Designing, validating, and contrasting conceptual quantum mechanics questions for spin states and spatial wave functions

Steven J. Pollock,[1] Homeyra Sadaghiani,[2] Adam Quaal,[3] and Gina Passante[3]

[1]*Department of Physics, University of Colorado, 390 UCB, Boulder, CO, 80309*
[2]*Department of Physics and Astronomy, California State Polytechnic University, 3801 W. Temple Ave, Pomona CA, 91768*
[3]*Department of Physics, California State University Fullerton, 800 N. State College Blvd., Fullerton, CA, 92831*

Student understanding of quantum mechanics is a topic of increasing interest to physics education researchers. One goal is to investigate differences in performance in various popular instructional paradigms. Towards this end, we have modified an existing, validated, conceptual multiple-choice assessment to include questions explicitly framed in two "contexts": spin states (typically using Dirac notation) and position (spatial wave functions represented in position space.) We discuss the development of new questions, and present preliminary results on student performance on questions which contrast related conceptual ideas. Our data come from "spins-first" instructional contexts. This effort serves as an initial validation for the new questions, provides insights into differing responses, and forms a basis for future work to allow instructors to tune assessment questions to best match and inform their own instructional environment choices.

I. INTRODUCTION

Upper-division quantum mechanics is a challenging and essential milestone in undergraduate physics education. There is a growing literature on students' conceptual difficulties [1–5], epistemology [6, 7], and much more [8–10]. A common curricular approach in many quantum courses is labeled Position First (PF) [11], starting with the Schrödinger equation in a continuous position basis. An increasingly popular alternative approach is Spin First (SF), which typically begins by considering sequential Stern-Gerlach Experiments in a discrete spin-1/2 basis to lay out the postulates of quantum mechanics in a less mathematically challenging context [11]. Pedagogical (and philosophical) arguments can be made for the relative advantages of these two approaches, and a variety of tools and instruments exist to assess students conceptual learning in such courses [12].

We had previously developed a conceptual, multiple choice instrument (QMCA [13]) to compare learning outcomes across classes. That instrument built on an earlier open-ended assessment (QMAT [14]), using largely PF notation and contexts. After teaching multiple semesters of SF courses, we began to wonder what impact the context and framing of questions might have on student performance. We label the contexts S- for questions about spin states (generally in Dirac notation), and P- for questions involving position-space wave functions. Here, we present preliminary efforts to develop and validate a new set of parallel questions, modifying existing QMCA items to address differences and commonalities in student performance and difficulties across contexts. One long term goal is to ultimately provide an assessment instrument suitable for use in either instructional context.

II. GOALS AND NEW QUESTIONS

Each QMCA question was categorized [13] into broad themes including "wave function properties", "time evolution", and "measurement". The latter include a variety of questions about different aspects of quantum measurements, and are particularly ripe for further investigation. The variation in performance on *measurement* questions across different courses and curricula was the largest of any category, and many questions readily lent themselves to alternative contexts and representations. We identified 9 such questions (of 31 on the QMCA) to target, all about particles in an infinite well. We formulated variants which ask analogous questions in the context of spin states. These questions target a variety of conceptual issues - interpretation of probability amplitudes, state after measurement, sequential measurements, impact of time evolution on measurements, and distinguishing physical measurement from the mathematical action of operators.

We discuss below our process of question development and summarize some initial outcomes. We were struck by the consistency across populations and time, the level of internal consistency for students' answers and reasoning elements, and (with some intriguing exceptions we discuss below) relatively small contextual differences. This work can aid development of future assessments to allow instructors to tailor the QMCA to better match their notation and style, without heavily compromising conclusions about student conceptual understanding. One important caveat is that the data we have collected come only from the authors' institutions, and thus are all from courses using SF curricula. Follow-up work collecting data from PF curricula will better inform future versions of the QMCA

III. METHODS AND QUESTION DEVELOPMENT

We followed an iterative process of modifying analogous questions based on exam results and post-test interviews. Several older P-questions were reworded slightly so that the S- and P-context versions could be phrased more similarly to each other, and one new pair of questions was introduced. Several sample questions are shown in Fig. 1, the rest of the questions can be found at [15].

A spin-1 particle is in a uniform magnetic field oriented in the +z direction. The Hamiltonian is proportional to S_z. The normalized eigenstates of S_z for this particle are labeled $|-1\rangle$, $|0\rangle$ and $|1\rangle$, with corresponding eigenvalues $-\hbar, 0,$ and $+\hbar$. **Consider a particle starting in a quantum state $|\psi\rangle = |1\rangle$.**

3-S. You make a measurement of S_x. You then immediately measure S_z. At this point, what value(s) could you get for the z-component of spin of the particle?
A. Any continuous angular momentum value between $-\hbar$ and $+\hbar$
B. 0 **C.** $+\hbar$ **D.** $-\hbar$ or $+\hbar$ **E.** $-\hbar, 0,$ or $+\hbar$

4-S. Instead of an immediate re-measurement of S_z, consider waiting between the S_x and S_z measurements. Would the probability of measuring S_z to be $+\hbar$ depend on how long you wait between the measurements?
A. Yes **B.** No

5-S. Choose the best explanation for your answer to the previous question.
A. Over time the particle will relax to its initial state.
B. Because it is precessing in a B-field, the probabilities of S_z change with time.
C. S_z commutes with the Hamiltonian, so its probabilities are time-independent.
D. All possible measurement probabilities in quantum mechanics are time independent because you take the absolute square of the coeficient.
E. Once you measure S_x the state is no longer in an eigenstate of the Hamiltonian.

FIG. 1. Sample S-context questions. Matching P-variants involve sequential position and energy measurements for a particle in a box. (Numbering and small formatting details changed from original)

A. Data collection

In the fall of 2016 (Fa16), we gave a first version of the new test, including most standard QMCA questions plus new questions. Students at three institutions participated in class at the end of their first upper-division Quantum course, given roughly 40 minutes, framed as an ungraded practice test. This resulted in 65 (33 at CUB, 14 at CPP, and 18 at CSUF) tests, representing roughly 75% of the enrolled students. Our classes are 10-20% female. CPP and CSUF are both Hispanic-serving primarily undergraduate institutions. CUB is a larger (R1, PhD granting) public University. We taught lecture-style classes using concept-tests and peer-instruction methods, with student-centered group activities. All classes followed McIntyre's (SF) text. CPP was on a quarter system, however the content of the course spanned the topics on the QMCA. The other two courses were semester-long and spanned significantly more topics than the QMCA covers.

The following year we modified questions based on informal interviews, comparing results to our earlier QMCA dataset, and considering faculty consensus learning goals. We revised the new S-context questions and some of their older P-counterparts, and converted the sole open-ended (S-context) "reasoning element" question (Q5-S above) to multiple choice format. The extra pair of questions added in the first version scored so well in all three classes that they were dropped. We show one sample page with S-context questions in Fig 1. The corresponding P-type questions change the context to a particle in an infinite square well, with sequential position and then energy measurements [1].

Below we present results from subsequent administrations to 135 students whose demographics and institutions were otherwise similar to the Fa16 data. These data were taken Fa17 (CUB, $N = 20$, CSUF, $N = 5$, CPP, $N = 38$) and a modified version in Sp18 (CUB, $N = 56$, and CSUF, $N = 16$, a second semester course) In section IV A below, we exclude students taking earlier or partial versions of new questions, leaving us with $N = 110$ for whole test comparison and analysis.

B. Reasoning elements and open-ended responses

For the first round of data collection, the S-context reasoning question (Q5-S) was in open-ended format to evaluate whether reasoning elements would match across contexts. In the original P-context, where sequential measurements are position and then energy for a particle in a well, the correct answer for the preceding question on time dependence (Q4-P: "...would the probability of measuring a particular energy depend on how long you wait", which it does not) is $\sim 60\%$. One distractor for the reasoning question, "over time, the wave function will relax to the initial state" in PF-classes historically attracts $\sim 15\%$ of students. In our sample (only SF-classes) this response now attracted $< 5\%$.

Q4-S scored almost identically to the P-version, $\sim 60\%$. However, students' open-ended responses showed very different reasoning elements. For instance, the S-context now elicited reasoning about precession. The later multiple-choice version (Q5-S) showed strong popularity for those new elements, although almost no students (<1%) selected "relaxation". Precession (which doesn't impact measurement probability along the B-field direction) attracted over 1/4 of students. An additional 20% chose "once you measure S_x the state is no longer in an eigenstate", a true statement but not an explanation for the lack of time-dependence, and almost five times more prevalent than the equivalent response in the P-context. In summary, although the overall correct rate on both versions of Q4 and Q5 were similar, reasoning elements were notably different. We believe qualitative research is needed to better understand and characterize how and why students activate different reasoning elements in different contexts.

C. Question design and false negatives

The first questions on the QMCA involve a superposition state $\Psi(x, t = 0) = \sqrt{4/5}\psi_1(x) + \sqrt{1/5}\psi_2(x)$, with a normalized energy eigenstate of a particle in a well. Students are asked for the most probable value for an energy measurement outcome (here, E_1), and then for the normalized state corresponding to a different measurement yielding the maximum possible energy value (here, $\psi_2(x)$). This latter question has historically generated moderately low scores [13]. Based on student feedback, we wondered whether there might be false negatives arising from students misinterpreting "maximum possible" with "most probable" value, and/or whether the sequence was generating some confusion regarding the starting state for the follow-up question.

We introduced a new sequence with an intermediate question asking separately for the "maximum possible" value, and also rewrote the starting (superposition) state in all questions. In both versions (S- and P-) students scored high ($\sim 90\%$) on both "most probable" and "max value" questions. As a result, we eliminated the new (intermediate) question. Consistent high performance on the final question (now $> 90\%$ across our sample) suggests that our earlier, lower results may have partly arisen from ambiguous question wording rather than inherent student confusion about the topic. Of course, the new results were in highly student-centered SF-courses, which may also play a role.

IV. RESULTS

A. Classical test validation of new questions

The final set of new S-context questions were constructed to match paired P-questions from an expert's perspective. We evaluated the new questions using standard statistical metrics. Because of question modifications, we restrict data to 2017-18 implementations, $N = 110$ students at 3 institutions. Table I summarizes some key statistics, comparing to previously published QMCA results [13]. The version analyzed here eliminates 2 older questions and adds the 9 new spin versions (S-context) discussed above.

TABLE I. A summary of classical test statistics, comparing previously published results with our new version.

Statistics	QMCA [13]	New version	S-context only
# students	263	110	(same, 110)
# questions	31	38	9
Standard deviation	16%	15%	19%
Item Difficulty	0.5	0.6	0.7
Point Biserial Coeff.	0.4	0.3	0.4
Kuder-Richardson	0.8	0.8	0.5
Ferguson's δ	0.97	0.97	0.90

Ferguson's δ measures how broadly total scores are distributed. Values above 0.9 are considered strong discrimination. Item-difficulty is the average score on each question. For the new S-questions, this ranges from 0.4 to 0.9, similar to the range for P-questions (0.2 - 0.9) and to the previous QMCA as a whole (0.2 - 0.9) The average new S-item difficulty (0.7) is similar to the average of the analogous P-questions (0.63), slightly above the published QMCA average (0.54) which represents a much broader range of questions, institutions and curricula. One P-question (Q3-P) was slightly modified to better match its S-counterpart and now stands out as distinctly low scoring (22% correct) and is discussed further below.

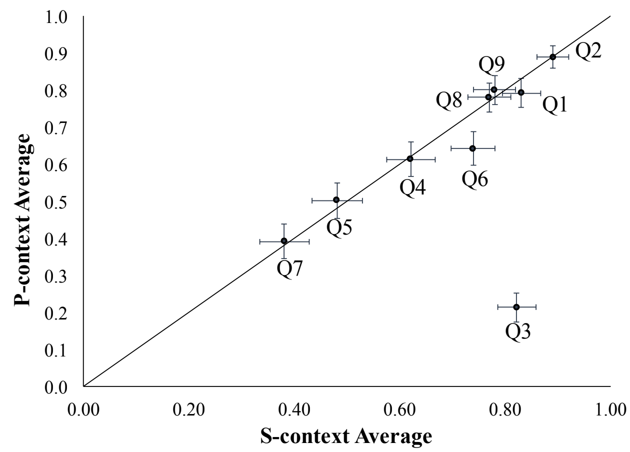

FIG. 2. Student scores on paired questions asked in both Spin (S) contexts or position (P) context. Horizontal axis shows the score on the S-version, vertical is the P-version score. Error bars are standard error of mean. The linear trend line is included for reference - points falling below this line show higher performance on the S-version.

Point-biserial coefficients measure individual item discrimination, correlating items with the entire test. These coefficients are similar for new and old questions. S-context questions' coefficients ranged from 0.1 to 0.5, with an average of 0.4. Kuder-Richardson (KR-20, equivalent to Chronbach's α) is considered a measure of internal consistency. The new test as a whole gives a value of 0.79, an acceptable indication of reliability. (This measure typically decreases for smaller numbers of tested items.) These broad test statistics track those from our previously published QMCA work. In summary, our results suggest reasonable coherence and no statistically problematic new spin-context questions, with performance comparable to previous (published) questions.

B. Comparison across contexts

Although the statistics above suggest average performance of new S-context questions is statistically similar to older questions, our driving motivation was to investigate possible differences of S- and P-context performance. Here, we look more closely at individual questions. Fig. 2 shows a comparison of 9 questions (combining recent data sets from 2 years across 3 institutions, $N = 110$ students). Each point allows a direct comparison of matched "S-context" (horizontal axis) and "P-context" (vertical axis) scores.

Although earlier implementations showed more variation than Fig 2, only a few questions show indication of being easier in one context. One strong example is Q3-P mentioned above. It is perhaps not too surprising that S-context scores are higher, since this topic is the first covered in SF curricula and heavily emphasized - measuring a component of spin "erases" knowledge of spin in orthogonal directions. But, the incompatibility of position and energy measurements was much less salient for students. We discuss this further below.

C. Triangulation with open-ended exam responses

One P-context question (Q3-P, on sequential measurements of incompatible operators) scored significantly lower than its spin counterpart (Q3-S in Fig. 1). In 2018 at CUB, we followed up with a related pair of final-exam questions. For this population, only 23% (13/56) got Q3-P right. This result is lower than historical averages, and significantly below the S-version (Q3-S) which scored 86% (48/56). In one final exam question, students were explicitly asked whether position and energy measurements commute for a particle in an infinite well: 80% (45/ 56) correctly stated that it does not. Our intent was to prime students, to see whether explicitly drawing attention to commutation assists with the next question, which replicates Q3-P: given an accurate position measurement ($\Delta x \ll$ well-width), if you then re-measure energy, what might you get (and on the final exam, why)?

The exam score for the re-measurement question (following the "priming" lead-in) was moderate, 66%, giving partial credit for consistency. Of the students who got this question right on the QMCA, 100% (13/13) had the earlier commutation exam question right, and 85% (11/13) got the subsequent measurement question correct with a solid written explanation. Of those who got QMCA Q3-P wrong, 74% (32/43) still correctly stated that $[x, H] \neq 0$, but only 37% (16/43) got the energy remeasurement question correct with correct explanation. And of those who got the commutation exam question wrong (11/43), only one got the energy re-measurement question correct. These results suggest that question Q3-P may indeed reflect common student difficulties regarding the impact of non-commuting observables on sequential measurements outside of a spins context. This may be attributable to aspects of the SF-curriculum, since the historical QMCA average for this question in (largely) PF-classes is considerably higher [13], and warrants future study.

V. CONCLUSIONS

We have created 9 new variant QMCA questions shifting context from position to spins. They yield overall statistically similar outcomes by conventional measures, suggesting that a "spins variant" of the QMCA might be feasible (e.g., for faculty teaching a spins-first curriculum). The questions pass preliminary classical test requirements, and may help us highlight and quantify prevalence of student difficulties. We largely do not see significant differences in scores in SF-curricula across question contexts, suggesting that whatever advantages SF-curricula might provide, it does not unambiguously show up on this particular instrument. We do see intriguing results in the connection between commutation properties to sequential measurement outcomes, particularly in the position context. We also note student challenges regarding time evolution and the interpretation of measurements, both of which show small but potentially interesting differences across question contexts. We believe such indicators deserve further study at a more qualitative level, to dig deeper into not merely "how many get it wrong" in different curricula, but what and how students are thinking about these questions, and what productive resources they bring to bear.

We are collectively interested in the affordances and challenges of a SF curriculum, and hope that such research can help guide ongoing curricular reforms in which either (or both) spin and position contexts might work better to deepen students understanding and facility with quantum mechanics concepts. Our ongoing research includes further investigations into the differences in student conceptual understanding in the two instructional paradigms and how targeted instruction can bridge the gaps.

ACKNOWLEDGMENTS

We thank all our respective PER groups, students, and colleagues for their invaluable help and support, particularly Benjamin Schermerhorn. This work was funded by the NSF DUE-1626280, DUE-1626594, and DUE-1626482. .

[1] C. Singh, and E. Marshman, Phys. Rev. ST - PER **11**, 020117 (2015)
[2] E. Marshman, and C. Singh, Phys. Rev. ST - PER **11**, 020119 (2015)
[3] G. Zhu and C. Singh, Phys. Rev. ST - PER, **8**, 10117 (2012)
[4] G. Passante et al., Phys. Rev. ST - PER **11**, 020135 (2015), and 020111 (2015).
[5] P. Emigh et al., Phys. Rev. ST - PER **11**, 020112 (2015)
[6] J. Hoehn, et al., N.,Phys. Rev. ST - PER**14**, 010122 (2018)
[7] B. Modir et al., Phys. Rev. ST - PER **13**, 020108 (2017)
[8] E. Gire and E. Price, Phys. Rev. ST - PER **11**, 20109 (2015)
[9] C. Singh, Am. J. Phys., **76**, 227 (2008)
[10] M. Dubson et al., PERC Proceedings 2009, AIP Press (2009) doi: 10.1063/1.3266697
[11] D. McIntyre et al., *Quantum Mechanics*(Pearson, 2012), D Griffiths, *Introduction to Quantum Mechanics* (Pearson, 2005)
[12] www.physport.org
[13] H. Sadaghiani and S. Pollock, Phys. Rev. ST - PER **11**,010110 (2015)
[14] S. Goldhaber et al., PERC Proceedings 2009, doi:10.1119perc.2016.pr.068
[15] Email steven.pollock@colorado.edu for a pdf of the QMCA.