expect a positive correlation between the Force Concept Inventory and the Force and Motion Conceptual Evaluation since they were designed to measure similar constructs (i.e., students' understanding of force and motion) [37–39]. Therefore, we examined

TABLE III.   Distribution of undergraduate and graduate student responses to individual questions on the QMFPS. Bolded percentages indicate the correct answer.

Question	Undergraduate students					Graduate students				
	A	B	C	D	E	A	B	C	D	E
1	**30%**	8%	15%	8%	39%	**53%**	2%	10%	3%	32%
2	**65%**	2%	13%	15%	5%	**59%**	1%	13%	22%	5%
3	9%	25%	2%	44%	**19%**	4%	11%	3%	39%	**43%**
4	**62%**	9%	13%	6%	9%	**90%**	3%	3%	3%	1%
5	7%	9%	9%	**40%**	33%	2%	2%	8%	**47%**	42%
6	4%	2%	29%	**37%**	27%	6%	4%	5%	**52%**	33%
7	22%	9%	5%	**35%**	29%	20%	5%	4%	**47%**	25%
8	36%	21%	**22%**	9%	12%	38%	11%	**26%**	9%	16%
9	11%	27%	25%	4%	**32%**	2%	36%	8%	4%	**50%**
10	9%	9%	4%	**55%**	24%	4%	1%	2%	**82%**	10%
11	37%	1%	**44%**	7%	11%	18%	0%	**77%**	3%	2%
12	6%	26%	**51%**	12%	5%	2%	15%	**80%**	1%	2%
13	7%	10%	25%	17%	**40%**	0%	3%	8%	9%	**80%**
14	**37%**	12%	28%	13%	9%	**71%**	1%	20%	6%	2%
15	14%	**30%**	19%	13%	23%	7%	**59%**	8%	13%	12%
16	10%	19%	17%	21%	**32%**	13%	6%	20%	26%	**36%**
17	**34%**	12%	6%	22%	25%	**61%**	6%	3%	14%	16%
18	16%	**37%**	25%	5%	15%	3%	**72%**	16%	1%	8%
19	9%	23%	28%	7%	**32%**	11%	8%	27%	6%	**49%**
20	24%	**29%**	11%	9%	25%	24%	**55%**	2%	2%	16%
21	2%	22%	7%	**56%**	11%	0%	4%	1%	**74%**	20%
22	11%	18%	21%	25%	**23%**	14%	18%	10%	25%	**32%**
23	18%	**70%**	5%	3%	3%	13%	**82%**	3%	0%	2%
24	**42%**	12%	12%	25%	5%	**56%**	10%	13%	19%	2%
25	21%	11%	17%	**33%**	13%	22%	11%	17%	**38%**	12%
26	12%	27%	7%	**47%**	3%	5%	13%	1%	**74%**	6%
27	28%	5%	**52%**	6%	5%	37%	1%	**53%**	0%	5%
28	12%	**41%**	16%	19%	6%	4%	**48%**	11%	26%	9%
29	11%	4%	**18%**	46%	15%	10%	4%	**24%**	35%	26%
30	**18%**	4%	9%	51%	12%	**35%**	2%	6%	45%	10%
31	13%	10%	**19%**	25%	27%	9%	14%	**45%**	13%	18%
32	37%	15%	**16%**	11%	15%	55%	13%	**16%**	7%	5%
33	15%	11%	**13%**	26%	28%	9%	10%	**49%**	11%	16%
34	38%	13%	**19%**	9%	15%	53%	14%	**18%**	5%	5%
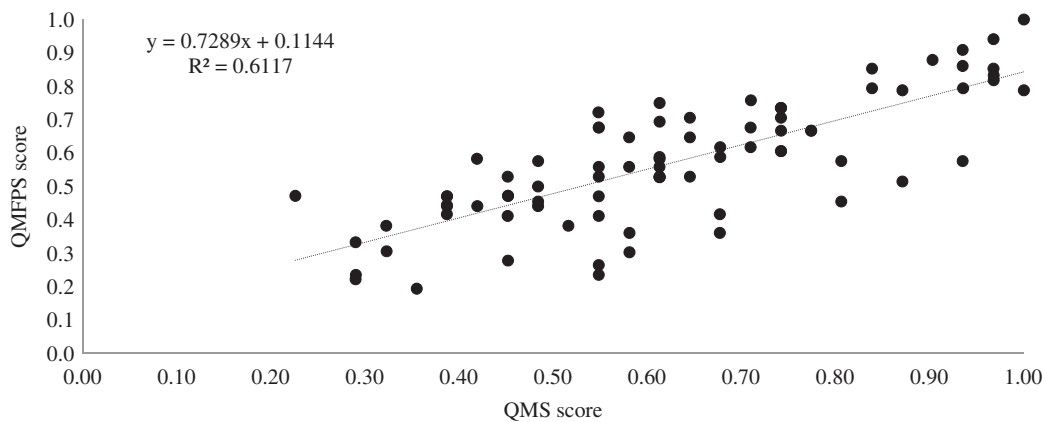
FIG. 2.   Correlation between 82 students' QMFPS scores and their QMS scores. The coefficient of determination is R^2 and the correlation coefficient is $R = 0.78$.
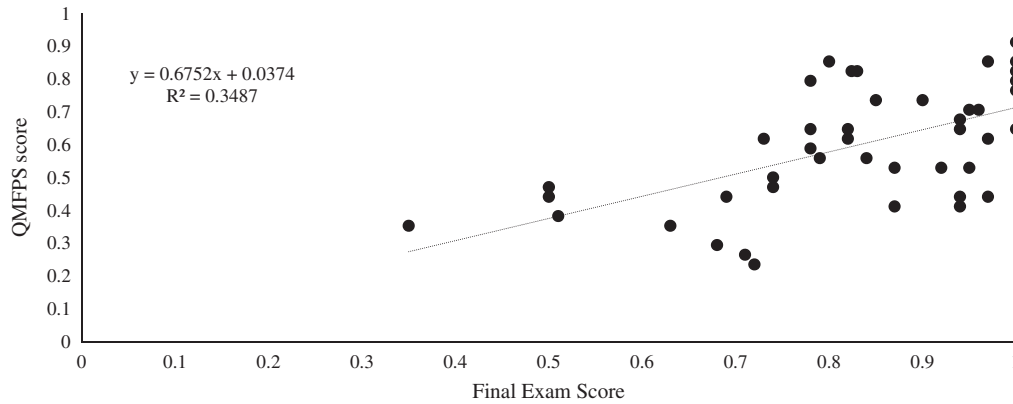
FIG. 3. Correlation between 44 graduate students' QMFPS scores and their final exam score in a core graduate level quantum mechanics course. The coefficient of determination is R^2 and the correlation coefficient is $R = 0.58$.

whether students' QMFPS scores were correlated with other validated QM surveys and their performance in quantum mechanics courses to ensure construct validity of the QMFPS. Eighty-two undergraduate students enrolled in the first semester of an undergraduate upper-division QM course were given both the QMFPS and the quantum mechanics survey (QMS), which is a previously validated survey that focuses on students' understanding of non-relativistic QM in one dimension, after traditional instruction in relevant concepts. Figure 2 shows that there is a strong correlation between students' scores on the QMS and the QMFPS. This correlation provides construct validity to the QMFPS survey because students who performed well on the QMS are generally likely to have a better foundation in formalism and postulates of quantum mechanics and perform better on the QMFPS. The QMFPS tends to be more difficult for students than the QMS, possibly because it covers more advanced topics as opposed to the QMS which covers QM in one spatial dimension.

In addition, 44 graduate students enrolled in the first semester of a graduate-level core quantum mechanics course were given the QMFPS at the end of the semester. Figure 3 shows that there is a moderate correlation between students' scores on the QMFPS and their final exam in the graduate-level QM course. This correlation provides further evidence of construct validity of the QMFPS since the concepts covered in the final exam were similar to those covered in the QMFPS.

III. SUMMARY

Learning QM is challenging for students partly because of the "paradigm shift" from classical mechanics to quantum mechanics as well as the mathematical expertise required to solve problems. Students in traditionally taught and evaluated QM courses may be able to "hide" their lack of conceptual understanding of the formalism and postulates of QM behind their mathematical skills [3]. However, in order for students to develop functional understanding,

it is important to close the gap [43] between conceptual and quantitative problem solving by assessing both types of learning. We have developed a conceptual survey that assesses students' conceptual knowledge of the formalism and postulates of QM, which are topics that instructors of QM courses agree are important to cover [25]. The development of the test followed the standards of multiple-choice test design [37–39], and we ensured that the test was valid and reliable, had good discrimination, was tailored to the population, and that we collected good comparative data [37–39].

Details of student difficulties found via QMFPS is beyond the scope of this paper and will be discussed elsewhere. Student responses to questions on the QMFPS can be used as a formative assessment [44,45] to help instructors identify common student difficulties and guide the design of instructional strategies and learning tools to improve students' understanding. This survey can also be administered to students after instruction in the relevant concepts to evaluate the effectiveness of instruction on relevant topics in a particular course.

Furthermore, we found that students who were enrolled in QM courses that used active learning methods, such as peer instruction and tutorials, performed better on the QMFPS than those who did not. These approaches included active learning techniques [46–81] such as peer instruction [52,53], tutorials [54–68], cooperative group problem solving [69], and Just-In-Time-Teaching [70,71], to help students develop a coherent knowledge structure of the formalism and postulates of QM. In addition, we found that, although graduate students performed significantly better than undergraduate students, their average overall score was not very high. This may partly be due to the fact that graduate students who were taught primarily via traditional approaches may have developed algorithmic skills to solve problems on their exams, which often reward "plug and chug" approaches, but lack a conceptual understanding of quantum mechanics. However, even graduate students may not be motivated to develop a coherent

knowledge structure of QM if course assessments only focus on quantitative reasoning. Therefore, in order to help students develop a functional knowledge of quantum mechanics, we suggest that the learning goals for upper-level QM include proficiency in concepts covered in the QMFPS and emphasize the connection between conceptual understanding and mathematical formalism. Furthermore, instructors of graduate-level QM courses can reflect on their students' responses on the QMFPS to design instruction that helps to "close the gap" between students' conceptual understanding and quantitative problem solving.

ACKNOWLEDGMENTS

[1] C. Singh, Student understanding of quantum mechanics, Am. J. Phys. **69**, 885 (2001).

[2] C. Singh, Student understanding of quantum mechanics at the beginning of graduate instruction, Am. J. Phys. **76**, 277 (2008).

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

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

[5] M. Wittmann, R. Steinberg, and E. Redish, Investigating student understanding of quantum physics: Spontaneous models of conductivity, Am. J. Phys. **70**, 218 (2002).

[6] M. Wittmann, J. Morgan, and L. Bao, Addressing student models of energy loss in quantum tunneling, Eur. J. Phys. **26**, 939 (2005).

[7] P. Jolly, D. Zollman, S. Rebello, and A. Dimitrova, Visualizing potential energy diagrams, Am. J. Phys. **66**, 57 (1998).

[8] P. C. Garcia Quijas and L. M. Arevala Aguilar, Overcoming misconceptions in quantum mechanics with the time evolution operator, Eur. J. Phys. **28**, 147 (2007).

[9] S. Sharma and P. K. Ahluwalia, Diagnosing alternative conceptions of Fermi energy among undergraduate students, Eur. J. Phys. **33**, 883 (2012).

[10] 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).

[11] C. Singh, Transfer of learning in quantum mechanics, AIP Conf. Proc. **790**, 23 (2005).

[12] C. Singh and G. Zhu, Cognitive issues in learning advanced physics: An example from quantum mechanics, AIP Conf. Proc. **1179**, 63 (2009).

[13] C. Singh, Assessing and improving student understanding of quantum mechanics, AIP Conf. Proc. **818**, 69 (2006).

[14] G. Zhu and C. Singh, Improving students' understanding of quantum measurement: I. Investigation of difficulties, Phys. Rev. ST Phys. Educ. Res. **8**, 010117 (2012).

[15] S. Lin and C. Singh, Categorization of quantum mechanics problems by professors and students, Eur. J. Phys. **31**, 57 (2010).

[16] C. Singh and E. Marshman, Analogous patterns of student reasoning difficulties in introductory physics and upper-level quantum mechanics, in *Proceedings of the 2013 Physics Education Research Conference, Portland, OR*, edited by P. V. Engelhardt, A. D. Churukian, and D. L. Jones (2014), https://doi.org/10.1119/perc.2013.inv.010.

[17] E. Marshman, C. Keebaugh, and C. Singh, Student difficulties with the corrections to the energy spectrum of the hydrogen atom for the intermediate field Zeeman effect, in *Proceedings of the 2018 Physics Education Research Conference, Washington, DC*, edited by A. Traxler, Y. Cao, and S. Wolf (2018), https://doi.org/10.1119/perc.2018.pr.Marshman.

[18] C. Singh, E. Marshman, and C. Keebaugh, Student difficulties with finding the fine structure corrections to the energy spectrum of the hydrogen atom using degenerate perturbation theory, *Proc. Phys. Educ. Res. Conf.* (2018), https://doi.org/10.1119/perc.2017.pr.088.

[19] E. Marshman, Ph.D. Thesis, University of Pittsburgh, 2015 accessed at http://d-scholarship.pitt.edu/25547/.

[20] E. Marshman and C. Singh, Student difficulties with quantum states while translating state vectors in Dirac notation to wave functions in position and momentum representations, in *Proc. Phys. Educ. Res. Conf.* (2015), https://doi.org/10.1119/perc.2015.pr.048.

[21] C. Singh and E. Marshman, Student difficulties with determining expectation values in quantum mechanics, in *Proceedings of the 2016 Physics Education Research Conference, Sacramento, CA*, edited by D. L. Jones, L. Ding, and A. Traxler (2016), https://doi.org/10.1119/perc.2016.pr.075.

[22] E. Marshman and C. Singh, Student difficulties with representations of quantum operators corresponding to observables, in *Proceedings of the 2016 Physics Education Research Conference, Sacramento, CA*, edited by D. L. Jones, L. Ding, and A. Traxler (2016), https://doi.org/10.1119/perc.2016.pr.049.

[23] C. Singh and E. Marshman, Investigating student difficulties with Dirac Notation, in *Proceedings of Physics Education Research Conference* (2014), https://doi.org/10.1119/perc.2013.pr.074.

[24] E. Marshman and C. Singh, Investigating student difficulties with time-dependence of expectation values in quantum mechanics, in *Proc. Phys. Educ. Res. Conf.* (2014), https://doi.org/10.1119/perc.2013.pr.049.

[25] S. Siddiqui and C. Singh, How diverse are physics instructors' attitudes and approaches to teaching undergraduate-level quantum mechanics?, Eur. J. Phys. **38,** 035703 (2017).

[26] A. Madsen, S. McKagan, and E. Sayre, Resource Letter RBAI-1: Research-based assessment instruments in physics and astronomy, Am. J. Phys. **85,** 245 2017.

[27] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, Phys. Teach. **30,** 141 (1992).

[28] D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke, and A. V. Heuvelen, Surveying students' conceptual knowledge of electricity and magnetism, Am. J. Phys. **69,** S12 (2001).

[29] L. Rimoldini and C. Singh, Student understanding of rotational and rolling motion concepts, Phys. Rev. ST Phys. Educ. Res. **1,** 010102 (2005).

[30] C. Singh and D. Rosengrant, Multiple-choice test of energy and momentum concepts, Am. J. Phys. **71,** 607 (2003).

[31] S. B. McKagan, K. K. Perkins, and C. E. Wieman, Design and validation of the quantum mechanics conceptual survey, Phys. Rev. ST Phys. Educ. Res. **6,** 020121 (2010).

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

[33] R. Robinett and E. Cataloglu, Testing the development of student conceptual and visualization understanding in quantum mechanics through the undergraduate career, Am. J. Phys. **70,** 238 (2002).

[34] S. Goldhaber, S. Pollock, M. Dubson, P. Beale, and K. Perkins, Transforming Upper Division Quantum Mechanics: Learning Goals and Assessment 2009 Physics Education Research Conference, AIP Conf. Proc. **1179,** 145 (2009).

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

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

[37] G. Aubrecht and J. Aubrecht, Constructing objective tests, Am. J. Phys. **51,** 613 (1983); A. J. Nitko, *Educational Assessments of Students* (Prentice-Hall/Merrill, Englewood Cliffs, NJ, 1996).

[38] P. V. Engelhardt, An Introduction to classical test theory as applied to conceptual multiple-choice tests, in *Getting Started in PER*, edited by C. Henderson and K. Harper (AAPT, College Park, MD, 2009).

[39] W. Adams and C. Wieman, Development and validation of instruments to measure learning of expert-like thinking, Int. J. Sci. Educ. **33,** 1289 (2011).

[40] https://www.physport.org/assessments/assessment.cfm?A=QMFPS.

[41] M. Chi, Thinking aloud, in *The think aloud method: A practical guide to modeling cognitive processes*, edited by M. W. Van Someren, Y. F. Barnard, and J. A. C. Sandberg (Academic Press, London, 1994).

[42] D. J. Griffiths, *Introduction to Quantum Mechanics* (Prentice Hall, Upper Saddle River, NJ, 1995).

[43] C. Singh, Closing the gap between teaching and assessment, written for the report of the Association of American Universities (AAU) 2014, http://rescorp.org/gdresources/publications/effectivebook.pdf.

[44] National Research Council 2001, Knowing what students know: The science, and design of educational assessment, *Committee on the Foundations of Assessment*, edited by J. Pellegrino, N. Chudwosky, and R. Glaser (National Academy Press, Washington, DC, 2001).

[45] P. Black and D. Wiliam, Assessment and classroom learning, Assess. Educ. Princ. Pol. Pract. **5,** 7 (1998); P. Black and D. William, Inside the black box: Raising standards through classroom assessment, Phi Delta Kappan **92,** 81 (2010); P. Black and D. William, Developing the theory of formative assessment, Educ. Asse. Eval. Acc. **21,** 5 (2009); P. Black, C. Harrison, C. Lee, B. Marshall, and D. William, *Assessment for Learning: Putting It Into Practice* (Open University Press, Buckingham, 2003).

[46] D. Zollman, N. Rebello, and K. Hogg, Quantum mechanics for everyone: Hands-on activities integrated with technology, Am. J. Phys. **70,** 252 (2002).

[47] A. Kohnle, I. Bozhinova, D. Browne, M. Everitt, A. Fomins, P. Kok, G. Kulaitis, M. Prokopas, D. Raine, and E. Swinbank, A new introductory quantum mechanics curriculum, Eur. J. Phys. **35,** 015001 (2014).

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

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

[50] C. Singh, M. Belloni, and W. Christian, Improving student's understanding of quantum mechanics, Phys. Today **59,** 43 (2006).

[51] B. Brown, C. Singh, and A. J. Mason, The effect of giving explicit incentives to correct mistakes on subsequent problem solving in quantum mechanics, in *Proc. Physics Education Research Conf.* (2015), https://doi.org/10.1119/perc.2015.pr.012.

[52] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).

[53] C. Singh and G. Zhu, Improving students' understanding of quantum mechanics by using peer instruction tools, AIP Conf. Proc. **1413,** 77 (2012).

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

[55] G. Zhu and C. Singh, Improving students' understanding of quantum mechanics via the Stern Gerlach experiment, Am. J. Phys. **79,** 499 (2011).

[56] C. Singh, Helping students learn quantum mechanics for quantum computing, AIP Conf. Proc. **883,** 42 (2007).

[57] G. Zhu and C. Singh, Improving students' understanding of quantum measurement: II. Development of research-based

learning tools, Phys. Rev. ST Phys. Educ. Res. **8,** 010118 (2012).

[58] G. Zhu and C. Singh, Improving student understanding of addition of angular momentum in quantum mechanics, Phys. Rev. ST Phys. Educ. Res. **9,** 010101 (2013).

[59] S. DeVore and C. Singh, Development of an interactive tutorial on quantum key distribution, in *Proc. Physics Education Research Conf.* (2015), https://doi.org/10.1119/perc.2014.pr.011.

[60] B. Brown and C. Singh, Development and evaluation of a quantum interactive learning tutorial on Larmor Precession of spin, in Proc. Physics Education Research Conf. (2015), https://doi.org/10.1119/perc.2014.pr.008.

[61] E. Marshman and C. Singh, Interactive tutorial to improve student understanding of single photon experiments involving a Mach-Zehnder Interferometer, Eur. J. Phys. **37,** 024001 (2016).

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

[63] E. Marshman and C. Singh, Investigating and improving student understanding of quantum mechanical observables and their corresponding operators in Dirac notation, Eur. J. Phys. **39,** 015707 (2018).

[64] E. Marshman and C. Singh, Investigating and improving student understanding of the expectation values of observables in quantum mechanics, Eur. J. Phys. **38,** 045701 (2017).

[65] E. Marshman and C. Singh, Investigating and improving student understanding of the probability distributions for measuring physical observables in quantum mechanics, Eur. J. Phys. **38,** 025705 (2017).

[66] A. Maries, R. Sayer, and C. Singh, Effectiveness of interactive tutorials in promoting "which-path" information reasoning in advanced quantum mechanics, Phys. Rev. Phys. Educ. Res. **13,** 020115 (2017).

[67] R. Sayer, A. Maries, and C. Singh, A quantum interactive learning tutorial on the double-slit experiment to improve student understanding of quantum mechanics, Phys. Rev. Phys. Educ. Res. **13,** 010123 (2017).

[68] C. Keebaugh, E. Marshman, and C. Singh, Developing and evaluating an interactive tutorial on degenerate perturbation theory, in *Proc. Phys. Educ. Res. Conf.* (2017), https://doi.org/10.1119/perc.2016.pr.041.

[69] P. Heller, R. Keith, and S. Anderson, Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving, Am. J. Phys. **60,** 627 (1992).

[70] G. Novak, E. Patterson, A. Gavrin, W. Christian, and K. Forinash, Just in time teaching, Am. J. Phys. **67,** 937 (1999).

[71] R. Sayer, E. Marshman, and C. Singh, A case study evaluating Just-in-Time Teaching and Peer Instruction using clickers in a quantum mechanics course, Phys. Rev. Phys. Educ. Res. **12,** 020133 (2016).

[72] C. Singh, Student difficulties with quantum mechanics formalism, AIP Conf. Proc. **883,** 185 (2007).

[73] C. Singh and G. Zhu, Improving students' understanding of quantum mechanics by using peer instruction tools, AIP Conf. Proc. **1413,** 77 (2012).

[74] A. J. Mason and C. Singh, Do advanced students learn from their mistakes without explicit intervention?, Am. J. Phys. **78,** 760 (2010).

[75] B. Brown, A. Mason, and C. Singh, Improving performance in quantum mechanics with explicit incentives to correct mistakes, Phys. Rev. Phys. Educ. Res. **12,** 010121 (2016).

[76] C. Keebaugh, E. Marshman, and C. Singh, Improving student understanding of corrections to the energy spectrum of the hydrogen atom for the Zeeman effect, Phys. Rev. Phys. Educ. Res. **15,** 010113 (2019).

[77] C. Keebaugh, E. Marshman, and C. Singh, Investigating and addressing student difficulties with the corrections to the energies of the hydrogen atom for the strong and weak field Zeeman effect, Eur. J. Phys. **39,** 045701 (2018).

[78] C. Keebaugh, E. Marshman, and C. Singh, Investigating and addressing student difficulties with a good basis for finding perturbative corrections in the context of degenerate perturbation theory, Eur. J. Phys. **39,** 055701 (2018).

[79] C. Keebaugh, E. Marshman, and C. Singh, Student difficulties with the number of distinct many-particle states for a system of non-interacting identical particles with a fixed number of available single-particle states, in *2018 Proc. Phys. Educ. Res. Conf.* http://doi.org/10.1119/perc.2018.pr.Keebaugh.

[80] C. Keebaugh, E. Marshman, and C. Singh, Improving student understanding of fine structure corrections to the energy spectrum of the hydrogen atom, Am. J. Phys. **87,** 594 (2019).

[81] C. Keebaugh, E. Marshman, and C. Singh, Improving student understanding of a system of identical particles with a fixed total energy, Am. J. Phys. **87,** 583 (2019).