

Case study evaluating Just-In-Time Teaching and Peer Instruction using clickers in a quantum mechanics course

Ryan Sayer,¹ Emily Marshman,² and Chandralekha Singh²

¹*Department of Physics, Bemidji State University, Bemidji, Minnesota 56601, USA*

²*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*
(Received 18 March 2016; published 18 October 2016)

Just-in-Time Teaching (JiTT) is an instructional strategy involving feedback from students on prelecture activities in order to design in-class activities to build on the continuing feedback from students. We investigate the effectiveness of a JiTT approach, which included in-class concept tests using clickers in an upper-division quantum mechanics course. We analyze student performance on prelecture reading quizzes, in-class clicker questions answered individually, and clicker questions answered after group discussion, and compare those performances with open-ended retention quizzes administered after all instructional activities on the same concepts. In general, compared to the reading quizzes, student performance improved when individual clicker questions were posed after lectures that focused on student difficulties found via electronic feedback. The performance on the clicker questions after group discussion following individual clicker question responses also showed improvement. We discuss some possible reasons for the improved performance at various stages, e.g., from prelecture reading quizzes to postlecture clicker questions, and from individual to group clicker questions.

DOI: [10.1103/PhysRevPhysEducRes.12.020133](https://doi.org/10.1103/PhysRevPhysEducRes.12.020133)

I. INTRODUCTION

Just-in-Time Teaching (JiTT) is an instructional strategy in which instructors receive feedback from students and use that feedback to tailor instruction [1]. Typically, students complete an electronic prelecture assignment in which they give feedback to the instructor regarding any difficulties they have had with the assigned reading material, lecture videos, and/or other self-paced instructional tools. The instructor then reviews student feedback before class and makes adjustments to the in-class activities. For example, during class, the instructor can focus on student difficulties found via electronic feedback. Students may engage in discussions with the instructor and with their classmates, and the instructor may then adjust the next prelecture assignment based on the progress made during class. When JiTT was first conceived in the late 1990s [1], the required internet technology for electronic feedback was still evolving; developments in digital technology since then have continued to make electronic feedback from students and the JiTT approach easier to implement in classes.

It has been hypothesized that JiTT may help students learn better because out-of-class activities cause students to engage with and reflect on the parts of the instructional material they find challenging [1]. For example, when the instructor focuses on student difficulties in lecture which

were found via electronic feedback before class, it may create a “time for telling” [2] particularly because students may be “primed to learn” better when they come to class if they have struggled with the material during prelecture activities. Although prior studies have shown that the JiTT strategy may be effective for helping introductory students develop expertise in introductory physics [1,3], the use of JiTT with students in upper-division courses has received less attention.

The JiTT approach is often used in combination with peer discussion in the classroom [1]. Peer collaboration has been used in many instructional settings in physics classes, and with various types and levels of student populations [[4–9]. Although the details of the implementation vary, students can learn from each other in many different environments. Integration of peer interaction with lectures has been popularized in the physics community by Mazur [4]. In Mazur’s approach, the instructor poses concrete conceptual problems in the form of conceptual multiple-choice clicker questions to students throughout the lecture and students discuss their responses with their peers. In addition to Mazur’s approach, Heller *et al.* have shown that collaborative problem solving with peers in the context of quantitative “context-rich” problems is valuable both for learning physics and for developing effective problem solving strategies [5].

One framework for explaining why the JiTT approach and peer discussion are effective learning strategies is the cognitive apprenticeship model. According to the cognitive apprenticeship model, students can learn effectively if the instructional design involves three essential components:

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

“modeling,” “coaching and scaffolding,” and “weaning” [10]. In this approach, modeling means that the instructor demonstrates and exemplifies the skills that students should learn (e.g., how to solve physics problems systematically). Coaching and scaffolding means that students receive appropriate guidance and support as they actively engage in learning the skills necessary for good performance. Weaning means gradually reducing the support and feedback to help students develop self-reliance.

In traditional physics instruction, especially at the college level, there is often a lack of coaching and scaffolding [11,12]. The situation is often akin to a piano instructor demonstrating for the students how to play the piano and then asking students to go home and practice. The lack of prompt feedback and scaffolding can be detrimental to learning. JiTT gives instructors the opportunity to receive student feedback on their difficulties and adjust their in-class activities accordingly, providing students with the necessary coaching and scaffolding to help them learn. Peer discussion also provides students an opportunity for being coached by peers who may even be able to discern their difficulties better than the instructor, and carefully designed targeted feedback from the instructor after the peer discussion can provide appropriate scaffolding.

It has been proposed that peer discussion may positively affect students’ self-efficacy, which is defined as students’ belief in their ability to succeed in accomplishing a given goal or task [8]. Likewise, students’ self-efficacy may also play a role in how students participate in peer discussion and how much they benefit from it. Miller *et al.* have shown that low self-reported self-efficacy may play an even greater role than their course performance up to that point in predicting how likely students are to switch their response to a clicker question from right to wrong after discussion with their peers [9]. It will be useful to investigate similar issues in upper-level courses using similar surveys.

Here, we discuss the findings of an investigation in a quantum mechanics course which employed a JiTT strategy including peer instruction with clickers as part of the in-class instruction. Learning quantum mechanics is challenging even for advanced students, partly because the subject matter is nonintuitive and abstract. Some investigations have focused on the difficulties upper-level students have with quantum physics [13–18] and how to help them learn quantum mechanics better [19–22]. In this case study, we compare students’ performance on prelecture reading quizzes, in-class conceptual clicker questions (concept tests) answered individually after lecture focusing on student difficulties, clicker questions answered after peer discussion, and open-ended retention quizzes given during a later class session after all relevant instruction on the particular topic. We then discuss some possible interpretations and implications of the findings to aid future research involving pedagogical interventions of similar type.

II. MOTIVATION, RESEARCH QUESTIONS, AND APPROACH TO ADDRESS THEM

Prior research on student learning in upper-division quantum mechanics courses suggests that students in these courses share some of the same characteristics as students in an introductory course in classical mechanics [23]. The diversity in student preparation and goals for majoring in physics has increased significantly, and advanced students in physics courses vary in their prior knowledge, skills, motivation, and self-efficacy in a manner similar to students in introductory physics courses [23–25]. Many students in advanced physics courses often struggle to develop a basic grasp of concepts, and they are not necessarily self-regulated learners [26,27], as some instructors might expect. They need the help of research-based teaching and learning strategies in order to repair, organize, and extend their knowledge structures and develop useful problem solving and reasoning skills. Moreover, the paradigm of quantum mechanics is significantly different from the classical paradigm that advanced students are familiar with and which is more intuitive. This paradigm shift introduces an additional obstacle in learning quantum mechanics unlike learning in the other advanced physics courses [23].

With this in mind, it is useful to understand how advanced students in a quantum mechanics (QM) course respond to pedagogical intervention that involves continuing feedback and active learning strategies in the classroom. The JiTT approach and in-class clicker questions involving peer instruction were implemented in an upper-division QM course in order to help students develop a robust knowledge structure of QM concepts while also helping them learn reasoning and metacognitive skills.

The study was designed to investigate the following research questions:

- (1) How do students in an advanced undergraduate QM course perform in “reading” quizzes administered right after a prelecture reading of the topics in the textbook (before in-class activities focusing on the concepts)?
- (2) How effective are lectures focusing on student difficulties in improving students’ performance on questions involving various QM concepts, as measured by their performance on clicker questions given after lecture on those concepts but before discussion with their peers?
- (3) Does peer discussion lead to better performance on the questions involving various QM concepts, as measured by students’ performance on clicker questions after discussion with their peers?
- (4) How do students perform after all relevant instruction on a particular topic, as evidenced by their performance on open-ended retention quizzes on those topics given later in the course?
- (5) Are students’ learning gains significantly larger after any particular learning activity than others?

- (6) What are some of the most challenging concepts for students who had this intervention, and what strategies in the instructional sequence appear to be effective in helping students overcome their difficulties?
- (7) Is there a correlation between advanced students' reported self-efficacy on a self-efficacy survey and their tendency to switch from an initially correct response on an in-class clicker question to an incorrect response on the clicker question after peer discussion?
- (8) Are students equally likely to not respond to in-class clicker questions at the beginning of the semester and later in the semester?

In order to investigate these questions, we compare students' performance on prelecture quizzes administered in multiple-choice format with their performance on identical clicker questions given first after lecture only and then again after peer discussion. We also compare these findings with students' performance on questions in open-ended retention quizzes focusing on similar topics that were given several times throughout the semester after all instruction in relevant concepts. We then focus on students' average performance on individual topics in QM after each learning activity in the instructional sequence in order to identify the concepts that are challenging for students and whether students' learning gains are significantly larger after a particular learning activity in the instructional intervention. We then discuss issues related to the correlation between students' self-efficacy and how students switch their responses between individual and group concept tests. Finally, we discuss some possible interpretations and implications of these findings to help future research to improve student learning with interventions of similar type.

III. METHODOLOGY

A. Instructional design and implementation

A JiTT strategy was implemented in an upper-division (junior-senior level) undergraduate quantum mechanics course taught at a large state-related research university. The course, which consisted of 20 students and met on Mondays, Wednesdays, and Fridays, was an advanced elective course mainly for physics juniors and seniors and focused on topics such as the hydrogen atom, identical particles, quantum statistical mechanics, time-independent and time-dependent perturbation theory, and other approximate methods for solving the time-independent Schrödinger equation (TISE). In addition to the traditional textbook homework problems assigned weekly on the material that was already discussed in the class, students were also assigned weekly prelecture reading from the textbook by Griffiths [28] as homework on the material not yet discussed in the class. In their "reflective homework assignment" on

the prelecture reading, they were asked to first summarize the assigned reading from the textbook in their own words focusing on the concepts and then identify the parts of the material they found challenging. Students electronically submitted to the instructor their written summaries of the pre-lecture reading and their feedback on the material they found challenging on the course website before the class. Participation in reflective homework assignments was generally good (the percentage of students completing the reading assignments each week was always greater than 75%). The reflective homework was graded for completeness, unlike the textbook homework problems from the previous week's material, which were graded for correctness. The instructor read students' reported difficulties and tailored the in-class lecture and concept tests to address the challenges identified by the students.

Each week, the students were administered a multiple-choice *reading quiz* (RQ) on Wednesdays at the beginning of the class soon after they had submitted the prelecture reading assignment but before any in-class lecture on the subject. In the RQ, students were typically given 10 multiple-choice questions to answer in 15 minutes. They were not allowed to consult their textbooks or class notes (or any other resource) while taking the quizzes. The time was sufficient for all students to complete the RQs. The students were not told the correct responses after they were administered the RQs. Student performance on the prelecture RQs was used to answer research question 1.

After lecture, which focused on student difficulties identified in the prelecture reading assignment, students were given a multiple-choice *individual concept test* (ICT) using clickers which repeated verbatim many of the questions from the reading quizzes. Students answered these individually without discussing them with a peer. The ICTs were given on the days when the RQ was not given. Since RQs were typically given on Wednesdays, ICTs were typically given on Mondays and Fridays. Student performance on ICTs compared to RQs was used to answer research question 2.

After answering the ICT, students were encouraged to discuss the questions in groups of two or three for 1–2 minutes and were told to try and convince their peers about why the response they chose was correct. Students were not shown a histogram with the distribution of student responses after the ICT. After peer discussion, each student individually answered the same clicker questions again. We refer to these clicker questions following peer discussion as the *group concept test* (GCT). Students' performance on GCTs was compared with their performance on ICTs to answer research question 3. After each GCT clicker response, there was a general discussion about each question as a whole class.

After the first week of classes, students typically settled down in a fixed seat in the class and they usually discussed the clicker questions with the same one or two peers seated

next to them throughout the semester before the GCT. We therefore divided the 20 students into nine groups based on their usual collaborations in the class during clicker questions, which we refer to as groups A–I. We will use these group identifiers to investigate the effectiveness of peer discussions in different groups.

Students were also given open-ended retention quizzes, referred to as *open quizzes* (OQs), to evaluate their learning after all activities related to a particular concept were completed (e.g., reflective homework, reading quizzes, clicker questions, whole class discussions, traditional textbook homework, and other out-of-class studying). These OQs were given several weeks after the same concepts were covered in prelecture reading, RQs, lectures, ICTs, GCTs, class discussions after GCTs and textbook homework. Students were told about the OQs at least a week ahead of time. A total of five OQs, which typically consisted of 8–10 questions in a free-response format, were given throughout the semester. Students' performance on OQs was analyzed to answer research question 4.

The RQ, ICT, GCT, and OQ questions were developed over a period of more than ten years using an iterative approach of development and evaluation. In particular, the questions were administered to students and faculty members, and went through multiple revisions based on both student and instructor feedback. The OQs together constituted about 4.5% of the students' grade and these OQ questions were graded for correctness. By comparison, the RQs and clicker questions counted as a bonus 5% added to the students' total grade, which comprised a 2.5% bonus for RQs and 2.5% bonus for clicker questions. Moreover, students were given 80% of the possible points on the RQ, ICT, and GCT for participating and 100% for answering the question correctly so there was less explicit incentive to be correct on these assessments compared to OQs.

After the first six weeks of the 14 week long course, the instructor was concerned about the amount of time left to cover all the remaining material. Therefore, from then on, the students were only given the clicker questions as GCT and asked to convince their peers of their reasoning immediately after the question was posed. Also, in the first six weeks, when students performed well on an ICT question as judged by the instructor (which typically meant that they scored above 75%), they were not given the corresponding GCT question. This occurred for seven of the 42 clicker questions given during the first six weeks of the course. These seven questions can be found in Appendix B. The remaining 35 clicker questions were given as both ICTs and GCTs. Eighteen of the clicker questions were most closely matched with free-response questions found in the OQs and were chosen for comparison in this study (since we wanted to evaluate the retention of the concepts learned a few weeks after all learning activities related to a particular concept were over). These questions, which are representative of the various QM

topics covered in the first six weeks of the course with RQs, ICTs, and GCTs, will be referred to as *comparison questions* in this paper (see Appendix A).

B. Data analysis

We took into account the possibility of guessing while grading the multiple-choice questions [29]. Although a one-to-one comparison of the multiple-choice questions with the corresponding open-ended OQ questions is not possible on the same scale, a qualitative comparison between the students' performance on OQ questions and on the multiple-choice clicker questions (RQ, ICT, and GCT) can be made after accounting for guessing. This qualitative comparison of the OQ scores with students' scores on earlier learning activities can provide some insight into robustness of student learning. However, we should keep in mind that students may perform poorly on an open-ended question because they may not have deep understanding to generate a response even though they can recognize the concept in the multiple-choice format. On the other hand, students may perform worse on a multiple-choice question if the alternative choices focus on common student difficulties. Thus, a comparison of RQ, ICT, and GCT with OQ cannot be taken as a one-to-one comparison on the same scale.

Guessing can occur on the multiple-choice clicker questions but is unlikely to occur on the open-ended questions in OQ since students had to generate their responses in the latter situation. Therefore, the multiple-choice questions were scored using a percentage of maximum possible or POMP technique described below in order to account for the possibility of guessing [29]. We used POMP scores to answer research questions 1–3 and for a qualitative comparison with OQ scores in order to answer research questions 4 and 5.

When considering how individual students performed on all of the comparison questions, *Individual POMP Scores* [29] in percent were calculated for each question using the following formula:

$$\text{Individual POMP score in percent} = \frac{(\text{individual\%} - \text{guessing\%})}{(100\% - \text{guessing\%})} \times 100.$$

In this example, the “individual %” is either 100% if the student selected one of the correct options or 0% if the student did not select one of the correct options. The “guessing %” corresponds to the probability that the student would guess one of the correct responses.

As an example, consider the following multiple-choice question:

I. Choose all of the following statements that are correct according to Hund's rules:

(1) The state with the highest total spin (S) will have the lowest energy.

(2) The state with the highest total spin (S) will have the highest energy.

(3) The state with the highest total orbital angular momentum (L), consistent with overall symmetrization, will have the lowest energy.

A. 1 only B. 2 only C. 3 only D. 1 and 3 only E. 2 and 3 only

Statement (1) is correct and is closely matched with an open-ended retention quiz question, so we calculated individual POMP scores for statement (1) based on whether the students selected either option A or option D, indicating that they agreed with statement (1). The guessing % in this case (for correctness of statement (1) only) is $2/5$ or 40% (option A or D out of the five options). Using the example shown above for a particular student, suppose a student chose option E. Since they did not select either option A or D, their individual % will be 0% (without POMP). The student's corresponding individual POMP score will then be $(0\% - 40\%)/(100\% - 40\%) \times 100\% = -66.6\%$. If the student had instead chosen option A or D, their individual POMP score would be 100% (same as their score without POMP). For each student, the individual POMP scores for all 18 comparison questions were averaged together (i.e., the sum of their individual POMP scores for the questions was divided by 18) to determine each student's overall individual POMP score that accounts for guessing [29].

When considering how all students in the class performed on average on a given question, an *average POMP score* in percent was calculated for each question by taking the average of the students' individual POMP scores for that question [29]. An average POMP score near 100% would indicate that most of the students selected the option with the correct statement. An average POMP score around 0% would indicate that, on average, the students were guessing on the question. A negative average POMP score would indicate that, on average, students were deliberately choosing incorrect responses over correct ones, possibly due to alternative conceptions associated with the topic.

In the OQ after all learning activities related to the concepts, questions were graded as either correct or incorrect based upon the students' responses (no partial credit). Agreement of greater than 90% was reached between two raters for all questions. If an open-ended question asked for more than what was asked for in the multiple-choice questions used in RQ, ICT, and GCT, we only graded the correctness of the OQ response for each student based upon the equivalent elements of the corresponding multiple-choice question (see Appendix A). Typically, the average OQ scores and POMP scores will both be high when the students know the correct responses and will both be low when students are guessing on both. However, if students are systematically choosing distractor options they may have a negative average POMP score

but they cannot have a negative average OQ score, so the comparison between the two formats is not on the same scale even with the POMP adjustments. While the quantitative features of our findings depend on whether the scores are unadjusted or adjusted via a POMP technique, the qualitative features are similar for both unadjusted and POMP scores. Here, we report findings involving POMP scores.

As an example, in one of the OQ questions, the students are asked to state Hund's rule used for determining total spin angular momentum quantum number S for the ground state of multielectron atoms. Students who responded that the state with the highest total spin S will have the lowest energy were counted as correct. This OQ question and the corresponding multiple-choice RQ/ICT/GCT questions are collectively referred to as question I in the discussion below. The comparison questions discussed in this research cover the following topics and are given in Appendix A:

- (I) Hund's rule for total spin (S).
- (II) Hund's rule for total orbital angular momentum (L).
- (III) Probability of finding an electron between a distance r and $r + dr$ from the nucleus of a hydrogen atom.
- (IV) Spin configuration of electrons for a helium atom in the ground state.
- (V) Spin configuration of electrons for a helium atom in an excited state.
- (VI) Fermi energy of copper cubes of different sizes at temperature $T = 0$ K.
- (VII) Total energy associated with valence electrons in copper cubes of different sizes at temperature $T = 0$ K.
- (VIII) Change in total energy associated with valence electrons as the volume of a copper cube is changed but the number of atoms is kept fixed.
- (IX) Noninteracting distinguishable particles in a one-dimensional infinite square well.
- (X) Noninteracting bosons in a one-dimensional infinite square well.
- (XI) Three noninteracting fermions in four single particle states.
- (XII) Is the perturbing Hamiltonian matrix \hat{H}' diagonal in the basis in which the unperturbed Hamiltonian matrix \hat{H}^o is diagonal?
- (XIII) Given that the perturbing Hamiltonian \hat{H}' and the unperturbed Hamiltonian \hat{H}^o both commute with some Hermitian operator \hat{A} , do they necessarily commute with each other?
- (XIV) Is an eigenstate $|a\rangle$ of \hat{H}^o corresponding to a degenerate subspace of \hat{H}^o necessarily a "good" state for a given perturbing Hamiltonian \hat{H}' ?
- (XV) Is an eigenstate $|c\rangle$ corresponding to a non-degenerate subspace of \hat{H}^o necessarily a "good" state for a given perturbing Hamiltonian \hat{H}' ?
- (XVI) Can one use the coupled representation $|n, l, s, j, m_j\rangle$ (the notation is standard) when

calculating 1st-order energy corrections to a hydrogen atom energy spectrum due to a perturbing Hamiltonian $\hat{H}' = \alpha \hat{L}_z$?

(XVII) Can one use the coupled representation $|n, l, s, j, m_j\rangle$ (the notation is standard) when calculating 1st-order energy corrections to a hydrogen atom energy spectrum due to a perturbing Hamiltonian $\hat{H}' = \alpha \delta(r)$?

(XVIII) Can one use the coupled representation $|n, l, s, j, m_j\rangle$ (the notation is standard) when calculating 1st-order energy corrections to a hydrogen atom energy spectrum due to a perturbing Hamiltonian $\hat{H}' = \alpha \hat{J}_z$?

In order to investigate research question 6, we compared students' average performance on the RQ, ICT, GCT for each of the 18 comparison question topics using the average POMP score for each question.

In addition, the students in this study were given a self-efficacy (S.E.) survey at the end of the semester which was the survey given by Miller *et al.* [9] adapted for QM. This survey asked students to rate how strongly they agreed or disagreed with 16 statements involving their perceived ability to perform the course activities [9]. For example, one of the questions adapted from Miller *et al.*'s survey states, "I am usually confident that I can convince my neighbor of my answer to a quantum mechanics concept test (clicker question)." Students were then asked to select whether they (5) strongly agree, (4) agree, (3) neither agree nor disagree, (2) disagree, or (1) strongly disagree with each statement. The responses were then scored on a scale of 1 to 5 points, where 5 points were given for a response corresponding to the greatest self-efficacy while 1 point was given for a response corresponding to the least self-efficacy. An average self-efficacy score was then determined for each student by averaging the points assigned to the students' responses on each question. A higher score corresponds to a higher reported self-efficacy [9]. We then determined the frequency with which each individual student switched from a correct response on the ICT to an incorrect response on the GCT after peer discussion using the following equation:

Switching%

$$= \frac{\text{\# of times switched from right ICT to wrong GCT}}{\text{\# of right ICT responses}} \times 100.$$

The students' switching frequencies were then matched with their reported S.E. score in order to investigate research question 7. In addition to switching frequencies, the number of times each student didn't respond to clicker questions when the student was present in class was determined for each week of instruction in order to answer research question 8. The attendance in class was generally very good (typically greater than 80%) throughout the semester.

IV. RESULTS

A. Results by student over the course of the semester

The overall individual POMP scores on the RQ, ICT, and GCT for all 18 comparison questions were averaged over all students. These average scores, as well as the students' average scores on the comparison questions in the OQs, are shown in Table I. Overall, there is an upward trend from RQ to ICT and from ICT to GCT. In Table I, average scores on OQ are indicated with decimals (out of a total score of 1) rather than percentages to highlight the difference in scoring for the open-ended questions. The average performance levels off from GCT to OQ. Median scores for the RQ, ICT, GCT, and OQ are also shown in Table I, and the same trend is observed with the medians as with the averages. In response to research question 1, students on average scored 20% on the RQ administered soon after they completed the prelecture reading assignment.

A comparison of the individual students' average performance on the RQ vs the ICT for the comparison questions is shown in Fig. 1. The symbols labeled A–I are chosen to represent the groups in which the students collaborated after the ICT to answer the GCT, e.g., students denoted by a dark blue circle worked in the same group A after ICT. While students did not work in groups to answer RQs or ICTs, it is useful to represent the members of different groups by different symbols in order to keep track of the student groups for future comparison and discussion. Students on average improved on the ICT (48% average) administered after lecture compared to the RQ immediately after completing the prelecture reading assignment (20% average). Comparison between RQ and ICT using a *t* test showed that the difference between the means was significant ($p = 0.004$). On an individual basis, some students exhibited high gains from the RQ to ICT (e.g., the two students represented by green triangles), while other students on average showed no improvement or even a decline in their scores. There was a noticeable decline in the

TABLE I. Average and median student scores (averaged over all students and all comparison questions) and standard deviations on the reading quiz (RQ), individual concept test (ICT), group concept test (GCT), and open quiz (OQ), with *p* values for comparisons between tests in the same format. The OQ score is in decimal (out of 1) as a reminder that it is in a different format. The *p* values show that the difference in means between RQ and ICT is significant ($p = 0.004$) and the difference in the means between ICT and GCT is significant ($p = 0.009$).

	RQ	ICT	GCT	OQ
Average	20%	48%	73%	0.78
Median	21%	51%	75%	0.74
St. Dev.	23%	33%	21%	0.13
<i>p</i> value		RQ → ICT 0.004	ICT → GCT 0.009	

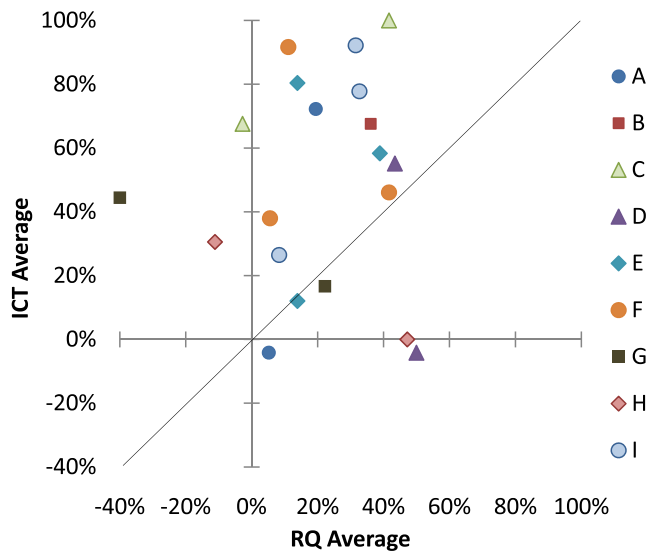


FIG. 1. Student performance on ICT vs RQ, averaged over all comparison questions. The difference between the means of the RQ and ICT scores is significant ($p = 0.004$). Error bars are not shown on the plot for clarity; the average standard error was $\pm 6\%$ for the RQ and $\pm 8\%$ for the ICT.

ICT performance vs RQ performance for two students (represented by the pink diamond and purple triangle near the bottom right corner). One possible reason for this decline may be that these students were mostly guessing on the RQ and got lucky in their responses. Another possibility is that these students did some cramming just before the RQ (on Wednesdays) when they turned in their reflective homework for that week but then forgot many of the concepts they had studied by the time they took the ICT (either Friday or next Monday). In response to our research question 2 (“How effective are the lectures focusing on student difficulties in improving students’ performance on various QM concepts?”), on average, students’ improvement in performance from RQ to ICT was statistically significant with $p = 0.004$.

Figure 2 shows a comparison of average student performances on the GCT vs the ICT. On average, students showed significant improvement from the ICT to GCT clicker questions after discussing the questions with their classmates ($p = 0.009$). In answer to research question 3 (“Does peer discussion lead to better performance on QM concepts as measured by students’ performance on clicker questions after discussion with their peers?”), Fig. 2 shows that a few of the groups were more productive in their collaborations than the others as measured by the group members’ GCT performance compared to their ICT performance. In many of the groups, all group members showed improvement after discussing the questions, as indicated by the symbols located above the diagonal line. However, sometimes the benefits of collaboration as measured by GCT scores appeared to be one-way, with a potentially stronger student helping a weaker student. In

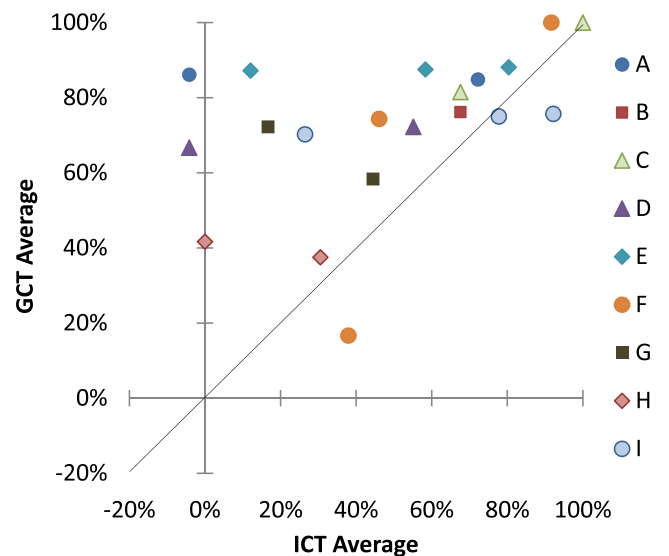


FIG. 2. Student performance on GCT vs ICT, averaged across all comparison questions ($p = 0.009$). Error bars are not shown on the plot for clarity; the average standard error was $\pm 8\%$ for the ICT and $\pm 5\%$ for the GCT.

group A, for example, one of the students performed better on the ICT questions than the other, but both members performed well on the GCT after their discussion. The discussions, in general, appear to have had a positive effect on the student who had a lower performance in the ICT. Additionally, Fig. 2 shows that for some groups, one of the members showed no improvement or even deteriorated after the discussions. In such a case, the group discussions could be considered ineffective for that student based on the comparison of ICT and GCT scores. This situation was observed with group F (represented by orange circles).

A comparison of students’ average GCT and OQ scores for all comparison questions is shown in Fig. 3. This plot suggests that most students performed relatively well on the OQ, regardless of how they performed on the same topics on the GCT. Indeed, the Pearson correlation coefficient $R^2 = 0.045$ between GCT and OQ suggests that students’ performance on the GCT was not correlated with their performance on the OQ. In response to research question 4 (“How do students perform after all relevant instruction, as evidenced by their performance on open-ended quizzes given later in the course?”), the average student performance on the OQ was 0.78 (as noted, OQ score is written as a decimal instead of a percentage to highlight its open-ended format). The reasonably high OQ score indicates that the lectures, class discussions that followed the ICT, and all other learning activities such as homework and self-study that students may have done in the intervening time had a cumulative positive effect on performance on the OQ.

Figure 4 compares students’ individual average performances on the OQ questions with their averages on the RQ. A Pearson correlation coefficient $R^2 = 0.039$ suggests that

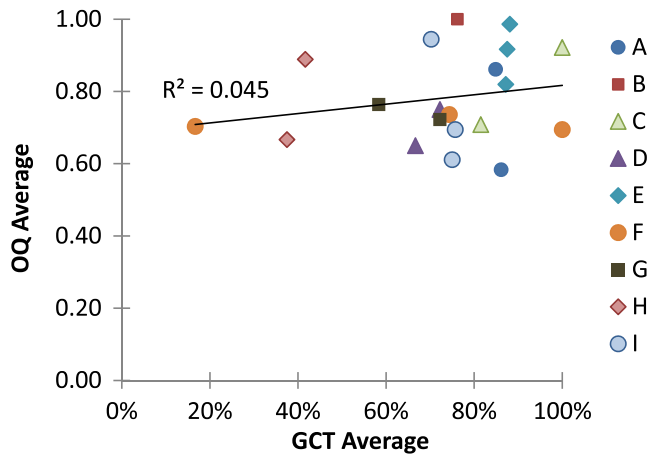


FIG. 3. Student performance on OQ vs GCT, averaged across all comparison questions, with linear regression and corresponding Pearson correlation coefficient ($R^2 = 0.045$). Error bars are not shown on the plot for clarity; the average standard error was $\pm 5\%$ for the GCT and ± 0.03 for the OQ.

students’ performance on the RQ was not correlated with their performance on the OQ. In response to research question 5 (“Are the students’ learning gains significantly larger after any particular learning activity?”), Figs. 1–4 suggest that there was no single learning activity that led to maximum learning gains for all students.

B. Results by topic

We now consider the average performance of all students taken together on individual topics. By considering data by topic, we can identify the concepts that were particularly difficult and investigate research question 6. Figure 5 shows

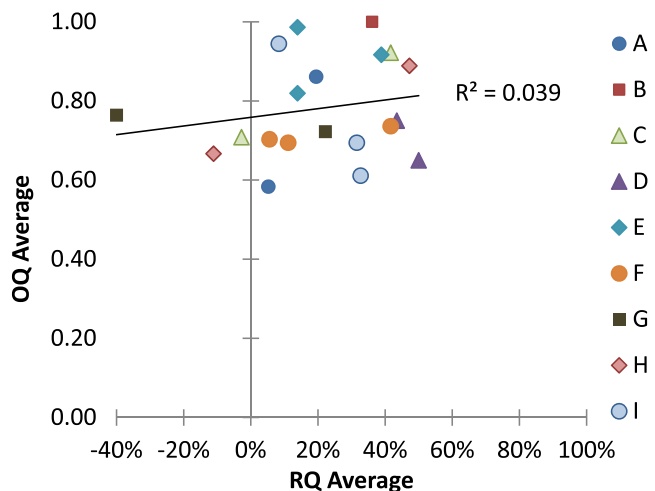


FIG. 4. Student performance on OQ vs RQ, averaged across all comparison questions, with linear regression and corresponding Pearson correlation coefficient ($R^2 = 0.039$). Error bars are not shown on the plot for clarity; the average standard error was $\pm 6\%$ for the RQ and ± 0.03 for the OQ.

the average ICT vs RQ scores for all comparison questions listed in Sec. III B. Each data point represents the average POMP score on a particular question. Figure 5 shows that students performed better on the ICT than on the RQ for most questions, although there were a few questions for which the scores either did not improve or declined from RQ to ICT. Figure 5 also shows that students improved greatly on some questions, e.g., question XV related to degenerate time-independent perturbation theory, which asks students to identify whether an eigenstate of the unperturbed Hamiltonian \hat{H}^o that is not part of a degenerate subspace of \hat{H}^o is a “good” state for finding first-order corrections to energy due to the perturbing Hamiltonian \hat{H}' . The only topic for which students performed worse on average on the ICT vs the RQ is question XIII, which asks students if \hat{H}^o and \hat{H}' must necessarily commute given that they both commute with another Hermitian operator \hat{A} . It appears that the lecture focusing on student difficulties was not very helpful in improving student understanding of this topic. (Note that questions III, IV, and V do not appear in Fig. 5 because there was no RQ for those questions.)

Figure 6 compares the average performances for each comparison question on the GCT vs the ICT. Each data point represents the average POMP score on a particular question. The students on average showed improvement for most of the questions after discussion with their peers. There was one question, however, for which peer discussions did not appear to be helpful: question III, which asks students to determine the probability of finding an electron in a hydrogen atom at a distance between r and $r + dr$ from the nucleus of the atom. This question is an example of a

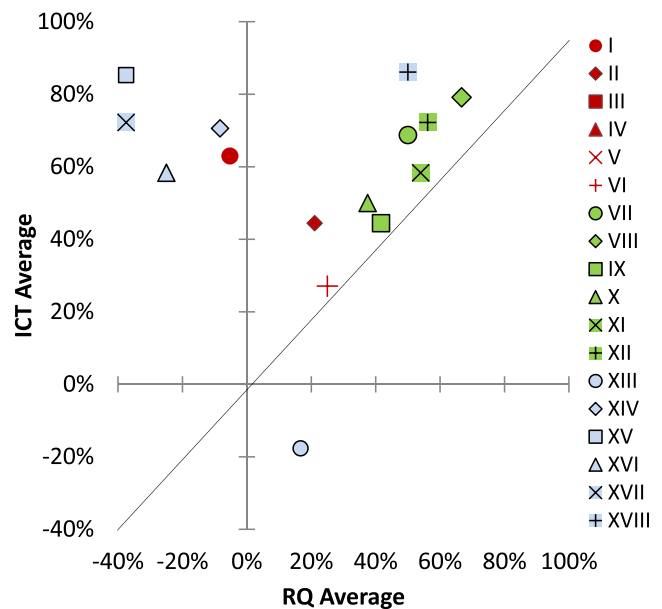


FIG. 5. Average scores on the comparison questions for the ICT vs the RQ. Error bars are not shown for clarity; the average standard error was $\pm 8\%$ for the RQ and $\pm 7\%$ for the ICT.

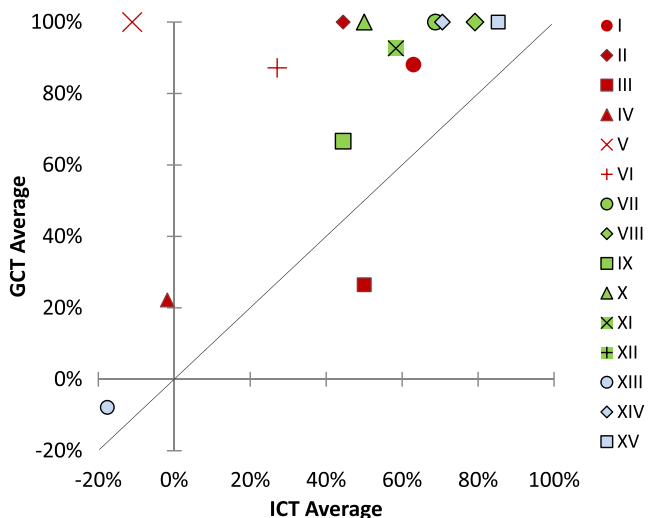


FIG. 6. Average scores on the comparison questions for the GCT vs the ICT. Error bars are not shown for clarity; the average standard error was $\pm 7\%$ for the ICT and $\pm 8\%$ for the GCT.

synthesis problem which is high on Bloom’s taxonomy [30]. In particular, question III involves synthesis of mathematical knowledge with knowledge of quantum physics. (Note that questions XII, XVI, XVII, and XVIII do not appear in Fig. 6 because there was no GCT for those questions.)

Figure 7 plots the average performances (averaged over all students) on each comparison question for the OQ vs GCT. The Pearson correlation coefficient $R^2 = 0.008$ indicates that there was no correlation between the performance on the GCT and the performance on the OQ. In particular, students performed reasonably well on most

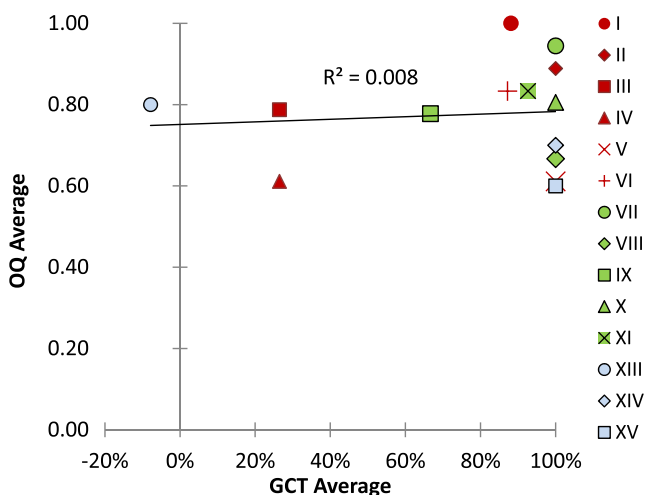


FIG. 7. Average scores on the individual comparison questions for the OQ vs the GCT, with linear regression and corresponding Pearson correlation coefficient ($R^2 = 0.008$). Error bars are not shown for clarity; the average standard error was $\pm 8\%$ for the GCT and ± 0.03 for the OQ.

questions on the OQ regardless of how well they performed on the GCT for the same topic. (Note that questions XII, XVI, XVII, and XVIII do not appear in Fig. 7 because there was no GCT for those questions.)

Finally, Fig. 8 compares the average performances (averaged over all students) on each comparison question for the RQ vs the OQ. A correlation coefficient $R^2 = 0.015$ indicates that there was no correlation between the performance on the RQ and the performance on the same topic on the OQ. In general, students benefitted from a variety of activities including lectures focusing on their difficulties, clicker questions and peer discussions, general class discussion after each clicker question, reflective and traditional homework assignments, etc. Figure 8 shows that students performed very well on OQ on topics such as those involved in answering question I, which asks students to state the Hund’s rule for determining the ground state spin configuration for a multielectron atom, i.e., the total spin angular momentum quantum number S is highest in the ground state. (Note that questions III, IV, and V do not appear in Fig. 8 because there was no RQ for those questions.)

C. Peer Instruction and clicker related results

In this section we will present some noteworthy findings related to students’ use of clickers in the advanced quantum mechanics class. These findings were used to answer research questions 7 and 8. We first define that “co-construction” of knowledge occurs when neither student who engaged in the peer interaction was able to answer the questions before the interaction, but both were able to answer them after working with a peer. In order to investigate whether co-construction of knowledge takes place, we analyzed performance of students on GCT depending upon the ICT performance of the peers in each

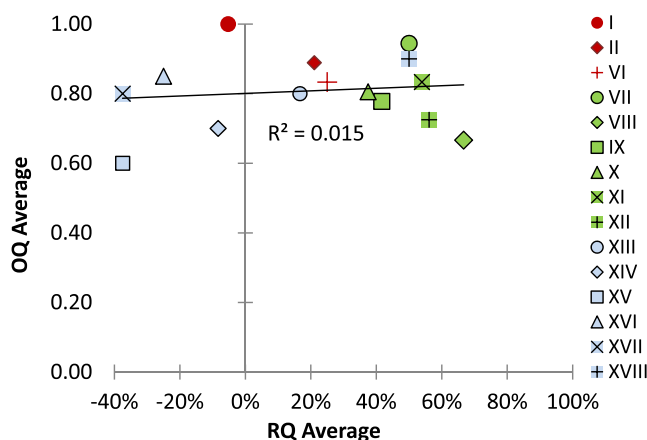


FIG. 8. Average scores on the individual comparison questions for the OQ vs the RQ, with linear regression and corresponding Pearson correlation coefficient ($R^2 = 0.015$). Error bars are not shown for clarity; the average standard error was $\pm 8\%$ for the RQ and ± 0.03 for the OQ.

TABLE II. Percentage of clicker questions for which (1) both group members answered incorrectly, (2) one member answered correctly and one incorrectly, and (3) both answered correctly, for the ICT and GCT.

		GCT			Total
		(1)	(2)	(3)	
ICT	(1)	61%	8%	31%	100%
	(2)	19%	4%	77%	100%
	(3)	2%	0%	98%	100%

group for all questions. Row 1 (with data) in Table II represents the situation in which all group members answered an ICT incorrectly and shows the percentages of all clicker questions for which all group members answered the corresponding GCT incorrectly (column 1 with data), one group member answered incorrectly (column 2 with data), and all group members answered correctly (column 3 with data). For example, row 1 (with data) in Table II shows that when all group members answered an ICT incorrectly they all answered the corresponding GCT correctly (i.e., they “co-constructed” knowledge) 31% of the time. Row 2 (with data) in Table II shows that when only one group member answered an ICT correctly, all group members answered a GCT correctly 77% of the time. Row 3 (with data) shows that when all group members answered an ICT correctly, all of them answered the corresponding GCT correctly 98% of the time.

Students in the QM course sometimes responded correctly to the ICT but then responded incorrectly to the corresponding GCT. Figure 9 shows a comparison of the fraction of times each student switched from a correct response on the ICT to an incorrect response on

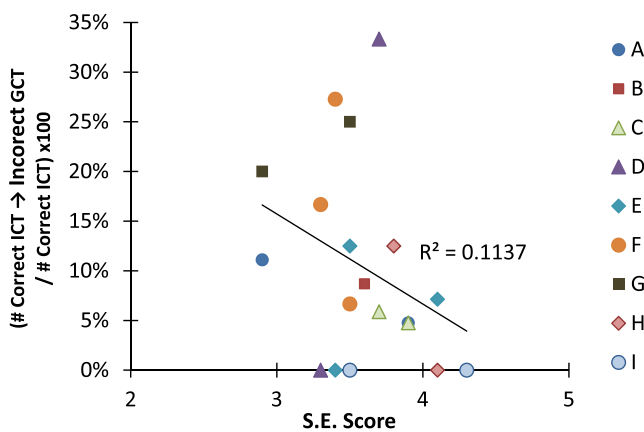


FIG. 9. The number of times each student switched from a correct ICT response to an incorrect GCT response divided by the total number of correct ICT responses for that student ($\times 100\%$) vs each student’s self-efficacy score. The average standard error for S.E. score was ± 0.086 , and for percentage of correct ICT switched to incorrect GCT was $\pm 2.30\%$.

the GCT vs each student’s reported self-efficacy (S.E.) score on the S.E. survey [9] administered at the end of the course. In other words, the y axis shows the number of correct ICT responses that were switched to incorrect GCT responses divided by the total number of correct ICT responses in percent for each student. Each data point in Fig. 9 represents an individual student, and colors denote the group to which the students belonged while discussing clicker questions. In response to Research Question 7 (“Is there a correlation between students’ reported self-efficacy and their tendency to switch from an initially correct response on an in-class clicker question to an incorrect response after peer discussion?”), Fig. 9 shows that there was no statistically significant correlation between higher S.E. score and a lower tendency to switch from the correct to incorrect answer on clicker questions after discussion with peers ($p = 0.157$). The Pearson correlation coefficient ($R^2 = 0.114$) in our study was comparable to that found in a prior study on self-efficacy in introductory physics [9]. However, since the number of students was large in introductory physics, the correlation was statistically significant in that study (unlike in this study). Also, the correlation between students’ S.E. scores and their performance on the final exam ($R^2 = 0.091$) in our study is not statistically significant ($p = 0.210$). On the other hand, when we compare the fraction of times students switched from correct ICT to incorrect GCT with each students’ performance on the final exam, the correlation ($R^2 = 0.255$) between the two is negative and is statistically significant ($p = 0.028$).

We also compared students’ average gains from the ICT to GCT for each of the first six weeks of class discussion, as shown in Fig. 10. We hypothesized that in addition to students having a better understanding of the group discussion protocol over time, student groups may become more cohesive and their discussions more productive as the semester goes, resulting in larger gains from ICT to GCT. Figure 10 shows that for the first five weeks of the course,

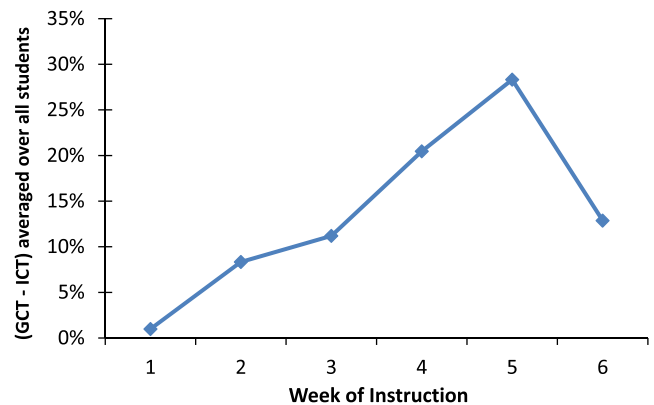


FIG. 10. (GCT—ICT) for each week of instruction (averaged over all students and all questions for that week). The average standard error was $\pm 8.56\%$.

the students on average improved more from ICT to GCT each week than they had in the previous week. We find that the increase in the amount of improvement in later weeks was due to a combination of more occurrences of co-construction of knowledge and fewer instances of switching from correct ICT to incorrect GCT. One possible reason for the dip in Fig. 10 in week 6 may be the difficulty associated with the concept of degenerate perturbation theory which was the focus.

Sometimes, a student who was present in class would not respond to one or more of the clicker questions, a trend that was more pronounced in the GCT than ICT. In particular, for a given student, the cumulative nonresponse rates for the entire semester was generally higher on the GCT than on the ICT. Since students received 80% of the points for participation and clicker responses are anonymous, it seems unlikely that they would not respond to a clicker question due to being unsure about the correct answer. Except for the first few weeks when students were still getting used to the various components of peer interaction (including familiarizing themselves with their peers and the instructor), we observed that most students participated in lively discussions with their peers after every ICT and then clicked for the GCT within the 1–2 minutes allotted for that discussion. One hypothesis for not clicking for the GCT (despite clicking for the ICT) is that students sometimes forgot to click for the GCT, e.g., due to being distracted by their discussion with their peers or not being used to the manner in which the instructor asked them to discuss their responses with their peers before the GCT. When students disagree with their peers about their responses in group discussion and get distracted in the heat of the discussion, the probability of not clicking increases. While other reasons are possible, this hypothesis is one that could result in a higher non-response rate on the GCT compared to the ICT. Figure 11 shows a comparison of how likely individual students were to not respond on the ICT vs the GCT. It shows the number of nonresponses on ICT and GCT questions for each student as a percentage of the total number of clicker questions given when the student was present. Each data point on the plot represents a particular student’s non-response percentage; e.g., the number of a student’s non-responses on GCT divided by the total number of times the students had the opportunity to answer a GCT clicker question along the vertical axis. We did not count non-responses for students who were absent on a particular day. As noted earlier, the attendance was typically greater than 80%. Figure 11 suggests that while a student who was more likely to not respond to ICT was also more likely to not respond to GCT, there was an overall tendency for most students to not respond to GCT more often than ICT.

Figure 12 shows the average nonresponse percentage for the whole class for each week of instruction. In response to research question 8 (“Are students equally likely to respond

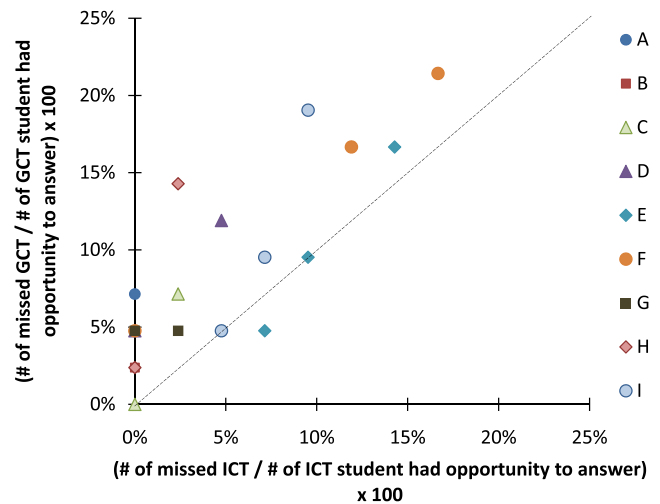


FIG. 11. The x axis denotes the number of times each student did not respond to an ICT divided by the number of ICT the student had the opportunity to answer $\times 100\%$; the y axis denotes the number of times each student did not respond to a GCT divided by the number of GCT the student had the opportunity to answer $\times 100\%$ for each student. The average standard error for missed ICT percentage was $\pm 1.19\%$ and for missed GCT percentage was $\pm 1.36\%$.

to in-class clicker questions at the beginning of the semester and later in the semester?”), Fig. 12 indicates that the first two weeks of the course had much higher nonresponse rates on both the ICT and GCT. However, the nonresponse rates declined greatly after the first two weeks of the course and stayed low for the rest of the course. A missed response to a clicker question is only counted as a nonresponse if the student was present in the classroom when the clicker question was given. There were roughly the same number of clicker questions (~ 6) given each week. It is possible that students needed time to familiarize themselves with the in-class clicker question procedures and with their peers

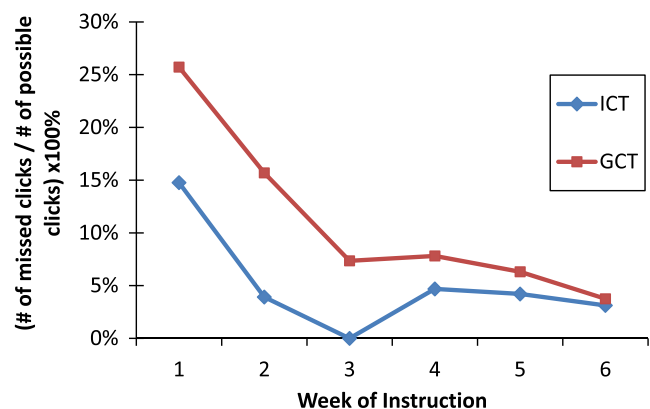


FIG. 12. Student nonresponse on ICT (blue) and GCT (red) as a percentage of total possible responses per week of instruction. The average standard error was $\pm 2.79\%$ for ICT and $\pm 3.86\%$ for GCT.

and develop the habit of regularly clicking in response to all clicker questions posed. Moreover, Fig. 12 is consistent with Fig. 11 in terms of the nonresponse rates being higher on average for the GCT than for the ICT.

V. DISCUSSION AND IMPLICATIONS

While the use of the JiTT approach at the introductory level has been a subject of prior studies [1,3], studies have not investigated its effectiveness when used in advanced courses such as quantum mechanics. Prior research suggests that similar to introductory mechanics, there is a large diversity both in the content knowledge and in the reasoning and self-regulatory skills of upper-level physics students in quantum mechanics [23]. The use of approaches that have been found effective at the introductory level may also be beneficial for advanced students in a quantum mechanics course. Our research suggests that lectures focusing on student difficulties, which were used in this case study as part of the JiTT-based instructional approach, resulted in improved performance on the ICT compared to the RQ for some students, but they were not sufficient for helping all students in the quantum mechanics course to have a time for telling [2]. Different students apparently experienced their time for telling at different stages of the instructional sequence and showed improved performance. However, a majority of students showed improved performance on various concepts at some point of time in the instructional design. Since the findings of this study suggest that an instructional design involving a variety of learning activities (including a JiTT approach and use of clicker questions with peer discussion) can lead to improvements in the performance of many advanced students in a QM course at different times, a related issue involves contemplating whether more students can be provided scaffolding support to learn and show improved performance earlier than they actually did. Instructors often work under tight time constraints to cover all of the relevant course materials. Learning activities which help a majority of students to have a time for telling as early as possible in an instructional sequence would be valuable since the later activities can be used to reinforce students' prior learning and help them apply learned concepts in diverse situations.

We now discuss some possible interpretations of some of the findings and implications for future research and pedagogical intervention.

(1) The pre-lecture JiTT activities did not sufficiently “prime” all students to learn from the lecture: Research by Schwartz *et al.* suggests that students who engage with learning materials in a deep and reflective manner are likely to be primed for future learning even via lectures [31]. Schwartz *et al.* have proposed invention tasks to prepare students for future learning via lecture because after their productive struggle students may be ready to learn from an instructor's lecture [31]. Also, research suggests that students who went through a productive failure cycle, in

which they worked in groups to solve complex ill-structured math problems without any scaffolding support, struggled to learn before a consolidation lecture by the instructor. However, those students significantly outperformed the students who did not struggle with the ill-structured problems before lectures [32].

It appears that the out-of-class activities in our investigation did not prepare all students sufficiently for future learning in the classroom setting. The average scores went from 20% on RQ to 48% on ICT after lectures specifically focusing on student difficulties. It is possible that the prelecture reading assignments did not cause some students to struggle productively, priming them to learn from the lectures and other in-class activities [31,32]. In their prelecture reading summaries, most students wrote at least a page summarizing what they read but it was unclear from those summaries what they had learned. Moreover, some of the difficulties that the students mentioned electronically about the prelecture reading did not convey deep productive struggle with the reading material. For example, one student wrote the following about his prelecture reading difficulty: “The most challenging part of this reading was definitely the section on degenerate perturbation theory. Perhaps I just need to work through it more, but I still don't feel very clear on why each step was taken.” This student did not delve deeply to specify what aspects of degenerate perturbation theory he found challenging, and only noted that he found the topic challenging. Another student wrote the following in their prelecture assignment related to quantum statistical mechanics: “One challenge this section posed is following Griffith's statement of the fundamental assumption of statistical mechanics (In thermal equilibrium, every distinct state with the same total energy, E , is equally probable). Indeed, whenever he suggests that the reader stop and think about what he just said, I can't help but feel like I missed something fundamental. I'm still not entirely sure that I understand why the assumption is a *deep* one, and it makes me question whether I'm thinking about the correct thing at all.” In quantum statistical mechanics, another student noted, “I thought that the most difficult and challenging part was the combinatorics of determining how many ways a distinct configuration can be achieved.” Another student wrote, “I found counting the states to be challenging.” These students were not the only ones who noted that they found the combinatorics challenging. In fact, 31% of the students mentioned combinatorics or counting states as their difficulty with the chapter on quantum statistical mechanics but they did not provide further elaboration on why it was challenging.

If the prelecture activities were more targeted and created opportunities for students to struggle productively with the material, they may have primed them better for learning from the lecture [31,32]. In particular, the JiTT approach may be more effective if instructors require students to elaborate more on their responses, which could prompt

students to be more cognitively engaged and reflect more deeply on the reading material before class and may better prime them to learn from the lectures. The reading assignment could ask the students more pointed questions, instead of only asking “What did you find challenging?” For example, the assignment could also ask “Why did you find it challenging?” or “Elaborate on the specific challenges you had with it.” Students could also be asked to write responses to specific conceptual questions related to the content of the reading. This type of specific questioning may help students to think more concretely about their difficulties and formulate more precise questions for which they would then actively seek answers in class. Another way to promote greater cognitive engagement in class could involve adding a question to each reflective homework assignment asking students what they learned from the in-class activities and how it helped them overcome difficulties with the part of the previous week’s reading they found challenging. Knowing that they will need to report on how they overcame their difficulties with each of their prelecture readings might prompt students to be better at self-regulating their learning and be more actively engaged with the lecture, clicker questions and in-class discussions.

(2) Some students lacked sufficient self-monitoring skills and intrinsic motivation to learn: Prior research suggests that even students in advanced quantum mechanics courses often vary in their motivation and in their problem-solving, reasoning, and self-regulation skills [23]. In particular, many advanced students in a quantum mechanics course lack the motivation and self-regulation skills to voluntarily engage with learning materials in a deep and reflective manner. They often focus only on their short term goals rather than on the long term goals such as developing robust knowledge structures and developing problem-solving, reasoning and metacognitive skills. Prior research also suggests that only providing students worked examples is insufficient [33], and effective approaches to learning involve students engaged in metacognition and self-monitoring while they solve problems [34–36].

The homework that was based upon prelecture reading and asked students to summarize what they read and what they found challenging was graded for completeness rather than correctness. This lack of grade incentive for correctness may have reduced the incentive for cognitive engagement with prelecture reading for some students. Providing a grade incentive for correctness may have encouraged those students to be more engaged with instructional activities. Similarly, some students may not have been cognitively engaged in learning from lectures (even though those lectures focused on their difficulties) since the in-class clicker questions were mainly graded for completeness rather than correctness. The grading policy for the reading quizzes and clicker questions was adopted in order to not penalize students for not knowing concepts they had

attempted to learn themselves either from the textbook or from the lecture recently. In particular, students were given 80% of the points for answering the clicker questions, even if they were not correct, and 100% for selecting the correct answer. The reading quizzes and clicker questions each counted for a bonus 2.5% to their grade and it was possible for students to get 4 out of 5 points simply by answering the question regardless of whether they were correct or not. While the grading policy was meant to encourage students to try their best on RQs and ICTs, it is possible that students were not reflecting as deeply on the prelecture reading and lecture (even though the lecture focused on their difficulties) as they would have if the grading for the RQ and ICT questions was for correctness instead of participation.

In fact, even graduate-level physics students report less motivation to complete out-of-class assignments if there is no grade incentive. For example, a similar JiTT strategy involving prelecture reading assignments before lectures was recently implemented in a first year graduate-level mathematical methods course in the physics department at the same university where this study took place. In class, the instructor focused on solving some problems on the board based upon the out-of-class reading in the first 30 minutes, and students were asked to work in groups of two in the last 20 minutes. The reading quizzes after prelecture reading were given online and were not graded, but it was suggested that students complete the prelecture quizzes in order to better prepare for the lecture which focused heavily on problem solving. At the end of the course, the students completed a course survey in which they were asked to select one of four statements describing their experience in the course regarding the prelecture reading assignments and quizzes. The percentage of students (out of 16 total students) who selected each statement is shown in parentheses: *Indicate which best describes your impression of the flipped course setup:*

- (A) *I usually completed the reading assignments and quiz and felt prepared when the topic was discussed in class. (18.75%)*
- (B) *I usually completed the reading assignments but found it difficult to absorb the information well enough to use it in class. (37.5%)*
- (C) *I tried to do all the reading assignments, but the lecture notes and book were not very good, and I learned little from them. (12.5%)*
- (D) *I often did not have enough time to complete the reading assignments in time. (50%)*

The percentages add up to 118.75% since some students selected more than one option. The important point here is that less than 20% of the students (3 out of 16) indicated that the reading assignments and quizzes prepared them so that they felt prepared when the topic was discussed in class, while 50% of the students indicated that they often didn’t complete the reading assignments. In the written

open-ended comments, some of the graduate students explicitly noted that since there was no grade incentive, the preclass reading assignment was their last priority among all the different things they had to do that week. Without grade incentive, only about half of the first-year physics graduate students took the time to complete the reading assignments even though the instructor specifically counseled them to regard the reading assignments as a valuable learning activity that would prepare them better for learning in class.

Returning to the undergraduates in our study, some students performed well in the OQs even though they did not perform well in the ICTs or GCTs. As mentioned in Sec. III, the OQ questions were graded for correctness, which may have incentivized students to prepare more for them. The grade incentive in conjunction with the homework and other discussions and study activities may partially explain the reasonable performance of most students on OQs.

Moreover, in future interventions, in addition to external motivation provided by grade incentives, students may benefit from instructors making an explicit effort to get student “buy in” at the beginning of the course (and several times during the course) by “framing” the instructional design and the importance of engaging actively with different activities, e.g., having a discussion about why the JiTT approach with peer instruction will help them learn, and why the students have to play a central role in their own learning with the instructor as their coach. An explicit class discussion (and preferably several throughout the course) related to self-efficacy and having a growth mindset rather than a fixed mindset may provide additional support to students to help them focus on learning and set appropriate goals for the course [37].

(3) Students had greater difficulty with some questions than others due to content involving a synthesis of different concepts. Student performance reached the ceiling for certain questions on the GCT involving simple application of principles, such as question II which concerns Hund’s rule for total orbital angular momentum. On the other hand, on average, students performed worse on the GCT after peer discussion than on the ICT on question III, which asked them to determine the probability of finding an electron in a hydrogen atom at a position between r and $r + dr$ from the nucleus. In future interventions, it may be advantageous to break down such multiple-choice problems that involve a synthesis of mathematic skills and quantum physics concepts (or synthesis of several quantum physics concepts) into separate multiple-choice subproblems (to be posed as ICT and GCT) to make them more manageable for students to think about and discuss with their peers. After students become proficient in the knowledge and skills involved in the subproblems, the original problem that combines them could then be posed as a clicker question.

(4) Reflection on optimizing the benefits of peer discussions: Prior research has shown that, even with minimal guidance from the instructors, students can benefit from peer discussions [6]. In particular, those who worked with peers not only outperformed an equivalent group of students who worked alone on the same task, but collaboration with a peer led to co-construction of knowledge in 29% of the cases [7]. In the present study, students were able to co-construct knowledge so that all members of the group chose the correct response on the GCT for 31% of the clicker questions for which all group members responded incorrectly on the ICT (see Table II).

Moreover, the comparison of students’ performance on the ICT vs the GCT shows that some student groups in QM appeared to benefit more from peer discussions than others. The cause for the differences was not immediately apparent. Consideration of the overall class grades of students in groups that were not as effective does not suggest any obvious academic reasons for the lack of benefit. We are also not aware of whether many of the students who worked together in groups were friends or worked with each other outside of class. Several factors foster productive group discussions. Interaction with peers provides opportunity for clarifying difficulties especially if there are diverse opinions. Also, students who have recently learned the concepts understand other students’ difficulties much better than the instructor and may be in a better position to help their peers, but students should be comfortable discussing their thought processes with their peers. In supportive environments, peer interaction generally helps all students since discussing and articulating concepts gives further clarity to thought processes and can help all students develop a better grasp of physics concepts. Also, since learning with peers is embedded in social context, it may be easier to retrieve that knowledge later.

To improve student learning further, investigations in the future can involve active learning using clicker questions and group problem solving for a greater portion of the class (or even the entire class with no lecture) [38]. In particular, in future interventions, the class could start with clicker questions focusing on student difficulties reported in the electronic feedback to the instructor instead of a lecture focusing on those difficulties first. The instructor could then clarify issues after a GCT related to the issue and follow it up with another clicker question. In this modified approach, more time in class would be devoted to clicker questions and peer discussions involving those questions rather than lectures focused on student difficulties. Topics that are easy for students as measured by the RQs could be omitted from clicker questions to save in-class time for discussion of more difficult topics.

VI. SUMMARY

Prior research suggests that students entering an upper-division quantum mechanics course share many

characteristics with introductory students in an introductory classical mechanics course [23]. The students vary greatly in their individual prior knowledge, problem-solving skills, mathematical skills, and motivation. Cognitive theory supports that instructors cannot force students to learn. Instead, they can motivate and engage students in the learning process and tailor activities to facilitate learning. The investigation using JiTT and Peer Instruction shows that overall, the instructional intervention led to improved student performance from the RQ to ICT and from the ICT to the GCT. If student performance is taken as the metric, the prelecture readings, lectures based on student difficulties, individual clicker questions and peer discussions varied in their usefulness for different students and for different topics and no single learning activity in the instructional sequence yield maximum learning gains for all students. In order for students in QM courses to maximally benefit from prelecture readings followed by in-class activities that build on the out-of-class activities, it will be useful to consider the suggestions for modifying the instructional intervention discussed in the preceding section in future investigations. Those modifications in the implementation of the instructional sequence may lead to more productive struggle and can better prepare students to have a time for telling [2,31,32].

Analysis of the ICT and GCT shows evidence of co-construction of knowledge in 31% of the cases. This level of co-construction is comparable to the level of co-construction previously reported in introductory physics [6]. We also find no significant correlation between higher student self-efficacy and tendency to switch from right to wrong answers in clicker responses after group discussion. In particular, although the Pearson correlation in this investigation was comparable to that found for introductory physics [9], since the number of students was large in introductory physics, the correlation was statistically significant in that case unlike in this study. Also, we find that the non-response rates on the in-class clicker questions started at or above 15% at the beginning of the semester but tended to decrease in later weeks of the course. One possible reason is that the students needed a few weeks to familiarize themselves with the in-class clicker procedures and group work. In addition, we find that for a given student, the cumulative nonresponse rates for the entire semester was generally higher on the GCT than on the ICT. These higher nonresponse rates on the GCT could partly be due to students disagreeing with their peers about their responses and getting distracted in the heat of the discussion and not clicking. To the best of our knowledge, these nonresponse rates have never been reported in introductory physics.

ACKNOWLEDGEMENTS

We thank the National Science Foundation for Award No. PHY-1505460.

APPENDIX A

This is a list of the comparison questions that were administered to the students. Each question is first given as it is found in the RQ, ICT, and GCT in multiple-choice format. In each case, the particular statement we are investigating and the responses corresponding to that statement are in bold. The question is then shown as in the open-ended retention quiz. The fully correct responses for the multiple-choice questions are given at the end of Appendix A. For questions VI-VIII, students were asked to treat the valence electrons within the free-electron gas model.

I. Choose all of the following statements that are correct according to Hund's rules:

(1) The state with the highest total spin (S) will have the lowest energy.

(2) The state with the highest total spin (S) will have the highest energy.

(3) The state with the highest total orbital angular momentum (L), consistent with overall symmetrization, will have the lowest energy.

A. 1 only

B. 2 only

C. 3 only

D. 1 and 3 only

E. 2 and 3 only

I. (open-ended quiz) Briefly explain the origin of the Hund's rules used for determining total spin angular momentum quantum number S for the ground state of multielectron atoms.

(Students who said that the state with the highest S will have the lowest energy received credit for this question regardless of how clear their full explanations were.)

II. Choose all of the following statements that are correct according to Hund's rules:

(1) The state with the highest total spin (S) will have the lowest energy.

(2) The state with the highest total spin (S) will have the highest energy.

(3) The state with the highest total orbital angular momentum (L), consistent with overall symmetrization, will have the lowest energy.

A. 1 only

B. 2 only

C. 3 only

D. 1 and 3 only

E. 2 and 3 only

II. (open-ended quiz) Briefly explain the origin of the Hund's rules used for determining total orbital angular momentum quantum number L for the ground state of multi-electron atoms.

(Students who said that the state with the highest L , consistent with overall symmetrization requirement, will have the lowest energy received credit for this question regardless of how clear their full explanations were.)

III. ψ_{nlm} are the energy eigenfunctions of the hydrogen atom (ignore spin). Choose all of the following statements that are correct about a hydrogen atom in the state $\psi_{220}(r, \theta, \phi) = R_{22}(r) \cdot Y_2^0(\theta, \phi)$. Note $Y_2^0(\theta, \phi) = Y_2^0(\theta)$. All notation is standard.

(1) The probability of finding the electron between r and $r+dr$ from the nucleus of the atom is $4\pi r^2 |\psi_{220}(r, \theta)|^2 dr$.

(2) **The probability of finding the electron between r and $r+dr$ from the nucleus of the atom is $\int_0^{2\pi} d\phi \int_0^\pi \sin \theta d\theta r^2 |\psi_{220}(r, \theta)|^2 dr$.**

(3) **The probability of finding the electron between r and $r+dr$ from the nucleus of the atom is $2\pi r^2 |R_{22}(r)|^2 dr \int_0^\pi |Y_2^0(\theta)|^2 \sin \theta d\theta$.**

A. 1 only

B. 2 only

C. 3 only

D. 2 and 3 only

E. None of the above

(Note: Since statements 2 and 3 are both true and answer the same question, when determining the POMP score we counted all students who selected either one or both of these statements.)

III. (open-ended quiz) The wave function for an electron in a hydrogen atom at time $t = 0$ is $\psi_{321}(r, \theta, \phi) = R_{32}(r) \cdot Y_2^1(\theta, \phi)$. What is the probability of finding the electron between r and $r+dr$ from the nucleus of the atom?

(Students who wrote a correct expression in terms of $R_{32}(r)$ or $\psi_{321}(r, \theta, \phi)$ received credit for this question.)

IV. Choose all of the following statements that are true about the helium atom:

(1) **The ground state of helium must have an antisymmetric spin configuration (singlet configurations).**

(2) The excited states of helium must have a symmetric spin configuration (triplet configuration).

(3) If the helium atom has a symmetric spin configuration (triplet configuration), it is known as orthohelium.

A. 1 only

B. 2 only

C. 3 only

D. 1 and 3 only

E. 2 and 3 only

IV. (open-ended quiz) If the electrons in a helium atom are in the ground state, write down the spin state of the two electrons, $\chi(s_1, s_2)$.

(Students who wrote an antisymmetric spin configuration received credit for this question.)

V. Choose all of the following statements that are true about the Helium atom:

(1) The ground state of helium must have an antisymmetric spin configuration (singlet configurations).

(2) **The excited states of helium must have a symmetric spin configuration (triplet configuration).**

(3) If the Helium atom has a symmetric spin configuration (triplet configuration), it is known as orthohelium.

A. 1 only

B. 2 only

C. 3 only

D. 1 and 3 only

E. 2 and 3 only

(Note: Since statement 2 is false about the excited states of Helium which comes in both symmetric and antisymmetric spin configurations, only students who did not choose statement 2 were counted as correct when determining the POMP score.)

V. (open-ended quiz) The excited spatial states of a helium atom consist of one electron in the hydrogenic ground state and the other electron in an excited state, $\psi_{nlm}(\vec{r}_1)\psi_{100}(\vec{r}_2)$. Explain why you agree or disagree with the following statement: For the composite wave function for the excited states, the spin state of the electrons, $\chi(s_1, s_2)$, must be symmetric.

(Students who disagreed with the statement received credit for this question regardless of the clarity of their explanation.)

VI. Cubes A and B with the same atom number density have N and $2N$ copper atoms, respectively. Choose all of the following statements that are correct.

(1) **At temperature $T = 0$ K, the Fermi energy of copper in cube B is larger than the Fermi energy of copper in cube A.**

(2) At temperature $T = 0$ K, the total energy of the valence electrons in cube B is larger than the total energy of the valence electrons in cube A.

(3) If we slowly compress the volume of cube A, the total energy of the valence electrons in cube A will increase.

A. 1 only

B. 2 only

C. 1 and 2 only

D. 2 and 3 only

E. all of the above

(Note: Since statement 1 is false, only students who did not choose statement 1 were counted as correct when determining the POMP score.)

VI. (open-ended quiz) Cubes A and B with the same atom number density have N and $2N$ copper atoms, respectively. At temperature $T = 0$ K, which cube has the higher Fermi energy?

(Students who said either that both cubes have the same Fermi energy or that the Fermi energy of cube B is NOT larger received credit for this question.)

VII. Cubes A and B with the same atom number density have N and $2N$ copper atoms, respectively. Choose all of the following statements that are correct.

(1) At temperature $T = 0$ K, the Fermi energy of copper in cube B is larger than the Fermi energy of copper in cube A.

(2) **At temperature $T = 0$ K, the total energy of the valence electrons in cube B is larger than the total energy of the valence electrons in cube A.**

(3) If we slowly compress the volume of cube A, the total energy of the valence electrons in cube A will increase.

- A. 1 only
- B. 2 only**
- C. 1 and 2 only
- D. 2 and 3 only
- E. all of the above

VII. (open-ended quiz) Cubes A and B with the same atom number density have N and $2N$ copper atoms, respectively. At temperature $T = 0$ K, which cube has the higher total energy associated with the valence electrons?

(Students who said that cube B has the higher total energy received credit for this question.)

VIII. Cubes A and B with the same atom number density have N and $2N$ sodium atoms, respectively. Choose all of the following statements that are correct.

(1) At temperature $T = 0$ K, the Fermi energy of sodium in cube B is larger than the Fermi energy of sodium in cube A.

(2) At temperature $T = 0$ K, the total energy of the valence electrons in cube B is larger than the total energy of the valence electrons in cube A.

(3) If we slowly compress the volume of cube A, the total energy of the valence electrons in cube A will increase.

- A. 1 only
- B. 2 only
- C. 1 and 2 only
- D. 2 and 3 only**
- E. all of the above

VIII. (open-ended quiz) Cube A has N copper atoms. How will the total energy of this solid associated with the valence electrons change if you increase the volume of the solid keeping the total number of atoms fixed?

(Students who said that the total energy will decrease as volume increases received credit for this question.)

IX. We have three noninteracting particles in a one-dimensional infinite square well. The energy of the three particle system is $E_{n_1 n_2 n_3} = (n_1^2 + n_2^2 + n_3^2)E_0$, in which E_0 is the ground state energy for a single particle system. If the total energy is $E = 27E_0$ and the particles are *distinguishable*, choose all of the following statements that are correct. Note: Three positive numbers, the sum of whose squares gives 27 are (1, 1, 5), (1, 5, 1), (5, 1, 1) and (3, 3, 3). Students were familiar with the notation.

(1) There are 4 distinct states of this many particle system with the energy $E = 27E_0$.

(2) If we randomly measure the energy of *one particle* when the total energy of the three particle system is $27E_0$, the probability of obtaining E_0 is $2/3$.

(3) If we randomly measure the energy of one particle when the total energy of the three particle system is $27E_0$, the probability of obtaining E_0 is $1/2$.

- A. 1 only
- B. 3 only**
- C. 1 and 2
- D. 1 and 3 only**
- E. none of the above

IX. (open-ended quiz) There are two non-interacting particles in a one-dimensional infinite square well. The total energy of the two particle system is $E_{n_1 n_2} = (n_1^2 + n_2^2)E_0$, in which E_0 is the ground state energy for one particle. The total energy of the system is $E = 50E_0$. If the particles are distinguishable particles and you randomly measure the energy of one particle, what is the probability of measuring $25E_0$? Note: Two positive numbers, the sum of whose squares gives 50, are (1, 7), (7, 1), and (5, 5).

(Students who said that the probability is $1/3$ received credit for this question.)

X. We have three non-interacting particles in a one-dimensional infinite square well. The total energy for the three particle system is $E_{n_1 n_2 n_3} = (n_1^2 + n_2^2 + n_3^2)E_0$, in which E_0 is the ground state energy for a single particle system. If the total energy is $E = 27E_0$ and the particles are *identical*, choose all of the following statements that are correct. Note: Three positive numbers, the sum of whose squares gives 27 are (1, 1, 5), (1, 5, 1), (5, 1, 1) and (3, 3, 3).

- (1) The particles can be either bosons or fermions.
- (2) If the particles are spinless bosons, there are 4 distinct states in this system.

(3) If the particles are bosons, when we measure the energy of one particle at random, the probability of obtaining $9E_0$ is $1/2$.

- A. 2 only
- B. 3 only**
- C. 1 and 2 only
- D. 1 and 3 only**
- E. all of the above

X. (open-ended quiz) There are two noninteracting particles in a one-dimensional infinite square well. The total energy of the two particle system is $E_{n_1 n_2} = (n_1^2 + n_2^2)E_0$, in which E_0 is the ground state energy for one particle. The total energy of the system is $E = 50E_0$. If the particles are identical bosons and you randomly measure the energy of one particle, what is the probability of measuring $25E_0$? Note: Two positive numbers, the sum of whose squares gives 50, are (1, 7), (7, 1), and (5, 5).

(Students who said that probability is $1/2$ received credit for this question.)

XI. Suppose you have three particles and four distinct one-particle states $\psi_1(x)$, $\psi_2(x)$, $\psi_3(x)$, and $\psi_4(x)$. How many different three-particle states can you construct if the particles are fermions?

- A. 4^3
 B. $\frac{4!}{3!1!} \cdot 4^3$
 C. $\frac{4!}{3!1!}$
 D. $\frac{6!}{3!3!}$

E. None of the above.

XI. (open-ended quiz) Suppose you have three particles and four distinct one-particle states $\psi_1(x)$, $\psi_2(x)$, $\psi_3(x)$, and $\psi_4(x)$. How many different three-particle states can you construct if the particles are identical fermions?

(Students who wrote either 4 or $4!(3!1!)$ received credit for this question.)

XII. Suppose \hat{H}^0 and \hat{H}' commute with each other. Choose all of the following statements that are correct.

(1) If \hat{H}^0 is diagonal in a given basis and there is no degeneracy in the eigenvalue spectrum of \hat{H}^0 and \hat{H}' , then \hat{H}' must be diagonal in that basis.

(2) If \hat{H}^0 is diagonal in a given basis and there is a degeneracy in the eigenvalue spectrum of \hat{H}^0 , then \hat{H}' must be diagonal in that basis.

(3) We can always find a special basis in which both \hat{H}^0 and \hat{H}' are diagonal simultaneously.

- A. **1 only**
 B. 1 and 2 only
 C. **1 and 3 only**
 D. 2 and 3 only
 E. All of the above

(Note: Since statement 2 is false, only students who did not choose statement 2 were counted as correct when determining the POMP score.)

XII. (open-ended quiz) Suppose that in an N dimensional vector space ($N > 2$), the energy spectrum of the unperturbed Hamiltonian \hat{H}_0 has a two-fold degeneracy. A perturbation \hat{H}' acts on this system. \hat{H}_0 and \hat{H}' commute with each other. Consider the following statement: "If we choose a basis in which \hat{H}_0 is diagonal, \hat{H}' MUST be diagonal in that basis." Explain why you agree or disagree with this statement.

(Students who disagreed with the statement received credit for this question.)

XIII. Suppose the unperturbed Hamiltonian \hat{H}^0 is two-fold degenerate, i.e., $\hat{H}^0\psi_a^0 = E_1^0\psi_a^0$, $\hat{H}^0\psi_b^0 = E_1^0\psi_b^0$, $\langle\psi_a^0|\psi_b^0\rangle = 0$. A perturbation \hat{H}' acts on this system and a Hermitian operator \hat{A} commutes with both \hat{H}^0 and \hat{H}' . Choose all of the following statements that are correct.

- (1) \hat{H}^0 and \hat{H}' must commute with each other.
 (2) If ψ_a^0 and ψ_b^0 are degenerate eigenstates of \hat{A} , they must be "good" states for finding perturbative corrections to the energy and wavefunction due to \hat{H}' .
 (3) If ψ_a^0 and ψ_b^0 are non-degenerate eigenstates of \hat{A} , they must be "good" states.

- A. 1 only
 B. **2 only**
 C. **3 only**

- D. 1 and 2 only
 E. 1 and 3 only

(Note: Since statement 1 is false, only students who did not choose statement 1 were counted as correct when determining the POMP score.)

XIII. (open-ended quiz) Consider the following statement: "If \hat{H}_0 and \hat{H}' each commute with a third Hermitian operator \hat{A} , then they must commute with each other." Explain why you agree or disagree with this statement.

(Students who disagreed with the statement received credit for this question.)

XIV. Consider the Hamiltonian $\hat{H}^0 + \varepsilon\hat{H}' = V_0 \begin{pmatrix} 1 - \varepsilon & \varepsilon & 0 \\ \varepsilon & 1 & \varepsilon \\ 0 & \varepsilon & 2 \end{pmatrix}$, where $\varepsilon \ll 1$. The basis vectors for

the matrix $|a\rangle$, $|b\rangle$, and $|c\rangle$ chosen in that order are the energy eigenstates of the unperturbed Hamiltonian \hat{H}^0 ($\varepsilon = 0$). Choose all of the following statements that are correct.

- (1) $|a\rangle$ is a "good" state for the perturbation \hat{H}' .
 (2) $|c\rangle$ is a "good" state for the perturbation \hat{H}' .
 (3) In the degenerate subspace of \hat{H}^0 , the perturbation

matrix is $V_0 \begin{pmatrix} -\varepsilon & 0 \\ 0 & 0 \end{pmatrix}$.

- A. 1 only
 B. **2 only**
 C. 1 and 3 only
 D. **2 and 3 only**
 E. All of the above.

(Note: Since statement 1 is false, only students who did not choose statement 1 were counted as correct when determining the POMP score.)

XIV. (open-ended quiz) Consider the Hamiltonian $\hat{H}^0 + \varepsilon\hat{H}' = V_0 \begin{pmatrix} 1 - \varepsilon & \varepsilon & 0 \\ \varepsilon & 1 & \varepsilon \\ 0 & \varepsilon & 2 \end{pmatrix}$, where $\varepsilon \ll 1$. The basis vec-

tors for the matrix chosen in the order $|a\rangle$, $|b\rangle$, and $|c\rangle$ are the energy eigenstates of the unperturbed Hamiltonian \hat{H}^0 ($\varepsilon = 0$). Explain in words how you would find the "good" basis states for the perturbation \hat{H}' and the first order corrections to the energy. Do not carry out the calculation.

(Students who either said that $|a\rangle$ is not a "good" basis state or correctly described how they would find "good" basis states received credit for this comparison question.)

XV. Consider the Hamiltonian $\hat{H}^0 + \varepsilon\hat{H}' = V_0 \begin{pmatrix} 1 - \varepsilon & \varepsilon & 0 \\ \varepsilon & 1 & \varepsilon \\ 0 & \varepsilon & 2 \end{pmatrix}$, where $\varepsilon \ll 1$. The basis vectors for

the matrix $|a\rangle$, $|b\rangle$, and $|c\rangle$ chosen in that order are the energy eigenstates of the unperturbed Hamiltonian \hat{H}^0 ($\varepsilon = 0$). Choose all of the following statements that are correct.

- (1) $|a\rangle$ is a “good” state for the perturbation \hat{H}' .
 (2) $|c\rangle$ is a “good” state for the perturbation \hat{H}' .
 (3) In the degenerate subspace of \hat{H}^0 , the perturbation

matrix is $V_0 \begin{pmatrix} -\epsilon & 0 \\ 0 & 0 \end{pmatrix}$.

- A. 1 only
 B. 2 only
 C. 1 and 3 only
 D. 2 and 3 only
 E. All of the above.

XV. (open-ended quiz) Consider the Hamiltonian $\hat{H}^0 + \epsilon \hat{H}' = V_0 \begin{pmatrix} 1 - \epsilon & \epsilon & 0 \\ \epsilon & 1 & \epsilon \\ 0 & \epsilon & 2 \end{pmatrix}$, where $\epsilon \ll 1$. The basis vec-

tors for the matrix chosen in the order $|a\rangle$, $|b\rangle$, and $|c\rangle$ are the energy eigenstates of the unperturbed Hamiltonian \hat{H}^0 ($\epsilon = 0$). Explain in words how you would find the “good” basis states for the perturbation \hat{H}' and the first order corrections to the energy. Do not carry out the calculation.

(Students who either said that $|c\rangle$ is a “good” basis state or correctly described how to find the other “good” basis states received credit for this comparison question.)

XVI. A perturbation \hat{H}' acts on a hydrogen atom with the unperturbed Hamiltonian $\hat{H}^0 = -\frac{\hbar^2}{2m} \nabla^2 - \frac{e^2}{4\pi\epsilon_0 r}$. To calculate the perturbative corrections, we use the coupled representation $|n, l, s, j, m_j\rangle$ as the basis vectors. Choose all of the following statements that are correct. (Students were familiar with the notation).

(1) If $\hat{H}' = \alpha \hat{L}_z$, where α is a suitable constant, we can calculate the first order corrections as $E^1 = \langle n, l, s, j, m_j | \hat{H}' | n, l, s, j, m_j \rangle$.

(2) If $\hat{H}' = \alpha \delta(r)$, the first order correction to energy is $E^1 = \langle n, l, s, j, m_j | \hat{H}' | n, l, s, j, m_j \rangle$.

(3) If $\hat{H}' = \alpha \hat{J}_z$ (z component of $\vec{J} = \vec{L} + \vec{S}$) we can calculate the first order correction as $E^1 = \langle n, l, s, j, m_j | \hat{H}' | n, l, s, j, m_j \rangle$.

- A. 1 only
 B. 1 and 2 only
 C. 1 and 3 only
 D. 2 and 3 only
 E. All of the above

(Note: Since statement 1 is false, only students who did not choose statement 1 were counted as correct when determining the POMP score.)

XVI. (open-ended quiz) A perturbation \hat{H}' acts on a hydrogen atom with the unperturbed Hamiltonian

$\hat{H}^0 = -\frac{\hbar^2}{2m} \nabla^2 - \frac{e^2}{4\pi\epsilon_0 r}$. For the perturbation $\hat{H}' = \alpha \hat{L}_z$, state whether to find the first order correction to the energy, coupled representation or uncoupled representation forms a good basis (or whether both coupled and uncoupled representations form a good basis, or neither representation forms a good basis).

(Students who said that the coupled representation does NOT form a good basis in this case received credit for this question.)

XVII. A perturbation \hat{H}' acts on a hydrogen atom with the unperturbed Hamiltonian $\hat{H}^0 = -\frac{\hbar^2}{2m} \nabla^2 - \frac{e^2}{4\pi\epsilon_0 r}$. To calculate the perturbative corrections, we use the coupled representation $|n, l, s, j, m_j\rangle$ as the basis vectors. Choose all of the following statements that are correct.

(1) If $\hat{H}' = \alpha \hat{L}_z$, where α is a suitable constant, we can calculate the first order corrections as $E^1 = \langle n, l, s, j, m_j | \hat{H}' | n, l, s, j, m_j \rangle$.

(2) If $\hat{H}' = \alpha \delta(r)$, the first order correction to energy is $E^1 = \langle n, l, s, j, m_j | \hat{H}' | n, l, s, j, m_j \rangle$.

(3) If $\hat{H}' = \alpha \hat{J}_z$ (z component of $\vec{J} = \vec{L} + \vec{S}$) we can calculate the first order correction as $E^1 = \langle n, l, s, j, m_j | \hat{H}' | n, l, s, j, m_j \rangle$.

- A. 1 only
 B. 1 and 2 only
 C. 1 and 3 only
 D. 2 and 3 only
 E. All of the above

XVII. (open-ended quiz) A perturbation \hat{H}' acts on a hydrogen atom with the unperturbed Hamiltonian $\hat{H}^0 = -\frac{\hbar^2}{2m} \nabla^2 - \frac{e^2}{4\pi\epsilon_0 r}$. For the perturbation $\hat{H}' = \alpha \delta(r)$, state whether to find the first order correction to the energy, coupled representation or uncoupled representation forms a good basis (or whether both coupled and uncoupled representations form a good basis, or neither representation forms a good basis).

(Students who said that the coupled representation forms a good basis in this case received credit for this question for comparison purposes although the correct answer for this question is both coupled and uncoupled representations.)

XVIII. A perturbation \hat{H}' acts on a hydrogen atom with the unperturbed Hamiltonian $\hat{H}^0 = -\frac{\hbar^2}{2m} \nabla^2 - \frac{e^2}{4\pi\epsilon_0 r}$. To calculate the perturbative corrections, we use the coupled representation $|n, l, s, j, m_j\rangle$ as the basis vectors. Choose all of the following statements that are correct.

(1) If $\hat{H}' = \alpha \hat{L}_z$, where α is a suitable constant, we can calculate the first order corrections as $E^1 = \langle n, l, s, j, m_j | \hat{H}' | n, l, s, j, m_j \rangle$.

(2) If $\hat{H}' = \alpha \delta(r)$, the first order correction to energy is $E^1 = \langle n, l, s, j, m_j | \hat{H}' | n, l, s, j, m_j \rangle$.

(3) If $\hat{H}' = \alpha \hat{J}_z$ (z component of $\vec{J} = \vec{L} + \vec{S}$) we can calculate the first order correction as $E^1 = \langle n, l, s, j, m_j | \hat{H}' | n, l, s, j, m_j \rangle$.

- A. 1 only
 B. 1 and 2 only
 C. 1 and 3 only
 D. 2 and 3 only
 E. All of the above

XVIII. (open-ended quiz) A perturbation \hat{H}' acts on a hydrogen atom with the unperturbed Hamiltonian

$$\hat{H}^0 = -\frac{\hbar^2}{2m}\nabla^2 - \frac{e^2}{4\pi\epsilon_0 r}.$$

For the perturbation $\hat{H}' = \alpha\hat{J}_z$, state whether to find the first order correction to the energy, coupled representation or uncoupled representation forms a good basis (or whether both coupled and uncoupled representations form a good basis, or neither representation forms a good basis).

(Students who said that the coupled representation forms a good basis in this case received credit for this question for comparison purposes.)

Fully correct responses: I.-D, II.-D, III.-D, IV.-D, V.-D, VI.-D, VII.-D, VIII.-D, IX.-D, X.-B, XI.-C, XII.-C, XIII.-C, XIV.-B, XV.-B, XVI.-D, XVII.-D, XVIII.-D.

APPENDIX B

This is a list of the seven questions for which students performed well enough on the ICT that they were not given as a GCT (which usually meant the score was generally greater than 75% on the ICT). The correct response for each question is in bold.

[1] Choose all of the following statements that are correct about the free electron gas model.

(1) The free electron gas model takes into account electron-electron repulsion.

(2) The free electron gas model ignores the charge of the free electrons.

(3) The “free electrons” in the free electron gas model refer to all of the electrons in each atom in the solid.

A. 1 only

B. 2 only

C. 1 and 2 only

D. 2 and 3 only

E. None of the above.

[2] Choose all of the following statements that are correct about the k -space for a free electron gas model in a three dimensional solid. (Students were familiar with the notation and convention used).

(1) In k space, each point (k_x, k_y, k_z) represents a wave vector.

(2) Because the momentum can be written as $p = \hbar k$, the k space can be treated as the momentum space.

(3) In k space, each single-particle state occupies a volume $\frac{\pi^3}{V}$, where V is the volume of the box in which the free electrons are.

A. 1 only

B. 2 only

C. 1 and 2 only

D. 2 and 3 only

E. All of the above

[3] Choose all of the following statements that are correct about doping of host materials. Here, q is the number of valence electrons contributed by each atom in a solid.

1) The only way to make an insulator behave as a semiconductor is to dope it with a small amount of atoms that have a larger q than the insulator.

2) Doping an insulator with a few atoms of larger q than the insulator will put extra electrons into the next higher energy band.

3) Doping an insulator with a few atoms of smaller q than the insulator will create “holes” in the previously filled energy band.

A. 1 only

B. 1 and 2 only

C. 1 and 3 only

D. 2 and 3 only

E. All of the above

[4] Choose all of the following statements that are correct about any configuration of a many-particle system (consisting of non-interacting particles) in thermal equilibrium:

1) It is a set of occupation numbers for all single-particle states.

2) The most probable configuration is the configuration with the maximum number of distinct many-particle states.

3) As the total number of particles gets large, the most probable configuration becomes so overwhelmingly probable that one can ignore other configurations.

A. 1 only

B. 1 and 2 only

C. 1 and 3 only

D. 2 and 3 only

E. All of the above

[5] Suppose a bookcase has three shelves, and each shelf could contain a number of books. If we want to put 4 books randomly selected from 10 different books into this bookcase, how many ways can we do that?

A. $\frac{10!}{4!} \cdot \frac{3^4}{3!}$

B. $\frac{10!}{4!} \cdot \frac{4^3}{3!}$

C. $\frac{10!}{4!(10-4)!} \cdot 3^4$

D. $\frac{10!}{4!(10-4)!} \cdot 4^3$

E. None of the above

[6] $Q(N_1, N_2, N_3, \dots)$ represents the number of microstates (distinct states) in a particular configuration (N_1, N_2, N_3, \dots) . To find the configuration for which the number of microstates $Q(N_1, N_2, N_3, \dots)$ is maximum, we can use the method of Lagrange multipliers. (Students were familiar with the notations).

Define

$$G = \ln Q + \alpha[N - \sum_{n=1}^{\infty} N_n] + \beta[E - \sum_{n=1}^{\infty} N_n E_n].$$

Choose all of the following statements that are correct about maximizing $Q(N_1, N_2, N_3, \dots)$.

1) If the particles are distinguishable, the most probable occupation numbers are $N_n = \frac{1}{e^{(\alpha+\beta E_n)} + 1}$.

2) The Lagrange multiplier α is related to the chemical potential as $\alpha = \frac{-\mu(T)}{k_B T}$.

3) The Lagrange multiplier β is related to the temperature as $\beta = \frac{1}{k_B T}$, where k_B is Boltzmann's constant.

- A. 1 only
- B. 1 and 2 only
- C. 1 and 3 only
- D. 2 and 3 only**
- E. All of the above.

[7] The Maxwell-Boltzmann distribution (MBD) is $n(\epsilon) = e^{-(\epsilon-\mu)/k_B T}$. Choose all of the following statements that are correct.

1) The Maxwell-Boltzmann distribution applies to distinguishable particles.

- 2) ϵ represents the energy of a single-particle state.
- 3) In the high temperature limit, Fermi-Dirac and Bose-Einstein statistics reduce to the MBD.
- A. 1 only
- B. 2 only
- C. 1 and 2 only
- D. 1 and 3 only
- E. All of the above**

-
- [1] G. Novak, E. Patterson, A. Gavrin, and W. Christian, *Just-In-Time-Teaching: Blending Active Learning with Web Technology* (Prentice Hall, Englewood Cliffs, NJ, 1999).
 - [2] D. Schwartz and J. Bransford, A time for telling, *Cognit. Instr.* **16**, 475 (1998).
 - [3] S. Formica, J. Easley, and M. Spraker, Transforming common-sense beliefs into Newtonian thinking through Just-In-Time Teaching, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020106 (2010).
 - [4] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).
 - [5] 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); P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, *Am. J. Phys.* **60**, 637 (1992).
 - [6] C. Singh, Impact of peer interaction on conceptual test performance, *Am. J. Phys.* **73**, 446 (2005); arXiv:physics/0207106.
 - [7] A. Mason and C. Singh, Helping students learn effective problem solving strategies by reflecting with peers, *Am. J. Phys.* **78**, 748 (2010); Using reflection with peers to help students learn effective problem solving strategies, *AIP Conf. Proc.* **1289**, 41 (2010); Impact of guided reflection with peers on the development of effective problem solving strategies and physics learning, *Phys. Teach.* **54**, 295 (2016).
 - [8] D. Zingaro, *Peer instruction contributes to self-efficacy in CSI, SIGCSE '14: Proceedings of the 45th ACM Technical Symposium on Computer Science Education* (2014), p. 373–378.
 - [9] K. Miller, J. Schell, A. Ho, B. Lukoff, and E. Mazur, Response switching and self-efficacy in Peer Instruction classrooms, *Phys. Rev. ST Phys. Educ. Res.* **11**, 010104 (2015).
 - [10] A. Collins, J. Brown, and S. Newman, Cognitive Apprenticeship: Teaching the crafts of reading, writing and mathematics, *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, edited by L. B. Resnick (Lawrence Erlbaum, Hillsdale, NJ, 1989), p. 453–494.
 - [11] For example, see J. Docktor and J. Mestre, Synthesis of discipline-based education research in physics, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020119 (2014).
 - [12] C. Singh, What can we learn from PER: Physics Education Research?, *Phys. Teach.* **52**, 568 (2014).
 - [13] P. Jolly, D. Zollman, S. Rebello, and A. Dimitrova, Visualizing potential energy diagrams, *Am. J. Phys.* **66**, 57 (1998).
 - [14] M. C. Wittmann, R. N. Steinberg, and E. F. Redish, Investigating student understanding of quantum physics: Spontaneous models of conductivity, *Am. J. Phys.* **70**, 218 (2002).
 - [15] J. Morgan and M. Wittmann, Examining the evolution of student ideas about quantum tunneling, *AIP Conf. Proc.* **818**, 73 (2006).
 - [16] C. Manogue and E. Gire, D. McIntyre, and J. Tate, Representations for a spins first approach to quantum mechanics, *AIP Conf. Proc.* **1413**, 55 (2012).
 - [17] E. Gire and C. Manogue, Making sense of operators, eigenstates, and quantum measurements, *AIP Conf. Proc.* **1413**, 195 (2012).
 - [18] C. Singh, Student understanding of quantum mechanics, *Am. J. Phys.* **69**, 885 (2001); C. Singh, M. Belloni, and W. Christian, Improving students' understanding of quantum mechanics, *Phys. Today* **59**, No. 8, 43 (2006); C. Singh, Assessing and improving student understanding of quantum mechanics, *AIP Conf. Proc.* **818**, 69 (2006) [arXiv:1602.05664]; Student difficulties with quantum mechanics formalism, *AIP Conf. Proc.* **883**, 185 (2007); Helping students learn quantum mechanics for quantum computing, *AIP Conf. Proc.* **883**, 42 (2007); G. Zhu and C. Singh, Improving students' understanding of quantum mechanics via Stern-Gerlach experiment, *Am. J. Phys.* **79**, 499 (2011); A. J. Mason and C. Singh, Do advanced students learn from their mistakes without explicit intervention?, *Am. J. Phys.* **78**, 760 (2010); C. Singh and E. Marshman, Review of student difficulties in upper-level quantum mechanics, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020117 (2015).
 - [19] M. Belloni, W. Christian, and D. Brown, Open source physics curricular material for quantum mechanics, *Comput. Sci. Eng.* **9**, 24 (2007); C. Singh and G. Zhu, Improving students' understanding of quantum mechanics

- by using peer instruction tools, [arXiv:1602.05461](#); S. Siddiqui and C. Singh, Surveying instructors' attitudes and approaches to teaching quantum mechanics, [arXiv:1603.06025](#).
- [20] 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).
- [21] P.J. Emigh, G. Passante, and P.S. Shaffer, Investigating student understanding of basic quantum mechanics in the context of time-dependent perturbation theory, in *2013 PERC Proceedings, Portland, OR, July 17–18, 2013*, edited by P.V. Engelhardt, A.D. Churukian, and D.L. Jones (AAPT, College Park, MD, 2014); Student ability to distinguish between superposition states and mixed states in quantum mechanics, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020135 (2015); Examining student ideas about energy measurements on quantum states across undergraduate and graduate level, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020111 (2015); Student understanding of time dependence in quantum mechanics, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020112 (2015).
- [22] C. Singh, Transfer of learning in quantum mechanics, *AIP Conf. Proc.* **790**, 23 (2005); Student understanding of quantum mechanics at the beginning of graduate instruction, *Am. J. Phys.* **76**, 277 (2008); Interactive learning tutorials on quantum mechanics, *Am. J. Phys.* **76**, 400 (2008); C. Singh and G. Zhu, Cognitive issues in learning advanced physics: An example from quantum mechanics, *AIP Conf. Proc.* **1179**, 63 (2009); G. Zhu and C. Singh, Surveying students' understanding of quantum mechanics in one spatial dimension, *Am. J. Phys.* **80**, 252 (2012); Improving students' understanding of quantum measurement I: Investigation of difficulties, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010117 (2012); Improving students' understanding of quantum measurement II: Development of Research-based learning tools, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010118 (2012); Improving Student understanding of addition of angular momentum in quantum mechanics, *Phys. Rev. ST Phys. Educ. Res.* **9**, 010101 (2013); S. Y. Lin and C. Singh, Assessing expertise in quantum mechanics using categorization task, *AIP Conf. Proc.* **1179**, 185 (2009); Categorization of quantum mechanics problems by professors and students, *Eur. J. Phys.* **31**, 57 (2010).
- [23] E. Marshman and C. Singh, A framework for understanding the patterns of student reasoning difficulties in quantum mechanics, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020119 (2015); Interactive tutorial to improve student understanding of single photon experiments involving a Mach-Zehnder Interferometer, *Eur. J. Phys.* **37**, 024001 (2016); 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); S. DeVore and C. Singh, Development of an interactive tutorial on quantum key distribution, [arXiv:1601.00730](#); C. Singh, E. Marshman, Developing an Interactive Tutorial on a Mach-Zehnder Interferometer with Single Photons, [arXiv:1510.01319](#); Investigating student difficulties with Dirac Notation, [arXiv:1510.01296](#); E. Marshman and C. Singh, Developing an interactive tutorial on a quantum eraser, [arXiv:1510.01300](#); Student difficulties with quantum states while translating state vectors in Dirac notation to wave functions in position and momentum representations, [arXiv:1509.04084](#); Investigating student difficulties with time-dependence of expectation values in quantum mechanics, [arXiv:1510.01299](#); R. Sayer, A. Maries, and C. Singh, Developing and evaluating a tutorial on the double-slit experiment, [arXiv:1509.07740](#); A. Maries, R. Sayer, and C. Singh, Investigating transfer of learning in advanced quantum mechanics, [arXiv:1509.07570](#); B. Brown and C. Singh, Development and evaluation of a quantum interactive learning tutorial on Larmor Precession of spin, [arXiv:1602.04869](#).
- [24] P. Pintrich, A motivational science perspective on the role of student motivation in learning, and teaching contexts, *J. Educ. Psychol.* **95**, 667 (2003); J. K. Ford, E. M. Smith, D. A. Weissbein, S. M. Gully, and E. Salas, Relationship of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer, *J. Appl. Psych.* **83**, 218 (1998); D. Moos and R. Azevedo, Exploring the fluctuation of motivation and use of self-regulatory processes during learning with hypermedia., *Instr. Sci.* **36**, 203 (2008); H. Song and B. Grabowski, Stimulating intrinsic motivation for problem solving using goal-oriented contexts and peer group composition, *Educ. Technol. Res. Dev.* **54**, 445 (2006); A. Wigfield, L. Hoa, and S. Klauda, The role of achievement values in the regulation of achievement behaviors, in *Motivation and Self-Regulated Learning: Theory, Research and Applications*, edited by D. H. Schunk and B. J. Zimmerman (Lawrence Erlbaum, New York, 2008), pp. 169–195; S. Sungur, Contribution of motivational beliefs and metacognition to students' performance under consequential and nonconsequential test conditions, *Educ. Res. Eval.* **13**, 127 (2007); C. S. Hulleman, A. M. Durik, S. B. Schweigert, and J. M. Harachiewicz, Task values, achievement goals, and interest: An integrative analysis, *J. Educ. Psychol.* **100**, 398 (2008).
- [25] A. Bandura, *Self-efficacy: The Exercise of Control* (W.H. Freeman, New York, 1997); M. Gerhardt and K. Brown, Individual differences in self-efficacy development: The effects of goal orientation and affectivity, *Learning and individual differences* **16**, 43 (2006); P. Hsieh, J. Sullivan, and N. Guerra, A closer look at college students: Self-efficacy and goal orientation, *J. Adv. Academics* **18**, 454 (2007); D. Moos and R. Azevedo, Learning with computer-based learning environments: A literature review of computer self-efficacy, *Rev. Educ. Res.* **79**, 576 (2009); P. H. M. Sins, W. R. van Joolingen, E. R. Savelsbergh, and B. van Hout-Wolters, Motivation and performance within a collaborative computer-based modeling task: Relation between students' achievement goal orientation, self-efficacy, cognitive processing and achievement, *Contemp. Educ. Psychol.* **33**, 58 (2008).
- [26] B. J. Zimmerman and D. H. Schunk, *Self-Regulated Learning and Academic Achievement: Theory, Research and Practice* (Lawrence Erlbaum, Mahwah, NJ, 2001); P. Winne, Self-regulated learning viewed from models of

- information processing, in *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives* 2nd ed. edited by D. H. Schunk and B. J. Zimmerman (Lawrence Erlbaum, Mahwah, NJ, 2001), pp. 153–189; B. J. Zimmerman, Development and adaptation of expertise: The role of self-regulatory processes and beliefs in *The Cambridge Handbook of Expertise and Expert Performance*, edited by K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman (Cambridge University Press, New York, 2006), pp. 705–722; J. A. Greene and R. Azevedo, A macro-level analysis of SRL processes and their relation to the acquisition of a sophisticated mental model of a complex system, *Contemp. Educ. Psychol.* **34**, 18 (2009); J. W. Fryer and A. J. Elliott, Self-regulation of achievement goal pursuit, in *Motivation and Self-Regulated Learning: Theory, Research and Applications*, edited by D. H. Schunk and B. J. Zimmerman (Lawrence Erlbaum, New York, 2008), pp. 53–75.
- [27] B. White and J. Frederiksen, A theoretical framework and approach for fostering metacognitive development, *Educ. Psychol.* **40**, 211 (2005); L. Jiang, J. Elen, and G. Clarebout, The relationships between learner variables, tool-usage behavior and performance, *Comput. Hum. Behav.* **25**, 501 (2009).
- [28] D. Griffiths, *Introduction to Quantum Mechanics* (Pearson Prentice Hall, Englewood Cliffs, NJ, 2004).
- [29] P. Cohen, J. Cohen, L. Aiken, and S. West, The problem of units and the circumstance for POMP, *Multivariate Behav. Res.* **34**, 315 (1999).
- [30] L. Anderson, D. Krathwohl, P. Airasian, K. Cruikshank, R. Mayer, P. Pintrich, J. Raths, and M. Wittrock, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, complete ed. (Longman New York, 2001).
- [31] D. Schwartz and T. Martin, Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction, *Cognit. Instr.* **22**, 129 (2004).
- [32] M. Kapur, Productive failure, *Cognit. Instr.* **26**, 379 (2008); Learning from productive failure, *Learning: Research and practice* **1**, 51 (2015).
- [33] R. Atkinson, S. Derry, A. Renkl, and D. Wortham, Learning from examples: Instructional principles from the worked examples research, *Rev. Educ. Res.* **70**, 181 (2000); E. Yerushalmi, E. Cohen, A. Mason, and C. Singh, What do students do when asked to diagnose their mistakes? Does it help them? I. An atypical quiz context, *Phys. Rev. ST PER* **8**, 020109 (2012); What do students do when asked to diagnose their mistakes? Does it help them? II. A more typical quiz context, *Phys. Rev. ST PER* **8**, 020110 (2012); A. Mason, E. Yerushalmi, E. Cohen, and C. Singh; Learning from mistakes: The effect of students' written self-diagnoses on subsequent problem solving, *Phys. Teach.* **54**, 87 (2016).
- [34] A. Schoenfeld, What's all the fuss about metacognition? in *Cognitive Science and Mathematics Education*, edited by A. Schoenfeld (Lawrence Erlbaum, Hillsdale, NJ, 1987); A. J. Mason and C. Singh, Surveying college introductory physics students' attitudes and approaches to problem solving, *Eur. J. Phys.* **37**, 055704 (2016); A. Mason and C. Singh, Using categorization of problems as an instructional tool to help introductory students learn physics, *Phys. Educ.* **51**, 025009 (2016); Assessing expertise in introductory physics using categorization task, *Phys. Rev. ST Phys. Educ. Res.* **7**, 020110 (2011).
- [35] G. Schraw, Promoting general metacognitive awareness, *Instr. Sci.* **26**, 113 (1998).
- [36] M. Chi, Self-explaining expository texts: The dual processes of generating inferences and repairing mental models, in *Advances in Instructional Psychology*, edited by R. Glaser (Lawrence Erlbaum Associates, Mahwah, NJ, 2000), pp. 161–238.
- [37] C. Dweck, *Mindset: The New Psychology of Success* (Random House, New York, 2006).
- [38] Z. Chen, T. Stelzer, and G. Gladding, Using multi-media modules to better prepare students for introductory physics lecture, *Phys. Rev. ST Phys. Educ. Res.* **6**, 010108 (2010).