REPLICATING EFFECTIVE PEDAGOGICAL APPROACHES FROM INTRODUCTORY PHYSICS TO IMPROVE STUDENT LEARNING OF QUANTUM MECHANICS

by

Ryan Thomas Sayer

B.S., Physics, Brigham Young University, 2010

M.S., Physics, Brigham Young University, 2012

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Ryan Thomas Sayer

It was defended on

July 13th, 2016

and approved by

Dr. Robert Devaty, Associate Professor, Department of Physics and Astronomy

Dr. Arthur Kosowsky, Professor, Department of Physics and Astronomy

Dr. Russell Clark, Senior Lecturer, Department of Physics and Astronomy

Dr. Larry Shuman, Professor, Department of Industrial Engineering

Dissertation Advisor: Dr. Chandralekha Singh, Professor, Department of Physics and Astronomy

REPLICATING EFFECTIVE PEDAGOGICAL APPROACHES FROM INTRODUCTORY PHYSICS TO IMPROVE STUDENT LEARNING OF QUANTUM MECHANICS

Ryan Thomas Sayer, PhD

University of Pittsburgh, 2016

Upper-level undergraduate students entering a quantum mechanics (QM) course are in many ways similar to students entering an introductory physics course. Numerous studies have investigated the difficulties that novices face in introductory physics as well as the pedagogical approaches that are effective in helping them overcome those difficulties. My research focuses on replicating effective approaches and instructional strategies used in introductory physics courses to help advanced students in an upper-level QM course. I have investigated the use of Just-in-time Teaching (JiTT) and peer discussion involving clicker questions in an upper-level quantum mechanics course. The JiTT approach including peer discussions was effective in helping students overcome their difficulties and improve their understanding of QM concepts. Learning tools, such as a Quantum Interactive Learning Tutorial (QuILT) based on the Double-slit Experiment (DSE) which I helped develop, have been successful in helping upper-level undergraduate students improve their understanding of QM. Many students have also demonstrated the ability to transfer knowledge from a QuILT based on the Mach-Zehnder interferometer while working on the DSE QuILT. In addition, I have been involved in implementing research-based activities during our semester-long professional development course for teaching assistants (TAs). In one intervention, TAs were asked to grade student solutions to introductory physics problems first using their choice of method, then again using a rubric designed to promote effective problem-solving approaches, then once more at the end of the semester using their choice of method. This intervention found that many TAs have ingrained beliefs about the purposes of grading which include placing the burden of proof on the instructor as well as a belief that grading cannot serve as a formative assessment. I also compared TAs grading practices and considerations when grading student solutions to QM problems versus when grading student solutions to introductory physics. Many TAs penalized students for not explicating the problem solving process more often in the QM context than in the introductory physics context. The implications of these interventions for promoting student learning in QM are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

First and foremost, I would like to thank my research advisor, Dr. Chandralekha Singh, for her seemingly endless supply of energy, wisdom, and patience towards me. She has devoted countless hours towards helping me achieve my research goals as a graduate student, and I am truly grateful for the privilege of working with her and learning from her these past couple of years.

I want to thank Dr. Alexandru Maries and Dr. Emily Marshman for their help and guidance in completing various research projects during my time as a graduate student. Their insight and dedication inspire me to be a better physics education researcher.

I would also like to thank the National Science Foundation for providing financial support of my research.

Finally, I would like to thank my parents, K. Ryan and Betsey Sayer, for all the love and emotional support they've given me, and for providing a sympathetic ear every time I've called them up to vent when school work was stressing me out.

1.0 INTRODUCTION

The primary goal of physics education research is to help students transition from an initial knowledge state to a desired final knowledge state [1,2]. Students entering an introductory physics course vary greatly in their prior knowledge and preparation. Numerous studies have investigated the difficulties that novices face in introductory physics, and many instructional strategies have been proposed and developed to help introductory physics students improve their problem-solving abilities and develop a robust knowledge structure [3-7]. Upper-level undergraduate students entering a quantum mechanics course are in many ways similar to students in an introductory physics course. For example, they also vary in their prior knowledge, preparation, and self-monitoring skills [8], and many of them do not have well-organized knowledge structures [9]. Advanced undergraduate students must be given support while they repair, extend, and organize their knowledge structure. In this chapter I will summarize some frameworks from cognitive science and prior research in physics education that inform the studies in this thesis involving advanced students in quantum mechanics.

1.1 QUANTUM MECHANICS VERSUS INTRODUCTORY PHYSICS

Richard Feynman once proclaimed that "nobody understands quantum mechanics [10]." Indeed, quantum mechanics (QM) is often a very difficult topic for physics students. Issues regarding interpretation, as well as the abstract formalism and nondeterministic results of QM, may present challenges for beginning students. The phenomena described by QM, such as wave-particle duality and the collapse of the wave function upon measurement in the most popular interpretation of QM, are often counter-intuitive for learners, and even advanced students may struggle with many of the concepts involved, just as beginning students struggle with the concepts in introductory physics. However, unlike classical mechanics, students do not encounter direct everyday observations and experiences with the quantum world to inform their reasoning.

In addition to the counter-intuitive nature of the theory and predictions of QM, the mathematical demands of the subject may present further challenges to learners and increase their overall cognitive load. Students in QM must quickly become fluent with mathematical topics such as linear algebra and differential equations or else they risk becoming overwhelmed by the mathematical demands of QM to the detriment of their conceptual understanding. Prior research shows that mathematical difficulties may hinder conceptual learning in students [11]. Moreover, students who possess alternative conceptions about the conceptual aspects of QM often make mathematical errors in their problem solving [11]. In order to help students develop a well-organized knowledge structure, students' difficulties with the conceptual and mathematical aspects of QM must be addressed.

1.2 RESEARCH IN COGNITIVE SCIENCE

Student difficulties in learning QM may be identified and understood through the tools and research of cognitive science. Cognitive research investigates how people learn and develop expertise in a particular domain as well as how they organize and retrieve their prior knowledge. The findings of cognitive research have implications for physics education research [12].

1.2.1 Memory

According to the information processing view of cognition, human memory is comprised of two different types: long term memory and short term memory. Long term memory is where prior knowledge is stored, and the limit to the amount of information that can be stored in this type of memory appears to be very high [13]. Short term memory, also known as "working memory," is where information is initially processed. Unlike long term memory, working memory is limited for most individuals to around seven "slots" to be used for storing information, though the number may vary between five and nine slots for certain individuals [14,15]. Both long term and working memory are crucial for learning and developing expertise. Individuals must access their prior knowledge from long term memory in order to make connections to the information being processed in their working memory.

1.2.2 Cognitive Load Theory and "Chunking"

Cognitive research also describes how people organize and retrieve their knowledge based on their level of expertise in a given domain. One study asked chess masters to place chess pieces on a board to reproduce a given configuration in a good game of chess, which they would often do by placing certain groups of pieces on the board at the same time [16]. Each group of chess pieces was considered to be a "chunk," and these chunks were usually based on meaningful relations among the pieces. These findings demonstrate that experts can recognize patterns in elements or data and then group the elements together based on those patterns. Experts are usually better than novices at recalling information about a given subject because they have developed many knowledge chunks and have organized those chunks into coherent knowledge structures [2,17]. By chunking elements together, experts reduce the number of slots in their working memory required for those elements. Due to the limited nature of working memory, novices with fewer chunks experience a greater load to their cognitive processing capacity when solving physics problems [18]. Instructional strategies should strive to reduce cognitive load in novices by helping them make connections between different concepts while they are learning, which will lead them to organize their knowledge hierarchically by forming chunks [18].

1.2.3 Knowledge Transfer

Transfer of learning occurs when an individual successfully applies knowledge acquired in one context to a novel situation. Novices often struggle to make connections between problems that appear different in their surface features but that are isomorphic, i.e. that share deeper similarities in their structure and solution method. This difficulty arises because knowledge is encoded and recalled within the context it is first acquired, making it difficult to recall without some form of explicit hint or prompting [19-21]. For example, a student may be told that when a ballerina pulls her arms in close to her body she spins faster due to conservation of angular momentum. However, that same student may not realize that when a neutron star is collapsing it starts to spin faster for the exact same reason. These two physical situations may appear different in their surface features (i.e., a ballerina vs. a star rotating in space), but the situations are structurally isomorphic and can both be understood using the same principle of angular momentum conservation. Novices tend to focus on the surface features of a problem and might not recognize when there is deep similarity between different problems and it is helpful to transfer a solution method from one problem to another. Adaptive experts, by comparison, are able to recognize deeper similarities between problems and thus apply their skills and learning in a variety of contexts and situations.

Conceptual frameworks of transfer, such as the taxonomy proposed by Barnett and Ceci [22] as well as the more recently proposed framework of Nokes-Malach and Mestre [23], suggest that framing and context play a vital role in the success or failure of transfer in a particular situation. According to these frameworks, by recognizing an isomorphism between two different problem spaces, individuals can form analogical mappings and transfer a solution method or reasoning from one context to another, although the success of the mapping depends on the extent to which the two contexts are isomorphic as well as the manner in which the problem is initially framed [24-26]. Instructors in physics should give students sufficient variation in their practice in order to help students recognize isomorphism between problems with different surface features and form new connections, which will help promote future transfer.

1.3 THEORETICAL FRAMEWORKS OF LEARNING FROM COGNITIVE SCIENCE USEFUL IN THIS DISSERTATION

1.3.1 Cognitive Apprenticeship Model

Cognitive researchers have developed different overarching learning frameworks to interpret the findings of their studies. In particular, the model of “Cognitive Apprenticeship” was used when designing and evaluating the studies in this thesis [26]. This model consists of three components:

1) An expert or instructor first models a task for students by carefully demonstrating how the task is performed. 2) The expert then provides coaching and guidance as the students attempt to follow the demonstrated model and learn. 3) Finally, the expert gradually reduces or “fades” the support they give until the students can carry out the task independently. The coaching and scaffolding is important for helping learners to develop expertise but is not always included as part of traditional instruction. In the context of physics, to help students learn well, an instructor may demonstrate how to solve a problem explicitly by highlighting all stages of problem solving (e.g., carrying out a conceptual analysis of the problem, planning the solution, implementing the plan, evaluating the plan and reflecting on the problem solving process to learn from the entire process of solving problems), then give students another problem to solve while the instructor is available to provide scaffolding and coaching. The provided support can be gradually reduced to help students develop self-reliance and be able to solve problems on their own in homework. The students achieve independence as they begin to develop their own knowledge structures and learn useful skills.

While the cognitive apprenticeship model serves as the overarching framework informing the studies in this thesis, three other frameworks have been used to specifically determine how to best help students make the transition from novice to expert while repairing, extending and organizing their knowledge structures. These frameworks are Vygotsky’s “zone of proximal development,” Piaget’s framework involving “optimal mismatch,” and Schwartz and Bransford’s “preparation for future learning.”

1.3.2 Zone of Proximal Development

Vygotsky's theoretical framework of learning [27] involves the concept of the zone of proximal development (ZPD), which is defined as the difference between what a learner can achieve without any support (i.e., their initial knowledge state) and what they can achieve through the guidance of an expert. By keeping instruction within the ZPD of their students, instructors can maximize their learning. In the context of physics, this requires an understanding of students' prior knowledge as well as the ability to design effective instruction that builds on students' prior knowledge. Students in both introductory physics and advanced quantum mechanics may vary greatly in their prior knowledge and skills. With appropriate support provided by an instructor or through interaction with peers, students may develop their knowledge structure of physics and learn useful skills and gradually move from their initial knowledge state to the desired final knowledge state, expanding their ZPD as they learn.

1.3.3 Assimilation, Accommodation, and Optimal Mismatch

Piaget's framework of learning involves the concepts of assimilation, accommodation and optimal mismatch [28]. When new knowledge conforms to the pre-existing knowledge of learners, the knowledge is assimilated in the knowledge structure. If the new knowledge, however, does not conform to the learner's pre-existing knowledge, accommodation is needed in order to incorporate the new knowledge in the knowledge structure. Novices in introductory physics often possess alternative conceptions of physics concepts that run contrary to the accepted ways of reasoning [29-31]. Likewise, students entering a quantum mechanics course may possess certain views of QM which must be accommodated and assimilated for learning to be meaningful. The instructional design should provide optimal mismatch to create a cognitive conflict and then provide guidance and support to help students accommodate and assimilate knowledge. A good instructional design also ensures that the learner does not experience cognitive overload, which might lead to frustration and disengagement from the learning process. Instructional activities which provide "optimal mismatch" can help students to make the transition from novice to expert-like knowledge structures. By carefully choosing instructional tasks to promote conceptual thinking and creating a state of disequilibrium in their students' minds, instructors may facilitate robust learning.

1.3.4 Preparation for Future Learning

As part of their framework known as "preparation for future learning," Schwartz et al. [32] proposed a two-dimensional learning space defined by orthogonal axes of "efficiency" and "innovation" which may be used to determine an optimal learning trajectory. While there are several interpretations of this model, efficiency can be described as "a high degree of consistency that maximizes success and minimizes failure" [32]. A task which is highly efficient may involve rote memorization of some procedure. Individuals who concentrate on efficient tasks eventually become "routine experts" who are good at performing a certain type of tasks but who cannot transfer their knowledge to a different context [33]. Innovation, on the other hand, involves confronting new and unfamiliar situations and solving problems under those situations. For example, in the context of physics, this may involve giving students complex problems which require them to adapt their prior physics knowledge to new situations. Tasks that focus solely on innovation may be too difficult and may lead to frustration in students and can interfere with robust learning. Therefore, both efficiency and innovation are important in helping to prepare students for future learning. Effective instruction should follow a “diagonal direction” in the 2D learning space by incorporating both of these elements. Instructors must balance between the novelty of the material students are learning with the prior knowledge of their students in order to maximize learning.

1.4 RESEARCH IN PHYSICS EDUCATION

1.4.1 Knowledge Structures: Novices and Experts

Physics education researchers have investigated differences between experts’ and novices’ knowledge structures [34-38]. Expertise is a continuous spectrum, with different individuals at different points between novice and expert. In experts, knowledge structures are highly connected and organized hierarchically with the most fundamental principles at the top of the hierarchy (e.g., Newton’s laws, conservation laws, etc.) and less fundamental principles at lower levels. On the other hand, novices’ knowledge structures are comprised of facts and formulas and are only loosely connected. Their learning is often dependent on context, which causes difficulties when students attempt to transfer learning from one context to another. Some upper-level students may fall on the “expert” side of the spectrum regarding their knowledge structures for topics such as quantum mechanics [3]. However, many upper-level undergraduate students do not have the hierarchically-organized knowledge structure of an expert for QM and may have inadequate problem-solving and metacognitive skills [4].

1.4.2 Pedagogical Content Knowledge

Teaching assistants (TAs) play a valuable role in the teaching of introductory students in many universities. They often interact closely with students and grade assignments. In addition to developing their content knowledge, TAs should acquire pedagogical content knowledge, or knowledge of the learning difficulties in their students. The process of developing pedagogical content knowledge is similar to the development of content knowledge. Professional development programs which have been implemented in training physics teachers [39-42] may also be of use to TAs. Effective professional development should build upon TAs' prior knowledge, including their past educational experiences and their beliefs about teaching and learning, which may be highly resistant to change [43,44]. However, limited training and feedback is usually given to new TAs, and many of them rely solely on their experiences in the classroom for learning how to teach [45].

1.5 EFFECTIVE INSTRUCTIONAL STRATEGIES FOR BOTH INTRODUCTORY AND ADVANCED STUDENTS

Physics education researchers have developed various instructional strategies to assist both introductory and advanced students in learning physics. These strategies make use of the learning frameworks mentioned above in order to reduce students' cognitive load and assist them in "chunking" information and developing a hierarchically-organized knowledge structure and in learning useful skills. These instructional strategies often incorporate modeling and coaching, and gradually "fade" support provided to students, allowing them to function effectively on their own. Research-based learning tools such as tutorials, Just-in-Time Teaching, and peer-instruction are effective scaffolding tools for introductory students [46-50], and prior research has shown that these learning tool are also effective in upper-level courses such as quantum mechanics [51-61]. They build on students' prior knowledge and explicitly address common difficulties students have in physics. These learning tools give students an opportunity to assimilate and accommodate new ideas while extending and organizing their knowledge structure.

In quantum mechanics, Quantum Interactive Learning Tutorials (QuILTs) use a guided approach to learning in which students predict what should happen in a particular situation and then are provided appropriate feedback. This feedback often involves visualization tools that students can use to check their predictions [51-54]. Each QuILT typically contains groups of questions that build on each other that students work on related to a certain topic. At the end of each group of questions, necessary feedback is provided to students, either through computer simulations or illustrations, or sometimes (at the discretion of the instructor) a general class discussion of the relevant issues. QuILTs may be used either in class or as a homework supplement or self-study tool by students, providing coaching and scaffolding support to the students.

As stated in Mazur's manual of peer instruction, the primary goal of implementing a peer instruction strategy in class is "to exploit student interaction during lectures and focus students' attention on underlying concepts" [50]. For a class using a traditional lecture format, students usually have little interaction with the instructor and their classmates during class, and they are often too busy taking notes to ask the instructor questions or identify any of their struggles. For a class using the peer instruction method, in-class time may be divided into several short presentations focusing on concepts [62], after which the students may be given multiple-choice questions designed to highlight common student difficulties. They may then discuss their answers to these questions with a partner and then respond to the questions, e.g., using electronic clickers [63]. These discussions often lead to co-construction of knowledge, which occurs when neither student working in a pair was able to answer a question before collaborating with each other, but both students were able to answer the question after their collaboration. Peer discussion may be incorporated into quantum mechanics courses to help advanced students develop their knowledge structures [56].

1.6 REPLICATING EFFECTIVE PEDAGOGICAL APPROACHES FROM INTRODUCTORY PHYSICS TO IMPROVE STUDENT LEARNING OF QUANTUM MECHANICS

In this thesis, I present the findings from investigations involving the use of learning strategies and approaches to improve student learning of quantum mechanics that are often found effective in the context of introductory physics. The first study investigates the use of Just-in-Time teaching and peer instruction in a quantum mechanics course. These instructional strategies are designed to help students develop useful skills and repair, organize, and extend their knowledge structure. The results of this study indicate that while the activities taken together were effective in helping most students learn, the students displayed different levels of learning in response to the different learning activities.

The second study investigates the development and evaluation of a QuILT on the double-slit experiment (DSE) involving single particles sent one at a time to the slits and a lamp whose photons scatter off the incident single particles at the slits. Before the study, there was a preliminary investigation of student difficulties with the de Broglie relation and the interference of macroscopic particles such as sand. In a pre-test and post-test, students were asked the following question:

"You are conducting a double-slit experiment in which you send a large number of non-relativistic electrons of the same kinetic energy one at a time towards a double-slit plate. The wavelength of the electrons is 9 pm, the slit width is 50 pm, the slit separation is 1 nm and the distance between the slits and the screen is 3 m. Suppose the experiment is modified by using protons instead of electrons while all of the following parameters are held fixed: kinetic energy, slit width/separation, and distance from slits to screen. How does the pattern change, if at all?"

The responses of undergraduate and graduate students to this question are shown in Table 1-I. The correct response is that the distance between interference fringes will become narrower since the de Broglie wavelength of protons is shorter than the wavelength of electrons if the particles have the same kinetic energy. Many students incorrectly assumed that if two particles have the same kinetic energy then they must have the same wavelength, so the distance between fringes would not change. Some students who responded that the distance between fringes will become wider claimed that the wavelength increases with mass, when in fact the wavelength will be shorter for more massive particles if the particles have the same kinetic energy.

Table 1-I Undergraduate (US) ($N = 46$) and graduate student (GS) ($N = 45$) responses to question 1 on the pre-test and post-test.

	Fringes Narrower	Fringes Wider	"Change"	Will Not Change	Interference	No Interference	Other/No Response
US Pre	34%	14%	2%	25%	5%	14%	7%
US Post	36%	2%	5%	14%	7%	11%	25%
GS Pre	25%	14%	11%	25%	5%	18%	2%
GS Post	36%	4%	11%	20%	7%	13%	9%

Students were also asked the following question regarding interference of macroscopic particles: “Consider particles of sand, which can be approximated as spheres of a radius of about 1/10 of a millimeter. Do you expect that a double slit experiment with well-chosen parameters would show an interference pattern? Explain your reasoning.”

For this question, student reasoning was the main criterion used to determine if a student’s response was deemed correct. For example, students who noted that they do NOT expect to observe an interference pattern regardless of the experimental parameters were counted as correct if they noted that the experimental parameters for this case are unrealistic. These students generally calculated the de Broglie wavelength for the sand particle and noted that interference is not possible because the wavelength of the sand particles is orders of magnitude smaller than any physical setup for double slit that can be constructed realistically to observe interference (e.g., how would the sand particles pass through a physical slit which is of the order of its de Broglie wavelength and the distance between the slits is also of the same order). These students understood that all particles have an associated wavelength and focused on the fact that the parameters for a double slit experiment with sand particles with very small wavelength could not be achieved in a realistic situation and the large sand particles would bump into the narrow slits of the order of a de Broglie wavelength even if such a slit could be constructed.

Furthermore, since an equivalent experiment to that of a DSE with sand particles would require that the distance between the slits and slit widths be comparable to the de Broglie wavelength of the sand particles which is very small, one needs to consider how to physically design an experiment that is equivalent to the DSE without using physical slits since the sand particles are large. Although we would have counted student responses to be correct if they had noted that such an experiment would be possible or at least could be conceived if one could achieve the conditions for a double slit experiment using some clever technique for sand particles with such small de Broglie wavelength without explicitly using physical slits (although such an experiment is not envisioned any time in the near future), no student provided this type of response.

The most common incorrect response was not focusing on the de Broglie wavelength of the sand particles and only focusing on their size. For example, some students who provided the incorrect response claimed that one needs slits with a size larger than that of the sand particles so that sand particles can pass through the slits to observe interference. The average scores and standard deviations of undergraduate and graduate students for question 2 about the sand particle are shown in Table 1-II, with p-values for comparison between the undergraduate and graduate students and also for the comparison of the pre-test and post-test scores for the undergraduates and graduate students separately. While the means of the pre-test scores of the undergraduates and graduate students are not statistically significantly different, the means of the post-test scores are statistically significantly different ($p = 0.031$).

Table 1-II Undergraduate (US) ($N = 46$) and graduate student (GS) ($N = 45$) averages and standard deviations (Std. Dev.) for question 2 on the pre-test and post-test, with p-values for comparison between pre-test and post-test scores of US and GS.

	Pre-test		Post-test		
	Average	Std. Dev.	Average	Std. Dev.	p
US	45%	50%	91%	29%	< 0.001
GS	41%	50%	73%	45%	0.002
p	0.671		0.031		

Overall, the QuILT on the DSE involving single particles was effective in helping undergraduate and graduate students learn these concepts, though the benefits were greater for the undergraduate students. (The difference between the means of the pre-test and post-test was statistically significant, with $p < 0.001$.) One reason for this difference may be the grade incentive provided to the undergraduate students.

The third study investigates the transfer of learning between a QuILT based on the DSE involving single particles and another QuILT based on the Mach-Zehnder Interferometer (MZI). The situations described in these two different QuILTs are isomorphic, which may facilitate students to transfer their reasoning from the MZI QuILT to successfully answer pre-test questions on the DSE QuILT (if students engaged with the MZI QuILT before DSE). The findings of this study indicate that advanced students were able to transfer their learning and reasoning involving "which-path" information from the MZI context to the context of the DSE.

The fourth study investigates learning activities in a TA training course which involve the use of a rubric designed to promote good problem-solving strategies in students whose work is being graded. This study shows that many TAs possess ingrained ideas about the purposes of grading. The fifth study, which also involves learning activities in a TA training course, investigates the grading beliefs and practices of TAs when grading introductory physics problems as compared to their beliefs and practices when grading problems from an advanced quantum mechanics course. The findings of this study indicate that many TAs believe that different grading criteria should be applied while grading introductory physics and quantum mechanics, though these differences vary to some extent. Leaders of professional development workshops can use these findings to better understand TAs' prior knowledge and beliefs as they help them develop pedagogical content knowledge.

1.7 CHAPTER REFERENCES

1. F. Reif, Systematic problem solving, in *Applying Cognitive Science to Education: Thinking and Learning in Scientific and Other Complex Domains* (MIT Press, Cambridge, MA, 2008), p 201.

2. F. Reif, Millikan Lecture 1994: Understanding and teaching important scientific thought processes, Am. J. Phys. **63**, 17 (1995).

3. W. Leonard, R. Dufresne, and J. Mestre, Using qualitative problem-solving strategies to highlight the role of conceptual knowledge in solving problems, Am. J. Phys. **64**, 1495 (1996).

4. J. Heller and F. Reif, Prescribing effective human problem solving processes: Problem description in physics. Cognition and Instruction **1**, 177 (1984).

5. A. Van Heuvelen, Overview, case study physics, Am. J. Phys. **59**, 898 (1991).

6. J. Mestre, R. Dufresne, W. Gerace, P. Hardiman, and J. Touger, Promoting skilled problem solving behavior among beginning physics students, Journal of Research in Science Teaching **30**, 303 (1993).

7. R. Dufresne, W. Gerace, P. Hardiman, and J. Mestre, Constraining novices to perform expert-like problem analyses: Effects on schema acquisition, Journal of the Learning Sciences **2**, 307 (1992).

8. A. Mason and C. Singh, Do advanced physics students learn from their mistakes without explicit intervention?, Am. J. Phys. **78**, 760 (2010); A. Mason and C. Singh, Reflection and self-monitoring in quantum mechanics, *Proceedings of the 2009 Phys. Ed. Res. Conference, Ann Arbor, MI*, edited by M. Sabella, C. Henderson, and C. Singh (AIP Conf. Proc., Melville, NY, 2009), p. 197; B. Brown, A. Mason, and C. Singh, Improving performance in quantum mechanics with explicit incentives to correct mistakes, Phys. Rev. PER **12**, 010121 (2016).

9. S. Lin and C. Singh, Categorization of quantum mechanics problems by professors and students, Eur. J. Phys. **31**, 57 (2010); S. Lin and C. Singh, Assessing expertise in quantum mechanics using categorization task, *Proceedings of the 2009 Phys. Ed. Res. Conference, Ann Arbor, MI*, edited by M. Sabella, C. Henderson, and C. Singh (AIP Conf. Proc., Melville, NY, 2009), p. 185.

10. R. Feynman, *The Character of Physical Law*, Chapter 6 (British Broadcasting Corp., London, 1965).

11. C. Singh, Student difficulties with quantum mechanics formalism, *Proceedings of the 2007 Phys. Ed. Res. Conference*, edited by L. Hsu, C. Henderson, and L. McCullough (AIP Conf. Proc., Melville, NY, 2007), p. 185.

12. E. Redish, The implications of cognitive studies for teaching physics, Am. J. Phys. **69**, 796 (1994).

13. J. Anderson, *Learning and Memory* (Wiley, New York, NY, 1995).

14. H. Simon, How big is a memory chunk?, Science **183**, 482 (1974).

15. G. Miller, The magical number seven, plus or minus two: Some limits on our capacity for processing information, Psychol. Rev. **63**, 81 (1956).

16. W. Chase and H. Simon, Perception in chess, Cog. Psych. **4**, 55 (1973).

17. M. Chi, Laboratory methods for assessing experts' and novices' knowledge, in *The Cambridge Handbook of Expertise and Expert Performance*, edited by K. Ericsson et al. (Cambridge University Press, London, 2006), pp. 167-184.

18. J. Sweller, Cognitive load during problem solving: effects on learning, Cog. Sci. **12**, 257 (1988).

19. M. Gick and K. Holyoak, The cognitive basis of knowledge transfer, in *Transfer of Learning: Contemporary Research and Applications*, edited by M. Cornier and J. Hagman (Academic Press, New York, NY, 1987), pp. 9-42.

20. M. Gick and K. Holyoak, Schema induction and analogical transfer, Cog. Psych. **15**, 1 (1983).

21. K. Holyoak, The pragmatics of analogical transfer, in *The Psychology of Learning and Motivation,* edited by G. Bower (Academic Press, New York, NY, 1985), pp. 59-87.

22. S. Barnett and S. Ceci, When and where do we apply what we learn? A taxonomy for far transfer, Psychological Bulletin **128**, 612 (2002).

23. T. Nokes-Malach and J. Mestre, Toward a model of transfer as sense-making, Educ. Psychol. **43**, 184 (2013).

24. H. Simon and J. Hayes, The understanding process: Problem isomorphs, Cog. Psych. **8**, 165 (1976).

25. J. Hayes and H. Simon, Psychological differences among problem isomorphs, in *Cognitive Theory*, edited by N. Castellan, D. Pisoni, and G. Potts (Lawrence Erlbaum, Hillsdale, NJ, 1977), pp. 21-41.

26. A. Collins, J. Brown, and S. Newman, Cognitive apprenticeship: Teaching the crafts of reading, writing and apprenticeship, in *Knowing, Learning and Instruction: Essays in Honor of Robert Glaser*, edited by R. Glaser and L. Resnick (Lawrence Erlbaum Associates, Hillsdale, NJ, 1989), pp. 453-494.

27. L. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes* (Harvard University Press, Cambridge, MA, 1978).

28. H. Ginsberg and S. Opper, *Piaget's Theory of Intellectual Development* (Prentice Hall, Englewood Cliffs, NJ, 1969).

29. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66**, 64 (1998).

30. N. Lasry, E. Mazur, and J. Watkins, Peer instruction: From Harvard to the two-year college, Am. J. Phys. **76**, 1066 (2008).

31. L. McDermott, M. Rosenquist, and E. van Zee, Student difficulties in connecting graphs and physics: Examples from kinematics, Am. J. Phys. **55**, 503 (1987).

32. D. Schwartz, J. Bransford, and D. Sears, Efficiency and innovation in transfer, in *Transfer of Learning: Research and Perspectives*, edited by J. Mestre (Information Age Publishing, Greenwich, CT, 2005), pp. 1-52.

33. G. Hatano and Y. Oura, Commentary: Reconceptualizing school learning using insight from expertise research, Educ. Res. **32**, 26 (2003).

34. K. Kotovsky, J. Hayes, and H. Simon, Why are some problems hard? Evidence from the Tower of Hanoi, Cog. Psych. **17**, 284 (1985).

35. B. Eylon and F. Reif, Effects of knowledge organization on task performance, Cog. Instruct. **1**, 5 (1984).

36. M. Chi, P. Feltovich, and R. Glaser, Categorization and representation of physics knowledge by experts and novices, Cog. Sci. **5**, 121 (1981).

37. A. Schoenfeld and D. Herrmann, Problem perception and knowledge structure in expert novice mathematical problem solvers, J. Exp. Psych.: Learning, Memory and Cognition **8**, 484 (1982).

38. A. Van Heuvelen, Learning to think like a physicist: A review of research-based instructional strategies, Am. J. Phys. **59**, 891 (1991).

39. L. Shulman, Those who understand: Knowledge growth in teaching, Educ. Res. **15**, 4 (1986).

40. L. McDermott, Millikan Lecture 1990: What we teach and what is learned—Closing the gap, Am. J. Phys. **59**, 301 (1991).

41. H. Borko, Professional development and teacher learning: Mapping the terrain, Educ. Res. **33**, 3 (2004).

42. B. Eylon and E. Bagno, Research-design model for professional development of teachers: Designing lessons with physics education research, Phys. Rev. ST PER **2**, 020106 (2006).

43. S. Lin, C. Henderson, W. Mamudi, C. Singh, and E. Yerushalmi, Teaching assistants' beliefs regarding example solutions in introductory physics, Phys. Rev. ST PER **9**, 010120 (2013); C. Singh, Rethinking tools for training teaching assistants, *Proceedings of the 2009 Phys. Ed. Res. Conference, Ann Arbor, MI*, edited by M. Sabella, C. Henderson, and C. Singh (AIP Conf. Proc., Melville, NY, 2009), p. 59; E. Yerushalmi, C. Henderson, W. Mamudi, C. Singh, and S. Lin, The group administered interactive questionnaire: An alternative to individual interviews, *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE*, edited by S. Rebello, C. Singh, and P. Engelhardt (AIP Conf. Proc., Melville, NY, 2012), p. 97; S. Lin, C. Singh, W. Mamudi, C. Henderson, and E. Yerushalmi, TA-designed vs. research-oriented problem solutions, *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE*, edited by S. Rebello, C. Singh, and P. Engelhardt (AIP Conf. Proc., Melville, NY, 2012), p. 255; E. Yerushalmi, E. Marshman, A. Maries, C. Henderson, and C. Singh, Grading practices and considerations of graduate students at the beginning of their teaching assignment, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN*, edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2015), p. 287; C. Henderson, E. Marshman, A. Maries, E. Yerushalmi, and C. Singh, Instructional goals and grading practices of graduate students after one semester of teaching experience, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN*, edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2015), p. 111.

44. J. Luft, J. Kurdziel, G. Roehrig, and J. Turner, Growing a garden without water: Graduate teaching assistants in introductory science laboratories at a doctoral/research university, J. Res. in Sci. Teach. **41**, 211 (2004).

45. A. Thompson and J. Zeuli, The frame and the tapestry: Standards-based reform and professional development, in *Teaching as the Learning Profession: Handbook of Policy and Practice*, edited by L. Darling-Hammond and G. Sykes (Jossey-Bass, San Francisco, CA, 1999), p. 341.

46. A. Mason and C. Singh, Assessing expertise in introductory physics using categorization task, Phys. Rev. ST PER **7**, 020110 (2011); C. Singh, Categorization of problems to assess and improve proficiency as teacher and learner, Am. J. Phys. **77**, 73 (2009); A. Mason and C. Singh, Revisiting categorization, in *Proceedings of the National Association of Research in Science Teaching (NARST) 2009 Annual Meeting* (Garden Grove, CA, 2009); A. Mason and

C. Singh, Categorization of mechanics problems by students in large introductory physics courses: A comparison with the Chi, Feltovich, and Glaser study, *Proceedings of the 2013 Phys. Ed. Res. Conference, Portland, OR*, edited by P. Engelhardt, A. Churukian, D. Jones (AIP Conf. Proc., Melville, NY, 2014), p. 35.

47. L. McDermott and P. Schaffer, *Tutorials in Introductory Physics* (Prentice Hall, Upper Saddle River, NJ, 1998).

48. G. Novak, E. Patterson, A. Gavrin, and W. Christian, *Just-in-Time-Teaching: Blending Active Learning with Web Technology* (Prentice Hall, Upper Saddle River, NJ, 1999).

49. 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).

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

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

52. G. Zhu and C. Singh, Improving students' understanding of quantum measurement, *Proceedings of the 2009 Phys. Ed. Res. Conference, Portland, OR,* edited by C. Singh, M. Sabella, and S. Rebello (AIP Conf. Proc., Melville, NY, 2010), p. 345; G. Zhu and C. Singh, Improving students' understanding of quantum measurement I: Investigation of difficulties, Phys. Rev. ST PER **8**, 010117 (2012); G. Zhu and C. Singh, Improving students' understanding of quantum measurement II. Development of research-based learning tools, Phys. Rev. ST PER **8**, 010118 (2012); G. Zhu and C. Singh, Students' difficulties with quantum measurement, *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE,* edited by S. Rebello, C. Singh, and P. Engelhardt (AIP Conf. Proc., Melville, NY, 2012), p. 387; C. Singh and G. Zhu, Surveying students' understanding of quantum mechanics, *Proceedings of the 2009 Phys. Ed. Res. Conference, Portland, OR*, edited by C. Singh, M. Sabella, and S. Rebello (AIP Conf. Proc., Melville, NY, 2010), p. 301; C. Singh and G. Zhu, Students' understanding of the addition of angular momentum, *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE,* edited by S. Rebello, C. Singh, and P. Engelhardt (AIP Conf. Proc., Melville, NY, 2012), p. 355; C. Singh and E. Marshman, Investigating student difficulties with Dirac notation, *Proceedings of the 2013 Phys. Ed. Res. Conference, Portland, OR*, edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2014), p. 345; E. Marshman and C. Singh, Investigating student difficulties with time-dependence of expectation values in quantum mechanics, *Proceedings of the 2013 Phys. Ed. Res. Conference, Portland, OR,* edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2014), p. 245; C. Singh and E. Marshman, Analogous patterns of student reasoning difficulties in introductory physics and upper-level quantum mechanics, *Proceedings of the 2013 Phys. Ed. Res. Conference, Portland, OR,* edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2014), p. 46; S. DeVore and C. Singh, Development of an interactive tutorial on quantum key distribution, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN,* edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2015), p. 59; B. Brown and C. Singh, Development and evaluation of a quantum interactive learning tutorial

on Larmor Precession of spin, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN*, edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2015), p. 47; E. Marshman and C. Singh, Developing an interactive tutorial on a quantum eraser, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN*, edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2015), p. 175; E. Marshman and C. Singh, Student difficulties with quantum states while translating state vectors in Dirac notation to wave functions in position and momentum representations, *Proceedings of the 2015 Phys. Ed. Res. Conference, College Park, MD*, edited by A. Churukian, D. Jones, and L. Ding (AIP Conf. Proc., Melville, NY, 2015), p. 210; A. Maries, R. Sayer, and C. Singh, Investigating transfer of learning in advanced quantum mechanics, *Proceedings of the 2015 Phys. Ed. Res. Conference, College Park, MD*, edited by A. Churukian, D. Jones, and L. Ding (AIP Conf. Proc., Melville, NY, 2015), p. 207; R. Sayer, A. Maries, and C. Singh, Developing and evaluating a tutorial on the double-slit experiment, *Proceedings of the 2015 Phys. Ed. Res. Conference, College Park, MD*, edited by A. Churukian, D. Jones, and L. Ding (AIP Conf. Proc., Melville, NY, 2015), p. 299; B. Brown, C. Singh, and A. Mason, The effect of giving explicit incentives to correct mistakes on subsequent problem solving in quantum mechanics, *Proceedings of the 2015 Phys. Ed. Res. Conference, College Park, MD*, edited by A. Churukian, D. Jones, and L. Ding (AIP Conf. Proc., Melville, NY, 2015), p. 67; E. Marshman and C. Singh, Interactive tutorial to improve student understanding of single photon experiments involving a Mach-Zehnder Interferometer, Eur. J. Phys. **37**, 024001 (2016); C. Singh and E. Marshman, Developing an interactive tutorial on a Mach-Zehnder interferometer with single photons, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN*, edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2015), p. 239.

53. 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, Physics Today **8**, 43 (2006); C. Singh, Assessing and improving student understanding of quantum mechanics, *Proceedings of the 2005 Phys. Ed. Res. Conference, Salt Lake City, UT*, edited by P. Heron, J. Marx, and L. McCullough (AIP Conf. Proc., Melville, NY, 2006), p. 69; C. Singh, Helping students learn quantum mechanics for quantum computing, *Proceedings of the 2006 Phys. Ed. Res. Conference, Syracuse, NY*, edited by L. McCullough, P. Heron, and L. Hsu (AIP Conf. Proc., Melville, NY, 2007), p. 42; E. Marshman and C. Singh, Framework for understanding the patterns of student difficulties in quantum mechanics, Phys. Rev. ST PER **11**, 020119 (2015); C. Singh and E. Marshman, Review of student difficulties in upper-level quantum mechanics, Phys. Rev. ST PER **11**, 020117 (2015); C. Singh, Student understanding of quantum mechanics at the beginning of graduate instruction, Am. J. Phys. **76**, 277 (2008); C. Singh and G. Zhu, Cognitive issues in learning advanced physics: An example from quantum mechanics, *Proceedings of the 2008 Phys. Ed. Res. Conference, Ann Arbor, MI*, edited by C. Henderson, M. Sabella, and C. Singh (AIP Conf. Proc., Melville, NY, 2009), p. 63; C. Singh, Transfer of learning in quantum mechanics, *Proceedings of the 2004 Phys. Ed. Res. Conference, Sacramento, CA* edited by P. Heron, S. Franklin, and J. Marx (AIP Conf. Proc., Melville, NY, 2005), p. 23; S. Siddiqui and C. Singh, Surveying instructors' attitudes and approaches to teaching quantum mechanics, *Proceedings of the 2009 Phys. Ed. Res. Conference, Portland, OR*, edited by C. Singh, M. Sabella, and S. Rebello (AIP Conf. Proc., Melville,

NY, 2010), p. 297; G. Zhu and C. Singh, Surveying students' understanding of quantum mechanics in one spatial dimension, Am. J. Phys. **80**, 252 (2012).

54. G. Zhu and C. Singh, Improving students’ understanding of quantum mechanics via the Stern-Gerlach experiment, Am. J. Phys. **79**, 499 (2011); G. Zhu and C. Singh, Students' understanding of Stern-Gerlach experiment, *Proceedings of the 2008 Phys. Ed. Res. Conference, Ann Arbor, MI*, edited by M. Sabella, C. Henderson, and C. Singh (AIP Conf. Proc., Melville, NY, 2009), p. 309.

55. G. Zhu and C. Singh, Improving students’ understanding of the addition of angular momentum in quantum mechanics, Phys. Rev. ST PER **9**, 010101 (2013).

56. C. Singh and G. Zhu, Improving students' understanding of quantum mechanics by using peer instruction tools, *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE*, edited by S. Rebello, C. Singh, and P. Engelhardt (AIP Conf. Proc., Melville, NY, 2012), p. 77.

57. M. Belloni, W. Christian, and A. Cox, *Physlet Quantum Physics: An Interactive Introduction* (Pearson Prentice Hall, Upper Saddle River, NJ, 2006).

58. J. Hiller, I. Johnston, and D. Styer, *Quantum Mechanics Simulations: The Consortium for Upper-level Physics Software* (Wiley, New York, NY, 1995).

59. P. Jolly, D. Zollman, N. Rebello, and A. Dimitrova, Visualizing motion in potential wells, Am. J. Phys. **66**, 57 (1998).

60. S. McKagan, K. Perkins, M. Dubson, C. Malley, S. Reid, R. LeMaster, and C. Wieman, Developing and researching PhET simulations for teaching quantum mechanics, Am. J. Phys. **76**, 406 (2008).

61. G. Zhu and C. Singh, Peer instruction for quantum mechanics, APS Forum on Education Newsletter, pp 8-10 (Fall 2009).

62. C. Crouch and E. Mazur, Peer Instruction: Ten years of experience and results, Am. J. Phys. **69**, 970 (2001).

63. I. Beatty, W. Gerace, W. Leonard, and R. Dufresne, Designing effective questions for classroom response system teaching, Am. J. Phys. **74**(1), 31 (2006).

2.0 A CASE STUDY EVALUATING JUST-IN-TIME TEACHING AND PEER INSTRUCTION USING CLICKERS IN A QUANTUM MECHANICS COURSE

2.1 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 pre-lecture 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 pre-lecture 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 pre-lecture 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: “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 non-intuitive 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 pre-lecture 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.

2.2 MOTIVATION, RESEARCH QUESTIONS, AND APPROACH

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 which 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 which 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 meta-cognitive 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 pre-lecture 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 pre-lecture 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.

2.3 METHODOLOGY

2.3.1 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 pre-lecture 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 pre-lecture 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 pre-lecture 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 pre-lecture RQs was used to answer Research Question 1.

After lecture, which focused on student difficulties identified in the pre-lecture 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 through 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 pre-lecture 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. 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 (please see Appendix A).

2.3.2 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 (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:

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

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 he/she did not select either option A or D, his/her "individual %" will be 0% (without POMP). The student's corresponding individual POMP score will then be $\frac{0\%-40\%}{100\%-40\%} \times 100\% = -66.6\%$. If the student had instead chosen option A or D, his/her individual POMP score would be 100% (same as his/her score without POMP). For each student, the individual POMP scores for all 18 comparison questions were averaged together (i.e., the sum of his/her 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 questions 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 OQ response for each student based upon the equivalent elements of the corresponding multiple-choice question (please see the notes in 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.

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 multi-electron 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 $T = 0K$.

VII. Total energy associated with valence electrons in copper cubes of different sizes at temperature $T = 0K$.

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. Non-interacting distinguishable particles in a one-dimensional infinite square well.

X. Non-interacting bosons in a one-dimensional infinite square well.

XI. Three non-interacting fermions in four single particle states.

XII. Is the perturbing Hamiltonian matrix $\widehat{H}'$ diagonal in the basis in which the unperturbed Hamiltonian matrix $\widehat{H}^{\mathrm{o}}$ is diagonal?

XIII. Given that the perturbing Hamiltonian $\widehat{H}'$ and the unperturbed Hamiltonian $\widehat{H}^o$ both commute with some Hermitian operator $\hat{A}$, do they necessarily commute with each other?

XIV. Is an eigenstate $|a\rangle$ of $\widehat{H}^o$ corresponding to a degenerate subspace of $\widehat{H}^o$ necessarily a "good" state for a given perturbing Hamiltonian $\widehat{H}'$?

XV. Is an eigenstate $|c\rangle$ corresponding to a non-degenerate subspace of $\widehat{H}^o$ necessarily a "good" state for a given perturbing Hamiltonian $\widehat{H}'$?

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 $\widehat{H}' = \alpha \hat{L}_z$?

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:

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

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.

2.4 RESULTS

2.4.1 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 2-I. Overall, there is an upward trend from RQ to ICT and from ICT to GCT. In Table 2-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 2-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 pre-lecture reading assignment, which is at the level of guessing.

Table 2-I Average and median student scores (averaged over all students and all comparison questions) and standard deviations (Std. Dev.) 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 scores are given as decimals (out of 1) as a reminder that the OQ is in a different format.

	RQ	ICT	GCT	OQ
Average	20%	48%	73%	0.78
Median	21%	51%	75%	0.74
Std. Dev.	23%	33%	21%	0.13
p-value		RQ→ICT 0.004	ICT→GCT 0.009	

A comparison of the individual students' average performance on the RQ versus the ICT for the comparison questions is shown in Fig. 2-1. The symbols labeled A through 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 pre-lecture 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 ICT performance versus 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 $p = 0.004$.

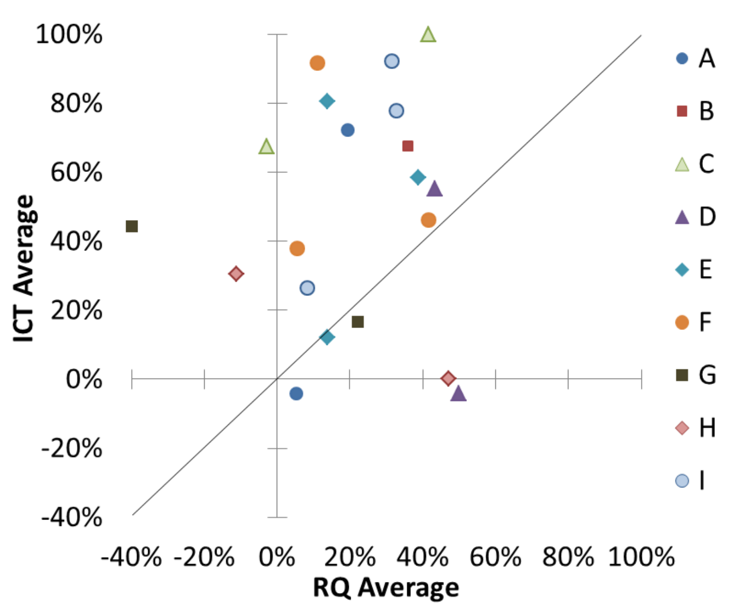


Figure 2-1 Student performance on ICT versus RQ, averaged over all comparison questions. The difference between the means of the RQ and ICT scores is significant ($p = 0.004$). The average standard error was ±6% for the RQ and ±8% for the ICT.

Figure 2-2 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. 2-2 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 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. 2-2 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).

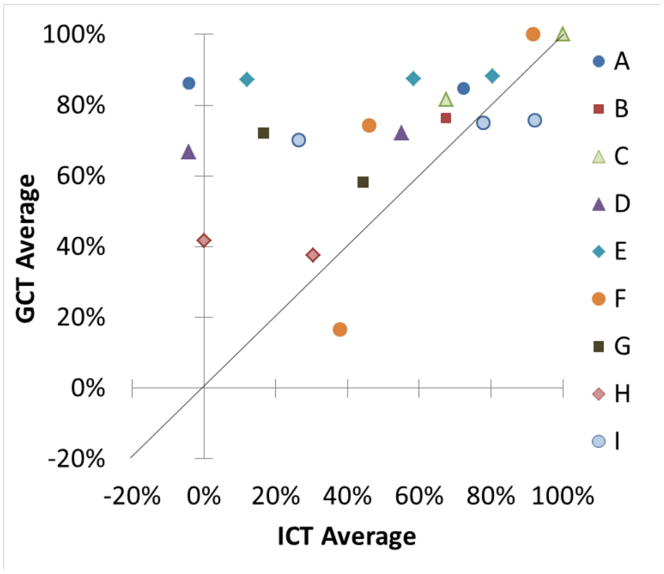


Figure 2-2 Student performance on GCT versus ICT, averaged across all comparison questions. The difference between the means of the ICT and GCT scores is significant ($p = 0.009$). The average standard error was ±8% for the ICT and ±5% for the GCT.

A comparison of students' average GCT and OQ scores for all comparison questions is shown in Fig. 2-3. 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, 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.

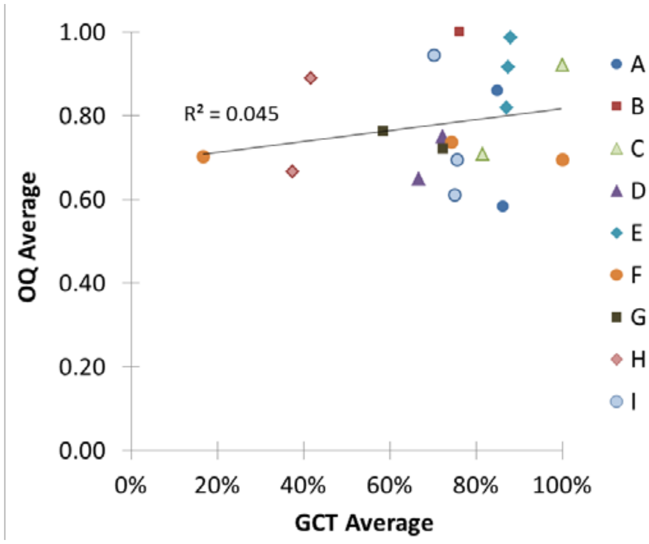


Figure 2-3 Student performance on OQ versus GCT, averaged across all comparison questions, with linear regression and corresponding Pearson correlation coefficient ($R^2 = 0.045$). The average standard error was ±5% for the GCT and ±0.03 for the OQ.

Figure 2-4 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 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. 2-1 through 2-4 suggest that there was no single learning activity that led to maximum learning gains for all students.

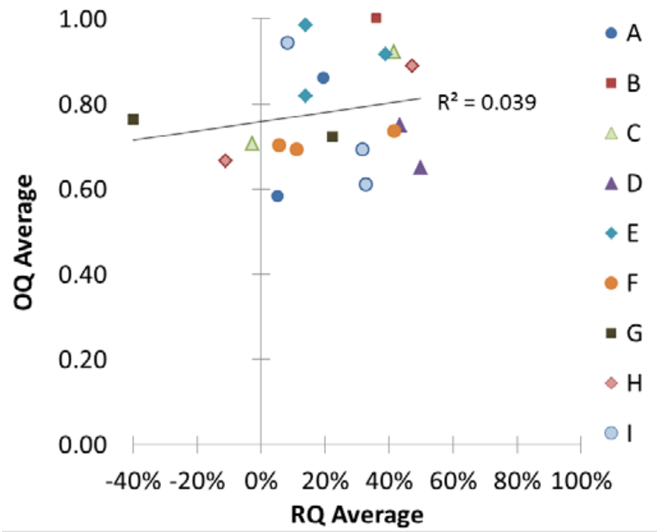


Figure 2-4 Student performance on OQ versus RQ, averaged across all comparison questions, with linear regression and corresponding Pearson correlation coefficient ($R^2 = 0.039$). The average standard error was ±6% for the RQ and ±0.03 for the OQ.

2.4.2 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 2-5 shows the average ICT vs. RQ scores for all comparison questions listed in section III (B). Each data point represents the average POMP score on a particular question. Figure 2-5 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 2-5 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 versus 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: Questions III, IV, and V do not appear in Fig. 2-5 because there was no RQ for those questions.)

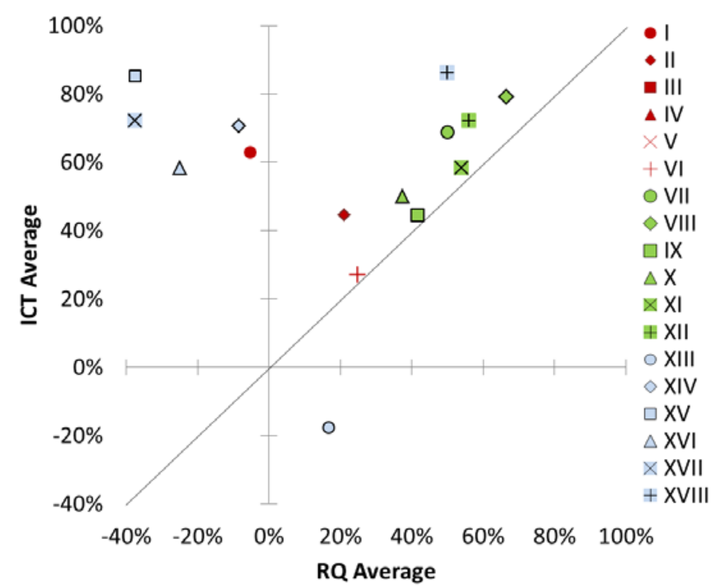


Figure 2-5 Average scores on the comparison questions for the ICT versus the RQ, averaged over all students. The average standard error was ±8% for the RQ and ±7% for the ICT.

Figure 2-6 compares the average performances for each comparison question on the GCT versus 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 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: Questions XII, XVI, XVII, and XVIII do not appear in Fig. 2-6 because there was no GCT for those questions.)

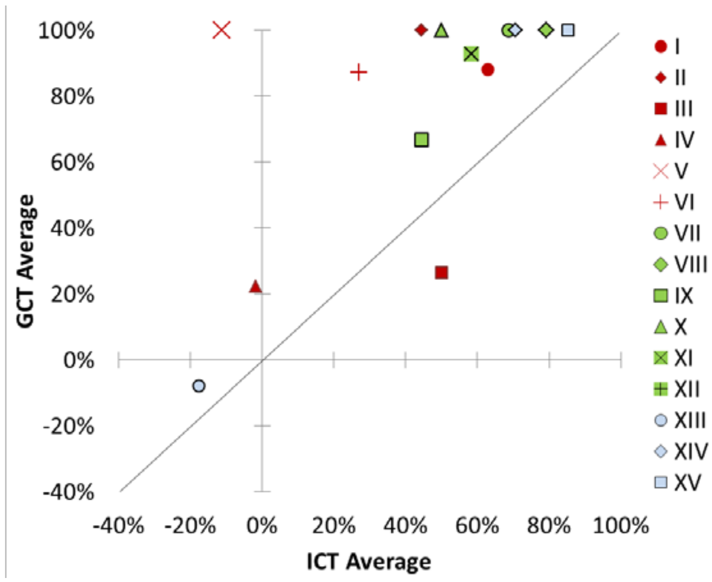


Figure 2-6 Average scores on the comparison questions for the GCT versus the ICT. The average standard error was ±7% for the ICT and ±8% for the GCT.

Figure 2-7 shows 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 questions on the OQ regardless of how well they performed on the GCT for the same topic. (Note: Questions XII, XVI, XVII, and XVIII do not appear in Fig. 2-7 because there was no GCT for those questions.)

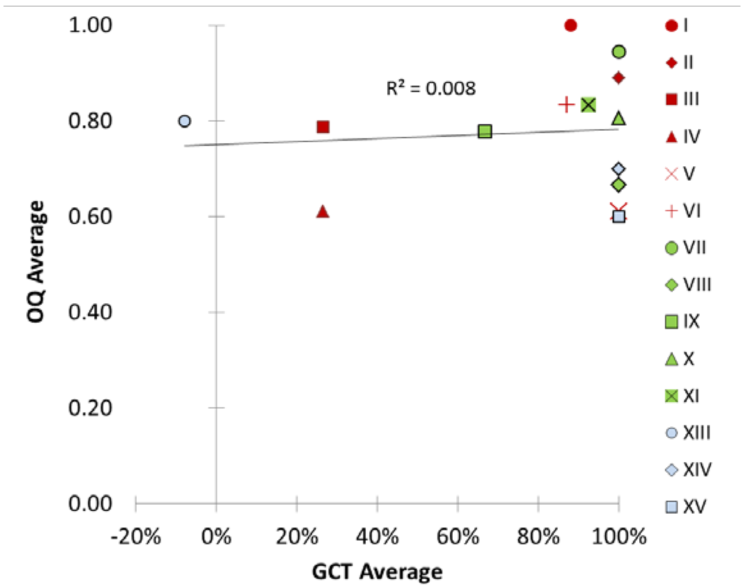


Figure 2-7 Average scores on the individual comparison questions for the OQ versus the GCT, with linear regression and corresponding Pearson correlation coefficient ($R^2 = 0.008$). The average standard error was ±8% for the GCT and ±0.03 for the OQ.

Finally, Fig. 2-8 compares the average performances (averaged over all students) on each comparison question for the RQ versus the OQ. 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, and reflective and traditional homework assignments, etc. Fig. 2-8 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 multi-electron atom, i.e., the total spin angular momentum quantum number S is highest in the ground state. (Note: Questions III, IV, and V do not appear in Fig. 2-8 because there was no RQ for those questions.)

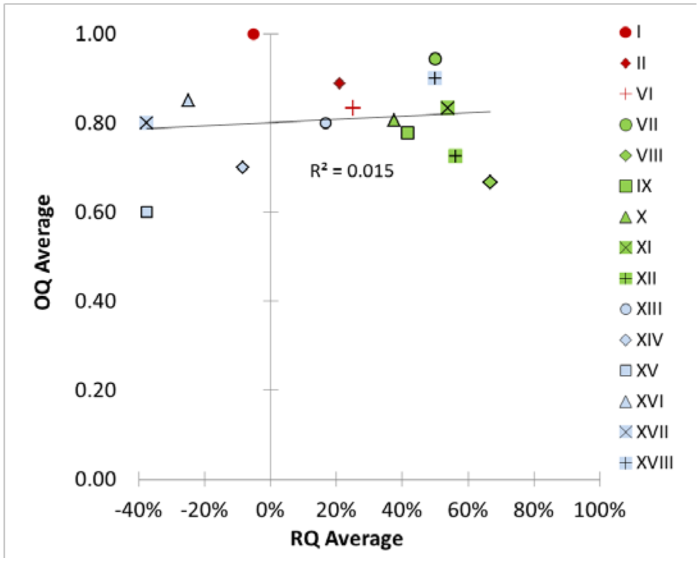


Figure 2-8 Average scores on the individual comparison questions for the OQ versus the RQ, with linear regression and corresponding Pearson correlation coefficient ($R^2 = 0.015$). The average standard error was ±8% for the RQ and ±0.03 for the OQ.

2.4.3 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 group for all questions. Row 1 (with data) in Table 2-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 2-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) 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.

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

Students in the QM course sometimes responded correctly to the ICT but then responded incorrectly to the corresponding GCT. Figure 2-9 shows a comparison of the fraction of times each student switched from a correct response on the ICT to an incorrect response on 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. 2-9 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. 2-9 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 ($p = 0.028$).

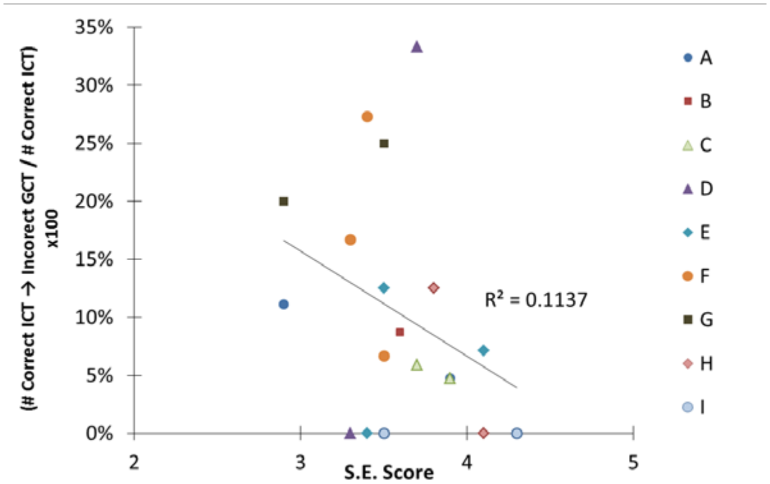


Figure 2-9 The number of times each student switched from a correct ICT response to an incorrect GCT response divided by the total number of correct ICT responses for that student (×100%) versus each student's self-efficacy score. The average standard error for S.E. score was ±0.086 and for percentage of correct ICT switched to incorrect GCT was ±2.30%.

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. 2-10. We hypothesized that in addition to students having a better understanding 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 2-10 shows that for the first five weeks of the course, 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. 2-10 in week 6 may be the difficulty associated with the concept of degenerate perturbation theory which was the focus.

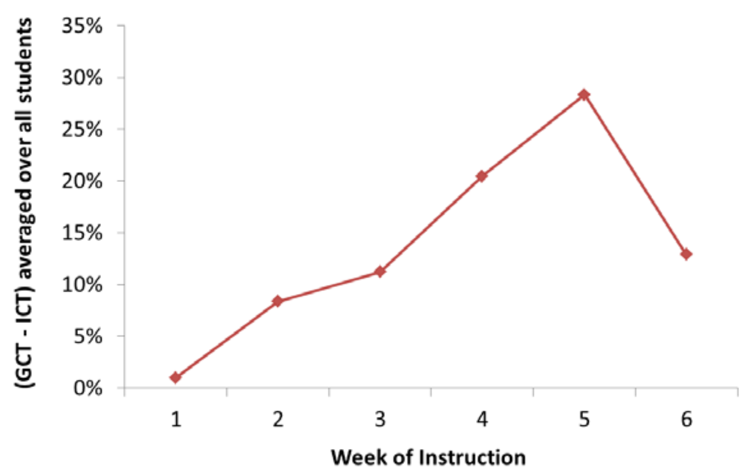


Figure 2-10 (GCT - ICT) for each week of instruction (averaged over all students and all questions for that week). The average standard error was ±8.56%.

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 non-response 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 peer discussion 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 2-11 shows a comparison of how likely individual students were to not respond on the ICT vs. the GCT. It shows the number of non-responses 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 2-11 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.

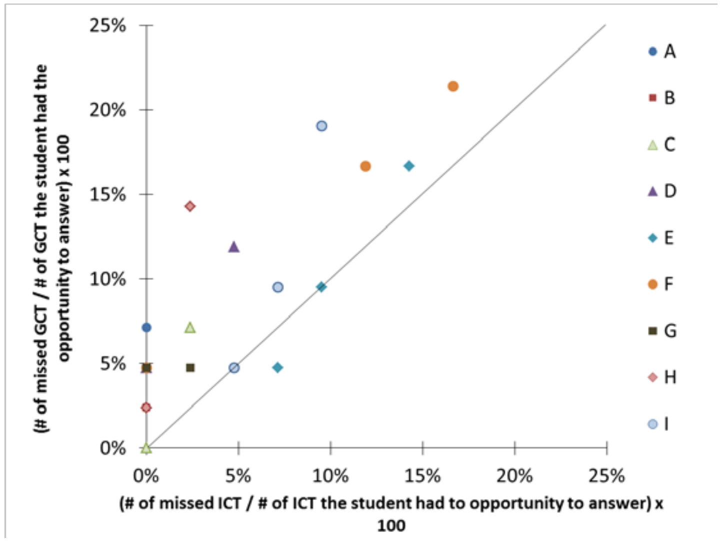


Figure 2-11 The *x*-axis denotes the number of times each student did not respond to an ICT divided by the number of ICT the student had the opportunity to answer × 100%; the *y*-axis denotes the number of times each student did not respond to a GCT divided by the number of GCT the student had the opportunity to answer × 100% for each student. The average standard error for missed ICT percentage was ±1.19% and for missed GCT percentage was ±1.36%.

Figure 2-12 shows the average non-response percentage for the whole class for each week of instruction. In response to Research Question 8 ("Are students equally likely to respond to in-class clicker questions at the beginning of the semester and later in the semester?"), Fig. 2-12 indicates that the first two weeks of the course had much higher non-response rates on both the ICT and GCT. However, the non-response 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 non-response 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 and develop the habit of regularly clicking in response to all clicker questions posed. Moreover, Fig. 2-12 is consistent with Fig. 2-11 in terms of the non-response rates being higher on average for the GCT than for the ICT.

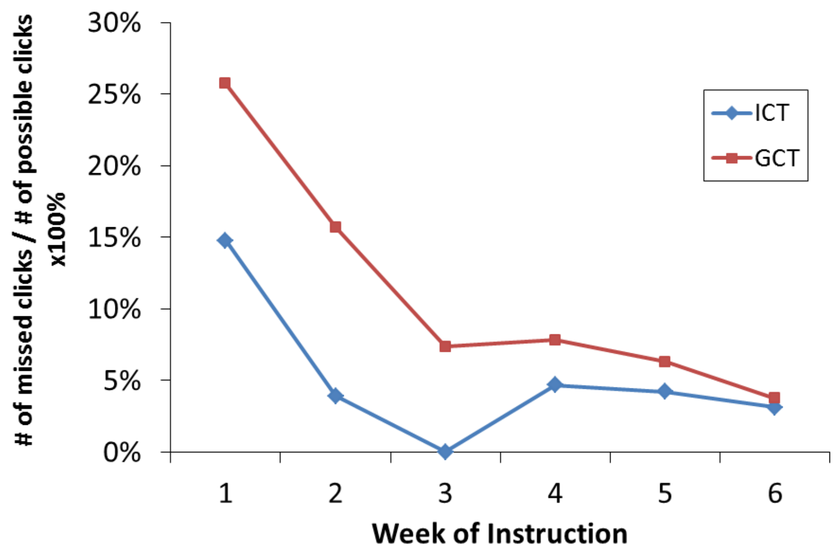


Figure 2-12 Student non-response on ICT (blue) and GCT (red) as a percentage of total possible responses per week of instruction. The average standard error was ±2.79% for ICT and ±3.86% for GCT.

2.5 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 in both 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 their prior learning and help 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]. In our investigation, the pre-lecture homework based upon out-of-class reading of the textbook asked students to summarize what they read and share with the instructor the parts of the reading material they found difficult via the course website. However, these pre-lecture assignments did not require students to explicitly elaborate upon and be specific about their difficulties or explicitly reflect on the reasons they found those parts of the reading to be difficult.

It appears that the out-of-class activities 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 pre-lecture 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 pre-lecture 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 pre-lecture reading did not convey deep productive struggle with the reading material. For example, one student wrote the following about his pre-lecture 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 pre-lecture 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 pre-lecture 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 pre-lecture 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 meta-cognitive skills. Prior research also suggests that only providing students worked examples is insufficient [33], and effective approaches to learning involve students engaged in meta-cognition and self-monitoring while they solve problems [34-36].

The homework that was based upon pre-lecture 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 pre-lecture 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 either attempted to learn themselves 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 pre-lecture 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 pre-lecture 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 pre-lecture reading were given online and were not graded, but it was suggested that students complete the pre-lecture 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 pre-lecture 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 pre-class 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 Section 2.3, 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.

Furthermore, 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 sub-problems (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 sub-problems, 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 2-II).

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. Peer interaction keeps students alert and on their toes because they must explain their reasoning to peers. Interacting with peers can also be fun. Also, since learning with peers is embedded in social context, it may be easier to retrieve that knowledge later.

For students who benefited significantly from peer interaction, struggling to answer the ICT may have been productive and prepared them to learn from interaction with their peers [8]. Another reason why peer interaction may have helped students learn is because the peer interaction was extended over a period of time and students may have begun to realize that their peers struggle with the same concepts. They could then attribute their struggles to the difficulty of the subject matter rather than personal factors, which might motivate them to be less anxious while learning the QM concepts.

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.

2.6 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 pre-lecture 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 pre-lecture 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 non-response rates for the entire semester was generally higher on the GCT than on the ICT. These higher non-response 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 non-response rates have never been reported in introductory physics.

2.7 CHAPTER REFERENCES

1. G. Novak, E. Patterson, A. Gavrin, and W. Christian, *Just-In-Time-Teaching: Blending Active Learning with Web Technology* (Prentice Hall, Upper Saddle River, NJ, 1999).

2. D. Schwartz and J. Bransford, A time for telling, Cognition and Instruction **16**(4), 475 (1998).

3. S. Formica, J. Easley, and M. Spraker, Phys. Rev. ST PER **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).

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, *Proceedings of the 2010 Phys. Ed. Res. Conference, Portland, OR* (AIP Conf. Proc., Melville, NY, 2010), p. 41; Impact of guided reflection with peers on the development of effective problem solving strategies and physics learning, The Physics Teacher **54**, 295 (2016).

8. D. Zingaro, Peer instruction contributes to self-efficacy in CS1, SIGCSE '14: Proceedings of the 45th ACM Technical Symposium on Computer Science Education, 373 (2014).

9. K. Miller, J. Schell, A. Ho, B. Lukoff, and E. Mazur, Response switching and self-efficacy in Peer Instruction classrooms, Phys. Rev. ST PER **11**, 010104 (2015).

10. A. Collins, J. Brown, and S. Newman, Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics, in *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, edited by L. Resnick (Lawrence Erlbaum, Hillsdale, NJ, 1989), pp 453-494.

11. For example, see J. Docktor and J. Mestre, Synthesis of discipline-based education research in physics, Phys. Rev. ST PER **10**, 020119 (2014).

12. C. Singh, What can we learn from PER: Physics Education Research?, The Phys. Teacher **52**, 568 (2015).

13. P. Jolly, D. Zollman, S. Rebello, and A. Dimitrova, Visualizing potential energy diagrams, Am. J. Phys. **66**(1), 57 (1998).

14. M. Wittmann, R. Steinberg, and E. 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, in *Proceedings of the 2005 Phys. Ed. Res. Conference, Salt Lake City, UT*, edited by P. Heron, L. McCullough, and J. Marx (AIP Conf. Proc., Melville, NY, 2006), pp. 73-76.

16. C. Manogue and E. Gire, Representations for a spins first approach to quantum mechanics, in *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE*, edited by N. Rebello, P. Engelhardt and C. Singh (AIP Conf. Proc., Melville, NY, 2012), pp.55-58.

17. E. Gire and C. Manogue, Making sense of operators, eigenstates, and quantum measurements, in *Proceedings of the 2011 Physics Education Research Conference, Omaha, NE*, edited by N. Rebello, P. Engelhardt, and C. Singh (AIP Conf. Proc., Melville, NY, 2011), pp. 195-198.

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, Physics Today **8**, 43 (2006); C. Singh, Assessing and improving student understanding of quantum mechanics, *Proceedings of the 2005 Phys. Ed. Res. Conference, Salt Lake City, UT* (AIP Conf. Proc., Melville, NY, 2006), p. 69; C. Singh, Student difficulties with quantum mechanics formalism, *Proceedings of the 2006 Phys. Ed. Res. Conference, Syracuse, NY* (AIP Conf. Proc., Melville, NY, 2007), p. 185; C. Singh, Helping students learn quantum mechanics for quantum computing, *Proceedings of the 2006 Phys. Ed. Res. Conference*, *Syracuse*, *NY* (AIP Conf. Proc., Melville, NY, 2007), p. 42; G. Zhu and C. Singh, Improving students' understanding of quantum mechanics via Stern-Gerlach experiment, Am. J. Phys. **79**(5), 499 (2011); A. Mason and C. Singh, Do advanced students learn from their mistakes without explicit intervention?, Am. J. Phys. **78**(7), 760 (2010); C. Singh and E. Marshman, Review of student difficulties in upper-level quantum mechanics, Phys. Rev. ST PER **11**, 020117 (2015).

19. M. Belloni, W. Christian, and D. Brown, Open source physics curricular material for quantum mechanics, Computing in Science and Engineering **9**, 24 (2007).

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. G. Passante, P. Emigh, and P. Shaffer, Investigating student understanding of basic quantum mechanics in the context of time-dependent perturbation theory, in *Proceedings of the 2013 Phys. Ed. Res. Conference, Portland, OR* (AIP Conf. Proc., Melville, NY, 2014), p. 269; Student ability to distinguish between superposition states and mixed states in quantum mechanics, Phys. Rev. ST PER **11**(2), 020135 (2015); Examining student ideas about energy measurements on quantum states across undergraduate and graduate level, Phys. Rev. ST PER **11**(2), 020111 (2015); Student understanding of time dependence in quantum mechanics, Phys. Rev. ST PER **11**(2), 020112 (2015).

22. C. Singh, Transfer of learning in quantum mechanics, *Proceedings of the 2004 Phys. Ed. Res. Conference, Sacramento, CA* (AIP Conf. Proc., Melville, NY, 2005), p. 23; C. Singh, Student understanding of quantum mechanics at the beginning of graduate instruction, Am. J. Phys. **76**, 277 (2008); C. Singh, 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, *Proceedings of the 2009 Phys. Ed. Res. Conference, Ann Arbor, MI* (AIP Conf. Proc., Melville, NY, 2009), p. 63; G. Zhu and C. Singh, Surveying students' understanding of quantum mechanics in one spatial dimension, Am. J. Phys. **80**(3), 252 (2012); G. Zhu and C. Singh, Improving students' understanding of quantum measurement I: Investigation of difficulties, Phys. Rev. ST PER **8**(1), 010117 (2012); G.

Zhu and C. Singh, Improving students' understanding of quantum measurement II: Development of Research-based learning tools, Phys. Rev. ST PER **8**(1), 010118 (2012); G. Zhu and C. Singh, Improving Student understanding of addition of angular momentum in quantum mechanics, Phys. Rev. ST PER **9**(1), 010101 (2013); S. Lin and C. Singh, Assessing expertise in quantum mechanics using categorization task, *Proceedings of the 2009 Phys. Ed. Res. Conference, Ann Arbor, MI* (AIP Conf. Proc., Melville, NY, 2009), p. 185; S. Lin and C. Singh, 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 PER **11**, 020119 (2015).

24. P. Pintrich, A motivational science perspective on the role of student motivation in learning and teaching contexts, J. Educ. Psych. **95**(4), 667 (2003); J. Ford, E. Smith, D. Weissbein, S. Gully and E. Salas, Relationship of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer, Journal of Applied Psychology **83**(2), 218 (1998); D. Moos and R. Azevedo, Exploring the fluctuation of motivation and use of self-regulatory processes during learning with hypermedia, Instructional Science **36**, 203 (2008); H. Song and B. Grabowski, Stimulating intrinsic motivation for problem solving using goal-oriented contexts and peer group composition, Educational Technology Research and Development **54**(5), 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. Schunk and B. Zimmerman (New York, Lawrence Erlbaum, 2008), pp 169-195; S. Sungur, Contribution of motivational beliefs and metacognition to students' performance under consequential and nonconsequential test conditions, Educational Research and Evaluation **13**(2), 127 (2007); C. Hulleman, A. Durik, S. Schweigert, and J. Harachiewicz, Task values, achievement goals, and interest: An integrative analysis, J. Educ. Psych. **100**(2), 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** (1), 43 (2006); P Hsieh, J. Sullivan, and N. Guerra, A closer look at college students: Self-efficacy and goal orientation, Journal of Advanced Academics **18**(3), 454 (2007); D. Moos and R. Azevedo, Learning with computer-based learning environments: A literature review of computer self-efficacy, Rev. of Educ. Res. **79** (2), 576 (2009); P. Sins, W. van Joolingen, E. 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. Psych. **33**(10), 58 (2008).

26. B. Zimmerman and D. 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. Schunk and B. Zimmerman (Lawrence Erlbaum, Mahwah, NJ, 2001), pp. 153-189; B. Zimmerman, Development and adaptation of expertise: The role of self-regulatory processes and beliefs.

In *The Cambridge Handbook of Expertise and Expert Performance* (Cambridge University Press, London, 2006), pp. 705-722; J. 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, Contemporary Educational Psychology **34**, 18 (2009); J. Fryer and A. Elliott, Self-regulation of achievement goal pursuit, in *Motivation and Self-regulated Learning: Theory, Research, and Applications*, edited by D. Schunk and B. Zimmerman (Lawrence Erlbaum, New York, 2009), pp.53-75.

27. B. White and J. Frederiksen, A theoretical framework and approach for fostering metacognitive development, Educational Psychologist **40**(4), 211 (2005); L. Jiang, J. Elen, and G. Clarebout, The relationships between learner variables, tool-usage behavior and performance, Computers in Human Behavior **25**, 501 (2009).

28. D. Griffiths, *Introduction to Quantum Mechanics* (Pearson Prentice Hall, 2004).

29. P. Cohen, J. Cohen, L. Aiken, and S. West, The problem of units and the circumstance for POMP, Multivariate Behavioral Research **34**(3), 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 edition) (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, Cognition and Instruction **22**, 129 (2004).

32. M. Kapur, Productive failure, Cognition and Instruction **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, Review of Educational Research **70**, 181 (2000).

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), pp. 189-215.

35. G. Schraw, Promoting general metacognitive awareness, Instructional Science **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 PER **6**, 010108 (2010).

2.8 APPENDIX A

This is a list of the comparison questions that were administered to the students. Each question is first shown as it appears in the RQ, ICT, and GCT in multiple-choice format. In each case, the particular statement we are investigating as well as the responses corresponding to that statement are in bold and the fully correct response is underlined. Each question is then shown as it appears in the open-ended retention quiz (OQ), with an explanation of the grading criteria in italics.

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

<u>**D. 1 and 3 only**</u>

E. 2 and 3 only

I. (OQ) Briefly explain the origin of the Hund's rules used for determining total spin angular momentum quantum number S for the ground state of multi-electron 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

<u>D. 1 and 3 only</u>

E. 2 and 3 only

II. (OQ) 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 $\psi_{220} = R_{22}(r) \cdot Y_2{}^0\,(\theta, \phi)$. 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 $2\pi r^2 |R_{22}(r)|^2 dr \int_0^{\pi} \left|Y_2^0(\theta, 0)\right|^2 \sin\theta\, d\theta$.

A. 1 only

B. 2 only

C. 3 only

<u>D. 2 and 3 only</u>

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. (OQ) 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

<u>**D. 1 and 3 only**</u>

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

<u>**D. 1 and 3 only**</u>

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. (OQ) 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, $\psi_{\mathrm{nlm}}(\mathrm{r}_1)\psi_{100}(\mathrm{r}_2)$. Explain why you agree or disagree with the following statement: For the composite wavefunction 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 2*N* copper atoms, respectively. Choose all of the following statements that are correct.

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

(2) At temperature $T = 0K$, 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

<u>**D. 2 and 3 only**</u>

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. (OQ) Cubes *A* and *B* with the same atom number density have *N* and 2*N* copper atoms, respectively. At temperature $T = 0K$, 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 2*N* copper atoms, respectively. Choose all of the following statements that are correct.

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

(2) At temperature $\boldsymbol{T = 0K}$, 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

<u>D. 2 and 3 only</u>

E. all of the above

VII. (OQ) Cubes *A* and *B* with the same atom number density have *N* and 2*N* copper atoms, respectively. At temperature $T = 0K$, which cube has the higher total energy associated with the valence electrons? *(Students who said that cube B his the higher total energy received credit for this question.)*

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

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

(2) At temperature $T = 0K$, 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

<u>D. 2 and 3 only</u>

E. all of the above

VIII. (OQ) 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 non-interacting 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 $\mathbf{27E_0}$, the probability of obtaining $\mathbf{E_0}$ is 1/2.**

A. 1 only

B. 3 only

C. 1 and 2

D. 1 and 3 only

E. none of the above

IX. (OQ) 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 spin-less 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. (OQ) There are two non-interacting particles in a one-dimensional infinite square well. The total energy of the two particle system is $E_{n_1n_2} = (n_1^2 + n_2^2)E_0$, 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 ½ 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. (OQ) 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 $\widehat{H}^0$ is diagonal in a given basis and there is no degeneracy in the eigenvalue spectrum of $\widehat{H}^0$ and $\widehat{H}'$, then $\widehat{H}'$ *must* be diagonal in that basis.

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

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

A. 1 only

B. 1 and 2 only

<u>**C. 1 and 3 only**</u>

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. (OQ) Suppose that in an N dimensional vector space ($N > 2$), the energy spectrum of the unperturbed Hamiltonian $\widehat{H}_0$ has a two-fold degeneracy. A perturbation $\widehat{H}'$ acts on this system. $\widehat{H}_0$ and $\widehat{H}'$ ***commute*** with each other. Consider the following statement: "If we choose a basis in which $\widehat{H}_0$ is diagonal, $\widehat{H}'$ 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 $\widehat{H}^0$ is two-fold degenerate, i.e., $\widehat{H}^0\psi_a{}^0 = E_1{}^0\psi_a{}^0$, $\widehat{H}^0\psi_b{}^0 = E_1{}^0\psi_b{}^0$, $\langle\psi_a{}^0|\psi_b{}^0\rangle = 0$. A perturbation $\widehat{H}'$ acts on this system and a Hermitian operator $\hat{A}$ commutes with both $\widehat{H}^0$ and $\widehat{H}'$. Choose all of the following statements that are correct.

(1) $\widehat{H}^0$ and $\widehat{H}'$must commute with each other.

(2) If $\psi_a{}^0$ and $\psi_b{}^0$ are degenerate eigenstates of $\hat{A}$, they must be "good" states for finding perturbative corrections to the energy and wavefunction due to $\widehat{H}'$.

(3) If $\psi_a{}^0$ and $\psi_b{}^0$are non-degenerate eigenstates of $\hat{A}$, they must be "good" states.

A. 1 only

B. 2 only

<u>**C. 3 only**</u>

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. (OQ) 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

<u>B. 2 only</u>

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. (OQ) 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 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 $V_0\begin{pmatrix} -\varepsilon & 0 \\ 0 & 0 \end{pmatrix}$.

A. 1 only

<u>B. 2 only</u>

C. 1 and 3 only

D. 2 and 3 only

E. All of the above.

XV. (OQ) 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 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 $|c\rangle$ 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 $\hat{H}^0 = -\frac{\hbar^2}{2m}\nabla^2 - \frac{e^2}{4\pi\varepsilon_0}\frac{1}{r}$. 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. (OQ) A perturbation $\hat{H}'$ acts on a hydrogen atom with the unperturbed Hamiltonian $\hat{H}^0 = -\frac{\hbar^2}{2m}\nabla^2 - \frac{e^2}{4\pi\varepsilon_0}\frac{1}{r}$. 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 $\hat{H}^0 = -\frac{\hbar^2}{2m}\nabla^2 - \frac{e^2}{4\pi\varepsilon_0}\frac{1}{r}$. 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 $\boldsymbol{\hat{H}' = \alpha\delta(r)}$, the first order correction to energy is $\boldsymbol{E^1 = \langle n, l, s, j, m_j|\hat{H}'|n, l, s, j, m_j\rangle}$.

(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. (OQ) A perturbation $\hat{H}'$ acts on a hydrogen atom with the unperturbed Hamiltonian

---

$\widehat{H}^0 = -\frac{\hbar^2}{2m}\nabla^2 - \frac{e^2}{4\pi\varepsilon_0}\frac{1}{r}$. For the perturbation $\widehat{H}' = \alpha\delta(r)$, 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.)*

XVIII. A perturbation $\widehat{H}'$ acts on a hydrogen atom with the unperturbed Hamiltonian $\widehat{H}^0 = -\frac{\hbar^2}{2m}\nabla^2 - \frac{e^2}{4\pi\varepsilon_0}\frac{1}{r}$. 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 $\widehat{H}' = \alpha\widehat{L}_z$, where α is a suitable constant, we can calculate the first order corrections as $E^1 = \langle n, l, s, j, m_j|\widehat{H}'|n, l, s, j, m_j\rangle$.

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

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

A. 1 only

B. 1 and 2 only

C. 1 and 3 only

<u>**D. 2 and 3 only**</u>

E. All of the above

XVIII. (OQ) A perturbation $\widehat{H}'$ acts on a hydrogen atom with the unperturbed Hamiltonian

$\widehat{H}^0 = -\frac{\hbar^2}{2m}\nabla^2 - \frac{e^2}{4\pi\varepsilon_0}\frac{1}{r}$. For the perturbation $\widehat{H}' = \alpha\hat{J}_z$, 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.)*

3.0 DEVELOPMENT AND EVALUATION OF A QUANTUM INTERACTIVE LEARNING TUTORIAL ON THE DOUBLE-SLIT EXPERIMENT

3.1 INTRODUCTION

According to a poll of Physics World readers, the interference of single electrons in a double slit experiment is “the most beautiful experiment in physics” [1]. The beauty of this experiment comes from its powerful illustration of the quantum nature of microscopic particles. This experiment (schematic diagram of the experimental setup shown in Fig. 3-1) is useful for helping students learn about foundations of quantum mechanics, including the wave-particle duality of a single particle, the probabilistic nature of quantum measurements, collapse of the wavefunction upon measurement, etc. It illustrates how information about which slit a particle went through, or “which-path” information, can destroy the interference pattern on the distant screen when a large number of single particles are sent [2,3]. Prior research on student learning of quantum mechanics has found that many students struggle with foundational concepts in quantum mechanics after instruction and many tools have been developed which can help improve student understanding of these concepts [4-10]. Here, we discuss the development and evaluation of a research-validated interactive tutorial designed to help students develop a good grasp of the foundational issues in quantum mechanics in the context of the double-slit experiment (DSE).

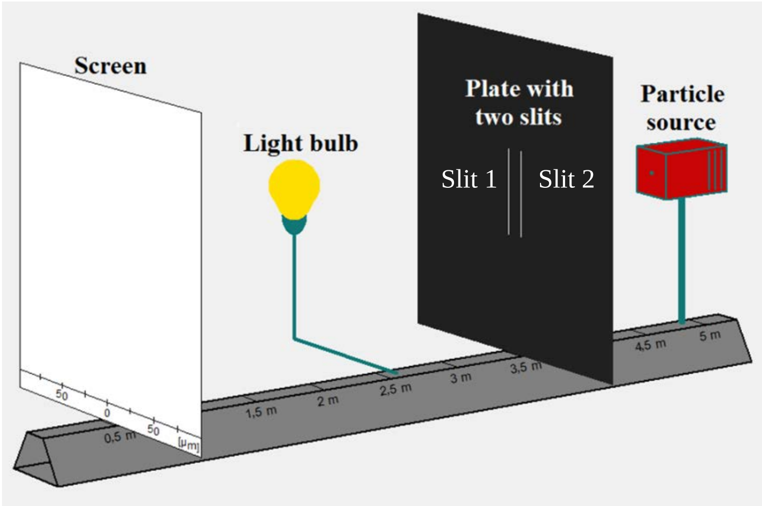


Figure 3-1 The basic setup of the double-slit experiment with single particles, consisting of a particle source, a plate with two narrow slits (labeled Slit 1 and Slit 2), a monochromatic lamp (light bulb) placed near the two slits, and a screen which detects the particles.

The development and use of research-based tools to help students learn upper-level quantum physics has been a subject of continuing interest. Our group has investigated the difficulties students have in learning various concepts in upper-level quantum mechanics, and developed and evaluated research-validated interactive tutorials or Quantum Interactive Learning Tutorials (QuILTs) [10]. The use of research-validated QuILTs in upper-level quantum mechanics courses shows that they help students develop a good grasp of quantum mechanics concepts [10]. The QuILTs use a guided approach to learning and often incorporate interactive simulations. They are structured in a way which allows students to make predictions and observe the outcome of a simulated experiment in a computer simulation, after which they are guided to reconcile the difference between what they predict and what they observe and extend and repair their knowledge structure. In other words, the students are asked to compare their observations with their predictions, and if their predictions do not agree with the simulation, they are given scaffolding support and feedback to reconcile the differences. The QuILTs provide students with appropriate guidance and prompt feedback as they strive to extend, organize, and repair their knowledge structure related to foundational issues in quantum mechanics using concrete examples. Previous QuILTs have been developed on topics such as the possible wavefunction, bound state and scattering state wavefunctions, time-development of wavefunction, uncertainty principle, Stern-Gerlach experiment, quantum key distribution, quantum measurement, Larmor precession of spin, addition of angular momentum, and the Mach-Zehnder interferometer with single photons and quantum eraser [11-14, 22-23].

Here, we discuss the development and evaluation of a research-validated QuILT on the DSE involving single particles sent one at a time through the slits [7-8]. We first discuss theoretical frameworks which inform our investigation. Next, we discuss common student difficulties we identified related to the DSE with single particles sent one at a time through the slits and describe how the DSE QuILT was developed. In particular, the development process took into account these difficulties via an iterative procedure to help students build a coherent knowledge structure of foundational concepts of quantum mechanics such as wave-particle duality, quantum measurement and collapse of the wavefunction using an inquiry-based approach. We then discuss the analysis of pre-test and post-test data to evaluate the improvement in student understanding of concepts covered in the DSE QuILT and to determine the extent to which the DSE QuILT is effective in addressing common difficulties of upper-level undergraduate and graduate students related to the quantum mechanics of the DSE.

3.2 THEORETICAL FRAMEWORKS

Research on student reasoning difficulties in learning upper-level quantum mechanics is inspired by cognitive theories that highlight the importance of knowing student difficulties in order to help them develop a functional understanding. According to the cognitive apprenticeship model, students can learn relevant concepts and develop effective problem-solving strategies if the instructional design involves three essential components: modeling, coaching and scaffolding, and weaning [15]. 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 [16-17]. Instructors typically give a lecture explaining the topics and demonstrate how to solve a few example problems. Students are then told to practice applying the skills on their own on homework with no guidance and little feedback (except for correct/incorrect after turning in the homework). Additionally, years of teaching experience and practice often make the instructor's reasoning and problem-solving skills implicit: they no longer have to think about what they are doing at each step, which is a hallmark of expertise. This suggests that as they are lecturing to students they may be deficient in modeling effective problem solving because they are no longer explicitly aware of their problem solving skills which have become automatic. In other words, students are often expected to learn and apply the expert-like practices not modeled explicitly by their instructors when working on the homework problems on their own. This situation is akin to a piano instructor demonstrating for the students how to play a particular musical piece and then asking students to practice on their own. The lack of prompt feedback and scaffolding support can be detrimental to learning. Advanced students are still developing expertise in quantum mechanics, and they need coaching and prompt feedback in order to develop expertise and build a robust knowledge structure. Research-validated QuILTs, which use a guided inquiry-based approach to learning, can provide students the opportunity to receive coaching and scaffolding as they engage in a guided exploration of quantum physics concepts.

Schwartz and Bransford's framework of "preparation for future learning" (PFL) suggests that in order to facilitate transfer of learning from one context to another, instructional design should include elements of both innovation and efficiency [18]. While there are many interpretations of the PFL framework, efficiency and innovation can be considered two orthogonal dimensions in instructional design. If the instructor only focuses on efficiently transferring information, cognitive engagement will be diminished and learning will be less effective. Conversely, if the instructional design is solely focused on innovation, students will struggle to connect what they are learning with their prior knowledge and learning and transfer will be inhibited. Incorporating the elements of efficiency and innovation into an instructional design based upon this framework demands that instruction build on students' existing knowledge and level of expertise. Innovation and efficiency are both incorporated in a guided active-learning approach via the QuILT: students are challenged to think through carefully designed questions (innovation) and are provided sufficient guidance (efficiency) to make progress. The QuILT strives to provide enough coaching and scaffolding to allow students to build a good knowledge structure while keeping them actively engaged in the learning process.

3.3 STUDENT DIFFICULTIES

During the development of the QuILT, we investigated the difficulties students have with the relevant concepts, including wave–particle duality, interference of a single particle with itself, and the collapse of a wavefunction upon measurement. Student difficulties involving the DSE with single particles were investigated by administering open-ended questions to upper-level undergraduate and graduate students in physics and conducting individual interviews with students in upper-level quantum mechanics courses after traditional instruction in relevant concepts. The traditional instruction in the undergraduate course included topics such as the de Broglie relation, calculation of the de Broglie wavelengths of different particles, an overview of the patterns that form on the distant screen in the DSE after a large number of single particles are sent one at a time through the slits, and a brief overview of the relevance of the information about which slit the particle went through to whether an interference pattern is observed on the screen. The open-ended questions were graded using rubrics which were designed to assess student understanding of relevant concepts by considering responses for multiple questions at once (an example of a specific question is provided later). A subset of the responses for all questions (20%-30%) was graded separately by two investigators. After comparing the grading of some students, the raters discussed any disagreements in grading and resolved them so that the inter-rater agreement after the discussions was better than 90%.

We conducted approximately 85 hours of individual interviews before, during, and after the development of different versions of the DSE QuILT and the corresponding pre-test and post-test. The interviews used a semi-structured, think-aloud protocol [19] and were designed to provide the researchers with a better understanding of the rationale students used to answer foundational questions related to the DSE. During the semi-structured interviews, upper-level undergraduate and graduate students were asked to verbalize their thought processes while answering the questions. Students read the questions related to the DSE setup and answered them to the best of their ability without being disturbed. They were prompted to think aloud if they became quiet for a long time. After students had finished answering a particular question to the best of their ability, they were often asked to further clarify and elaborate issues that they had not clearly addressed earlier. Below, we present a brief background for the DSE and discuss common difficulties identified in students' written work and think-aloud interviews related to the DSE.

3.3.1 Background on the DSE

Before discussing common student difficulties identified, we provide a brief background on the DSE shown in Fig. 3-1. In particular, we discuss how one may reason in terms of "which-path information" (WPI) to predict the pattern observed on the screen after a large number of single particles are emitted by the source. In this setup, the particle source emits single particles one at a time towards a plate with two narrow slits and are finally detected on the distant screen. We will use electrons for this discussion, but the reasoning we discuss can be applied to any other particle that is sufficiently small (e.g., protons, neutrons, Na atoms, etc.) to create an interference pattern under appropriate conditions with this setup. We assume that the parameters of the experiment, e.g., the distance between the narrow parallel slits and wavelength of the electrons are such that when the monochromatic lamp is turned off, an interference pattern is observed on the screen after a large number of electrons are detected. When the lamp is turned on, it emits photons of a certain wavelength which scatter off the electrons. For simplicity, we assume that this scattering process occurs very near or at the slits only. We also assume that a single particle only scatters a single photon, i.e., multiple scattering is neglected. The lamp has an intensity which can be varied from 0% to 100%, where 100% means that all of the electrons at the slits scatter off photons emitted by the lamp. Scattering between a photon and an electron corresponds to a measurement and it can localize the electron's position depending upon the wavelength of the photon emitted by the lamp. In other words, the scattering process localizes the electron in a region of length scale comparable to the wavelength of the photon. Therefore, if the wavelength of the photon is smaller than the distance between the slits, since we are assuming that the scattering process occurs at the slits, the scattering will provide information about the position of the electron during the scattering process: at one slit or at the other slit. This is what is referred to as "having WPI": knowing that the electron went through one slit or the other, but not both. In this case, if the intensity of the lamp is 100%, the interference pattern that would otherwise be observed on the screen (when the lamp is turned off) is destroyed due to scattering between an electron and a photon emitted by the lamp when the lamp is turned on. If the lamp is of intermediate intensity, say 50%, only half of the electrons scatter off of photons and do not interfere, whereas the other half do interfere. Therefore, the pattern observed on the screen after a large number of electrons are detected will be an interference pattern (50% of electrons that do not scatter) on top of a uniform background due to the 50% of the electrons which do not interfere (so overall, there will be a reduced contrast in the interference pattern).

If, instead, the wavelength of the photons is larger than the distance between the slits, scattering between an electron and a photon does not provide WPI because the length scale at which the photon can be used to resolve the electron's position at the slits is not small enough to be able to know that the electron goes through one slit or the other. In this case, the electron goes through both slits and we observe an interference pattern on the screen indistinguishable from when the lamp is turned off. Furthermore, the intensity of the lamp is irrelevant because regardless of whether an electron scatters off a photon or not, it will still interfere with itself.

We now discuss common difficulties that students have with this reasoning in various DSE setups which include a monochromatic lamp. The different setups corresponded to different lamp intensities and different wavelengths of the photons emitted by this lamp.

3.3.2 Difficulty Reasoning in Terms of "Which-path" Information

Many students struggle with the concept of WPI and its relevance to whether interference is observed on the screen if particles are sent one at a time through the slits. The concept of WPI at a detector (such as a screen) is useful when the state of the system is a superposition of two different spatial path states as in the DSE. In general, when a detector can project both components of the path state, then WPI is unknown. On the other hand, when a detector can project only one component of the path state, then we have complete WPI. For example, a single electron that is delocalized in space can go through both slits before reaching the screen and interfere with itself. In this case we do not have WPI for the electron, and interference of single electrons is observed on the screen. In other words, interference occurs because, as the electron wavefunction evolves when the electron travels from the slits to the screen, the two components related to the different path states pick up different phases that are related to the path lengths – from one or the other slit to the point on the screen where it is detected. Depending on the path length difference, the probability of detecting the electron (corresponding to the wavefunction absolute squared) varies as $\cos(\Delta\phi)$ (where $\Delta\phi$ is the phase difference) which results in an interference pattern. However, if we measure which slit the electron went through, the wavefunction collapses to one or the other path state at the slits, and when the electron reaches the screen, the detector (screen) can only project that particular path state and no interference is observed. In this case, by measuring the electron near one of the slits, we obtain the WPI for the electron and the electron cannot interfere with itself when it reaches the screen.

There is no analogue to the concept of WPI in classical mechanics, and many students find it difficult to reconcile their intuition with the quantum effects observed in the DSE when considering WPI and whether interference will be observed on the screen in a particular situation. The concept of WPI and its relation to the interference at the screen in the DSE can be difficult for students if they are not given appropriate scaffolding support as they learn these counter-intuitive concepts. We find that some students explain single particle interference by saying that one electron going through one slit interferes with another electron going through the other slit, even though they have been told in the beginning that a single electron is sent at a time. In other words, this concept of single electron interference is so difficult for students to grasp that they ignore relevant information provided (one electron at a time) and explain it in their own way.

3.3.3 Difficulty Recognizing the Effect of Lamp Wavelength on the Interference Pattern

When a monochromatic lamp is placed between the slits and the screen, the interaction between the incoming particles and the photons emitted by the lamp can localize the particles in some situations. For example, when the wavelength of the photons is significantly smaller than the distance between the two slits, the scattered photons will localize the incoming particles to one of the two slits, which provides WPI about the particles. This will destroy the interference pattern on the screen. Conversely, when the wavelength of the photons is much larger than the distance between the slits, the scattered photons will not localize the particles sufficiently to provide WPI.

Many students struggle to incorporate the wavelength of the lamp's photons into their responses to the DSE questions before working on the QuILT. In interviews, students were asked to predict the pattern that will be observed on the screen in a DSE when the wavelength of the photons is smaller than the distance between the slits. Many students claimed that the wavelength of the scattered photons is not important, and that only the intensity of the lamp matters. For example, in an interview, when asked explicitly about why he did not incorporate the wavelength of the scattered photon, one student simply noted that he thought that the answer to what happens to the interference pattern should be independent of the photon wavelength and only depend on how many photons are interacting with the single particles incident on the slits (the student felt that every incident particle that interacts with a photon will not show interference regardless of the photon's wavelength).

Some students also struggled to clearly differentiate the wavelength associated with particles such as electrons or atoms emitted by the particle source from the wavelength associated with the photons emitted by the lamp. This lack of differentiation led some students to claim that if a photon emitted by the lamp and a particle emitted by the particle source have the same wavelength, they can interfere with each other destructively and annihilate each other. For example, when asked to describe a situation in which the presence of the lamp will lead to the destruction of the interference pattern on the screen, several students described a scenario in which the photon and the particle from the slit destructively interfere with each other. Discussions suggest that these students were familiar with the concept of a particle behaving as a wave but had not yet developed a deeper understanding to realize that an electron and a photon with the same wavelength but with opposite phase cannot destructively interfere with each other. When asked in interviews to explain his reasoning, one student simply asserted that he feels this way because “*that’s just what I’ve been told [by his instructor in class]*.” This type of response suggests that even advanced students are likely to misinterpret what they learn from lectures particularly if what the instructor tells them is not consistent with their existing knowledge structure. Furthermore, these types of responses also convey an epistemology about learning quantum physics in which the advanced student views the instructor as an authority figure and accepts what the instructor says without questioning or making sense of it and integrating it with his or her existing knowledge structure.

3.3.4 Difficulty Recognizing the Effect of Lamp Intensity on the Interference Pattern

In addition to the difficulties incorporating photon wavelength into the DSE, many students have difficulty accounting for the role of lamp intensity in the DSE. When the wavelength of the photons is smaller than the distance between the two slits, WPI is available for each particle sent through the slits that scatters off a photon. The intensity of the lamp will then determine the fraction of the incoming particles that scatter off a photon, leading to WPI for those particles. If the lamp intensity is 50%, such that only half of the particles scatter a photon, then WPI will be available for half of the particles. In this case, the pattern on the screen will be a combination of interference fringes for half of the particles and a uniform background on the screen for the other half of the particles, i.e., an interference pattern with reduced contrast. When they first encounter this scenario, many students do not recognize that WPI is only available for the fraction of particles which scatter off a photon.

When the wavelength of the photons is much larger than the distance between the two slits, the scattered photons will not provide WPI about the particles incident on the slits. In this case, the scattered photons cannot resolve the particle sufficiently to localize it to one slit, so an interference pattern will be observed on the screen regardless of the intensity of the lamp. Many students struggle with the fact that lamp intensity does not matter when the wavelength of the lamp's photons is very large. Students were asked in individual interviews to predict the pattern that will form on the screen in the case in which the wavelength of the photons was much larger than the slit separation and the intensity of the lamp was initially 100% and then reduced to 50%. In response to this question, many of the students predicted that the patterns observed on the screen would be different in the two cases. When one student was asked to explain why he predicted different patterns in the two cases, instead of explaining a causal relation of some kind, he emphatically stated, "*It [the pattern] HAS to change in some way*." This type of a response from advanced students in the context of quantum mechanics illustrates a powerful phenomenological primitive that many beginning students possess [20], that when you change the input of a system, the output must always change in some way in response.

3.4 DEVELOPMENT OF THE QUILT, ITS STRUCTURE, AND LEARNING OBJECTIVES

3.4.1 Development and Validation of the DSE QuILT

The difficulties discussed above indicate that upper-level undergraduate and graduate students struggle to develop a coherent understanding of the foundational issues in quantum mechanics relevant for understanding whether interference will be observed on the screen after a large number of single particles pass through the slits in the DSE under various conditions. These students can benefit from a research-validated tutorial which uses a guided approach to help them learn these concepts involving single particles passing through a DSE. Therefore, we were motivated to develop a research-validated QuILT on the DSE with single particles.

The development of the QuILT was a cyclical, iterative process which included the following stages: (1) development of a preliminary version of the QuILT based upon a cognitive task analysis of the underlying concepts and knowledge of common student difficulties found via research; (2) implementation and evaluation of the QuILT by administering it to individual students, asking them to think aloud as they worked on it, and measuring improvement via their performance on pre-/post-tests; and (3) after determination of its impact on student learning and assessment of what difficulties were not adequately addressed by a particular version of the QuILT, making refinements and modifications based upon the feedback from the implementation and evaluation of the previous version.

Different versions of the QuILT were also iterated several times with five physics faculty members to ensure that experts agreed with the content and wording. The faculty feedback complemented the feedback obtained by having advanced students work on the QuILT in individual think-aloud interviews. These interviews helped to ensure that the guided approach was effective and the questions were unambiguously interpreted by students, as well as to better understand students' reasoning as they answered the questions. A total of approximately 85 hours of individual interviews were conducted with students during the development and assessment phases of the DSE QuILT.

3.4.2 Structure of the DSE QuILT

The guided approach used in the DSE QuILT helps students build on their prior knowledge and accounts for common student difficulties to help them develop a robust knowledge structure of foundational issues in quantum mechanics using the context of the DSE. The QuILT consists of these components to be used in the following order: a pre-test, a warm-up, a main tutorial, an associated homework component, and a post-test, as shown in Fig. 3-2. The pre-test consists of free-response questions involving the DSE with single particles and a monochromatic photon source placed between the slits and the screen. The photon source emits photons of a particular wavelength which scatter off the single particles at the slits. The warm-up serves to help students learn about the double slit experiment without the photon source placed between the slits and the screen and focuses on the de Broglie relation, wave-particle duality as manifested in the DSE, how the registering of a particle on the distance screen can be viewed as a measurement of position, and the impact of measurement on the wavefunction of the particle. The warm-up helps prepare students to learn the pre-requisite concepts and engage effectively with the main tutorial in the sequence. Students work on the QuILT in class in groups, and whatever they do not finish in class they work on at home. After working on the main tutorial which is conceptual in nature, students work on a homework component which connects the conceptual and mathematical aspects of the DSE to help students connect the conceptual and quantitative aspects of quantum mechanics involved in the experiment [2]. Finally, students work on a post-test which is identical to the pre-test.

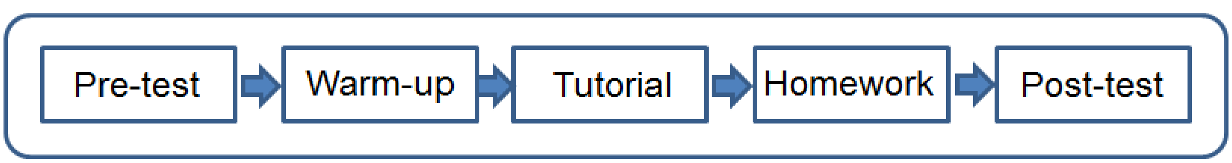


Figure 3-2 Sequence of components comprising the entire DSE QuILT suite.

The warm up and main tutorial make use of a computer simulation in which students can manipulate the DSE setup and observe the resulting pattern. This setup involves a plate with two slits, a particle source, a screen which serves as the detector for the experiment, and a photon source (light bulb) placed near the two slits, as shown in Fig. 3-1. Students are asked to predict the pattern that will appear on the screen based on the type of particles emitted by the source, their energy, the width and separation of the two slits, the wavelength of the photons emitted by the photon source, and the intensity of the photon source. Students then use the simulation to check their predictions. When a large number of particles has reached the screen, students can observe an interference pattern consisting of several dark and bright fringes or a featureless distribution without any interference fringes. Figure 3-3 shows a screenshot of the simulation in which an interference pattern has formed on the screen. Students can use the computer simulation to verify that there are interference fringes on the screen when the chosen parameters are used (as shown in Fig. 3-3). Figure 3-4 shows a screenshot of the simulation in which no interference pattern has formed on the screen after a large number of particles has reached it. Students can also observe a combination of the two in which the dark and bright fringes are still visible but they are on top of a uniform background of scattered particles that arrive at the screen (in which case, there is WPI for some photons but not for others and there is a reduced contrast in the interference pattern due to some photons, for which WPI is known, not displaying interference). Students are then given an opportunity to reconcile the difference between their predictions and observations before proceeding further in the tutorial. They are also provided checkpoints to reflect upon what they have learned and to make explicit connections with their prior knowledge.

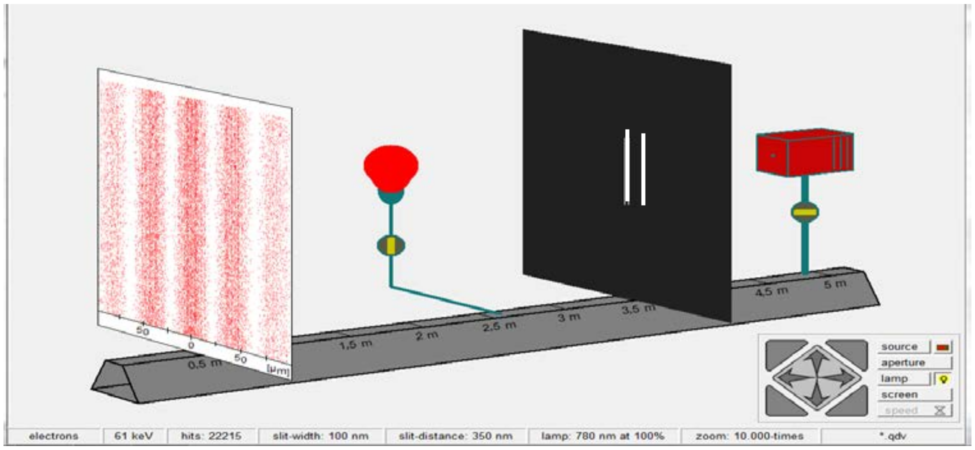


Figure 3-3 Screenshot of the computer simulation of the DSE (which is part of the QuILT) for a situation in which an interference pattern has formed on the screen after a large number of single particles have been sent through the slits to the screen.

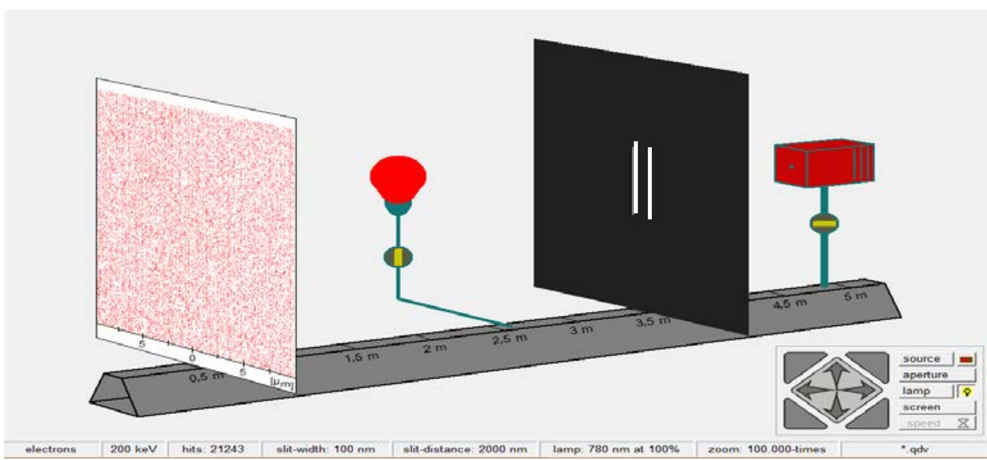


Figure 3-4 Screenshot of the computer simulation of the DSE for a situation in which no interference pattern is formed on the screen after a large number of single particles have been sent through the slits to the screen.

3.4.3 DSE QuILT Learning Objectives

The DSE QuILT focuses on helping students learn about interference of single particles in a DSE with a photon source placed near the two slits. In particular, the DSE QuILT was designed to address common student difficulties and help students develop a robust knowledge structure of the foundational quantum mechanical concepts involved (wave-particle duality, quantum measurement and collapse of wavefunction, etc.) by focusing on the following learning objectives:

Learning Objective 1: Recognize and understand why the photons which scatter off particles at the slits may provide WPI about the particles if the wavelength of the photon is shorter than the distance between the two slits.

As discussed in Section 3.3, many students struggle with reasoning in terms of WPI, which is a convenient conceptual framework for considering whether interference is observed in a particular situation or not. The QuILT is designed to help students make the connection between WPI and the presence or absence of an interference pattern on the screen in the DSE. The QuILT also helps students learn how to calculate the number density of electrons (and other particles incident on the two slits) when a large number of those particles arrive at a small region on the screen. In particular, students learn to incorporate WPI for the electrons to determine whether interference is observed when the electrons arrive at the screen and its impact on the pattern and the number density. They are first guided to find the number density for the case in which no WPI is available for the electrons sent through the slits and an interference pattern is observed on the screen. The students are then provided scaffolding support and appropriate feedback for the case in which WPI is available and determine the number density on the screen for this case (in this case, there is no interference pattern on the screen).

For example, in order to scaffold student learning, the following question in the QuILT asks students to think about what changes occur in the number density of particles on the screen based on whether WPI is available for the particles incident on the slits. (Note: In the notation used in the QuILT, $\psi_1(x)$ and $\psi_2(x)$ represent the wavefunction at point x on the screen when slit 2 or slit 1 is closed, respectively, and $\Delta\phi$ represents the phase difference between $\psi_1(x)$ and $\psi_2(x)$ at point x on the screen. When both slits are open, students must take into account both $\psi_1(x)$ and $\psi_2(x)$.)

"Circle all of the following statements about the double-slit experiment that are correct:

I. If the cross term $(2|\psi_1(x)| \cdot |\psi_2(x)| \cdot \cos\Delta\phi)$ in the expression for the expected number density of electrons is negligible in a given situation, interference effects will be negligible.

II. If we obtain WPI, i.e., information about which slit the electron went through, the cross term in the expression for the expected number density of electrons vanishes.

III. If we first square the wavefunction from each slit and then add the results to obtain the total probability density for a single electron, i.e., $|\psi(x)|^2 = |\psi_1(x)|^2 + |\psi_2(x)|^2$, and then sum over all electrons to obtain the expected number density of electrons at each point x on the screen, we would conclude that there are no interference effects.

Explain your reasoning." (Answer: *All three statements are correct*.)

After this question, which prompts students to connect their conceptual understanding of the WPI with the number density of electrons on the screen, students are provided guidance and support to help them build a coherent understanding of relevant concepts.

Learning Objective 2: Predict the qualitative features of the pattern that will form on the screen after a large number of particles have been sent through the slits depending on the <u>wavelength</u> of the photons that scatter off the particles.

After the QuILT guides students to reason about WPI and to incorporate it into the DSE to predict the pattern that forms on the screen after a large number of single particles are detected at the screen (without a monochromatic lamp between the slits and the screen), students learn about the role of photon wavelength in determining WPI for the particles incident on the slits. The QuILT then builds upon students' understanding of WPI by incorporating the simulation. It asks students to make predictions about the pattern that will form on the screen after a large number of particles reach the screen based upon the wavelength of the photons from the lamp. Students are then asked to use the simulation to check their predictions. If the simulation does not agree with their predictions, the students must reconcile the difference by reconsidering their reasoning when making the prediction. When students work in small groups in class, they discuss their predictions and observations with their peers. The QuILT then provides guidance and support to help them develop a good understanding of these issues.

Many students do not realize that the wavelength of the photons is related to the length scale over which the scattered photons can resolve an object. The QuILT often discusses using single electrons in the DSE, but students learn that everything that follows can be applied in a very similar manner to other particles. The QuILT includes other particles, and the source used in the simulation can be used to select between electrons, Na atoms, muons, etc. The following question in the QuILT uses a hypothetical conversation between three students to scaffold student learning about the role of photon wavelength in the DSE:

"Consider the following conversation between Pria, Mira and Nancy about why an important consideration in the loss of the interference fringes is the comparison of the slit separation with the wavelength of the photons emitted by the lamp:

- Pria: I think that we will always have WPI regardless of the wavelength of the photons emitted by the lamp as long as the lamp has high intensity. If the lamp has high intensity, virtually every electron will scatter off one photon. Therefore, we will be able to determine where each electron scattered from (which slit it went through) based upon the information about the scattered photon.

- Mira: I disagree with your conclusion. If the photon had very large wavelength compared to the distance between the slits, it would not matter if an electron scatters off a photon because diffraction will limit our ability to resolve length scales smaller than the wavelength of the photon. In this case, scattering does not provide information about which slit the electron went through. For example, due to diffraction, one cannot use an optical microscope to examine viruses because their size is smaller than the shortest wavelength of visible light.
- Nancy: I agree with Mira that you may not be able to resolve two things by using photons of a wavelength larger than the length you are trying to resolve. In this context, if we are using photons with a wavelength larger than the distance between the slits, from the point of view of a photon, those two slits overlap and could be regarded as indistinguishable. If instead, the wavelength of the photon is smaller than the distance between the slits, a photon which scatters off an electron at one slit or another can provide information about which slit the interference occurred.

Do you agree with Pria and/or Mira and Nancy? Explain your reasoning." (Answer: *Mira and Nancy are correct.*)

Learning Objective 3: **Predict the pattern that will form on the screen after a large number of single particles have been sent based on the intensity of the lamp from which photons are emitted and scatter off the particles**.

As discussed in Section 3.3, many students struggle to incorporate the intensity of the lamp from which photons are emitted and scatter off the particles in the DSE. The second part of the tutorial specifically addresses this difficulty by helping students make predictions about the pattern that will form on the screen based upon the intensity of the lamp from which photons are emitted that scatter off the electrons. The students first consider the limiting cases of 100% intensity (meaning that every electron scatters off a photon) and 0% intensity (meaning that none of the electrons scatter off a photon). The students are then guided to think about intermediate cases in which only some of the electrons scatter off a photon.

For example, the following question asks students to incorporate a lamp with an intensity such that half of the electrons scatter off the photons (but scattered electrons still arrive at the screen) into the DSE and make a prediction about the pattern that will form on the screen:

"Consider a case in which the lamp has intermediate intensity such that half of the electrons do not scatter off photons. Which one of the following statements is correct if the wavelength of the photons emitted by the lamp is significantly <u>less</u> than the distance between the slits?

A. The interference pattern will go away.
B. The interference pattern essentially remains unchanged.
C. The interference pattern is still visible, however, it is harder to discern because of reduced contrast.
D. The interference pattern becomes easier to discern because of increased contrast.

Explain your reasoning for your answer." (Answer: *C is correct*.)

Students are then prompted to use the simulation to check their prediction and reconcile differences, if any, between their prediction and observation. For example, after running the simulation, students are asked the following question:

“What happened to the interference pattern as you lowered the intensity? Is this observation consistent with your answer to the preceding question? If it is not, reconcile the difference between your prediction and observation.”

The QuILT provides guidance and scaffolding support and strives to help students develop a good grasp of foundational concepts in quantum mechanics using the concrete context of the DSE. After working on the QuILT, students are expected to be able to qualitatively reason about how a single particle can exhibit the properties of both a wave and a particle, and be able to determine the de Broglie wavelength of a particle based on its mass or energy. They should be able to describe how scattering between a photon and a particle can provide WPI depending on the wavelength of the photon and whether a particle can be localized over a distance smaller than the distance between the slits depending on the situation, and also describe how measurement of a particle's position at the screen collapses the wavefunction. Students are also expected to be able to explain the role of the photons from the lamp and how the photons from the lamp that scatter off the incoming particles can affect the presence of an interference pattern at the screen. Students should be able to reason about whether scattered photons give WPI about the particles after passing through the slits based on the wavelength of the photons, and be able to incorporate the intensity of the lamp into their predictions about what fraction of the particles incident on the slits will create interference fringes on the screen.

3.5 EVALUATION OF THE QUILT

Once it was determined that the QuILT was effective in meeting the learning objectives in individual administration, it was administered to students in two upper-level undergraduate quantum mechanics courses ($N = 46$) and graduate students who were simultaneously enrolled in the first semester of a graduate-level core quantum mechanics course and a course for training teaching assistants (TAs) ($N = 45$). First, the students were administered a pre-test. After the students worked on the pre-test, they worked through the warm-up and the main part of the QuILT in groups. They were given one week to work through the rest of the QuILT (including the homework component) and then submit it to the instructor as homework. They were then given a post-test in class. Any students who did not work through the QuILT for any reason were omitted from the post-test data.

The upper-level undergraduate students who were enrolled in a quantum mechanics course received full credit for taking the pre-test, the tutorial counted as a small portion of their homework grade for the course and their post-tests were graded for correctness as a quiz. In addition, the upper-level undergraduates were aware that topics discussed in the tutorial could also appear in future exams since the tutorial was part of the course material. The graduate students were enrolled in a TA training course along with the graduate level core quantum mechanics course. In the TA training course, the graduate students learned about instructional strategies for teaching introductory physics courses. They were asked to work through the QuILT in one TA training class to learn about the effectiveness of the tutorial approach to teaching and learning. It was considered that the graduate students would recognize the value of the tutorial approach better if they discussed tutorials on topics which they are familiar with but do not fully understand (as opposed to discussing tutorials in introductory physics for which many graduate students are likely to be experts). If graduate students engage with these tutorials, they can learn the topics discussed and understand the value of utilizing these tools as supplements to instruction. They were given credit for completing the pre-test, tutorial, and post-test. However, their scores did not contribute to the final grade for the TA training course (which was a pass/fail course).

The students' performance on the pre- and post-tests administered before and after they worked through the tutorial were used to assess the extent to which the learning objectives outlined in Section 3.4 were achieved. The pre-/post-test questions involve the following situations (the entire pre/post-test is given in Appendix B):

Question 1 (Q1) presents a DSE set-up with single electrons and asks students to describe a situation in which the introduction of a lamp between the slits and the screen close to the slits would destroy the interference pattern (although the electrons still arrive at the screen). A correct response mentions that the wavelength of the photons emitted by the lamp must be smaller than the separation between the two slits in order to localize the incoming electron sufficiently close to one of the two slits so that when the electron arrives at the screen we have WPI about which slit the electron went through.

Question 2 (Q2) presents a DSE using sodium (Na) atoms and asks students to calculate the number density at a point *x* on the screen and to describe the pattern observed after a large number of atoms reaches the screen. In the situation presented, the wavelength of the photons emitted by the lamp is *significantly smaller* than the slit separation, while the intensity of lamp is such that each Na atom scatters off a photon (but still arrives at the screen). The correct number density is $\frac{N}{2} \cdot |\psi_1(x)|^2 + \frac{N}{2} \cdot |\psi_2(x)|^2$ and the pattern on screen is no interference, which may be reasoned using WPI.

Question 3 (Q3) repeats the setup described in Q2, but now the wavelength of the photons emitted by the lamp is *significantly larger* than the slit separation. The correct number density is $\frac{N}{2} \cdot (|\psi_1(x)|^2 + |\psi_2(x)|^2 + 2|\psi_1(x)| \cdot |\psi_2(x)| \cdot \cos\Delta\phi)$ and the pattern on the screen is an interference pattern since the photons' wavelength is not small enough to localize the Na atoms sufficiently to provide WPI (about which slit each particle went through) after the scattering takes place.

Question 4 (Q4) and Question 5 (Q5) repeat Q2 and Q3, but now the intensity of the lamp has been decreased so that only half of the Na atoms scatter off the photons. The correct number densities are $\frac{N}{2}\cdot(|\psi_1(x)|^2+|\psi_2(x)|^2)+\frac{N}{2}\cdot|\psi_1(x)|\cdot|\psi_2(x)|\cdot\cos\Delta\phi$ and $\frac{N}{2}\cdot(|\psi_1(x)|^2+|\psi_2(x)|^2+2|\psi_1(x)|\cdot|\psi_2(x)|\cdot\cos\Delta\phi)$, and the patterns are partial interference (only Na atoms that do not scatter a photon show interference) and full interference (scattering does not localize Na atoms sufficiently to give WPI), respectively. The parameters for the photons that scatter off the Na atoms in the DSE situations for Q2 through Q5 are summarized in Table 3-I.

Table 3-I Summary of relevant properties of photons from the lamp that interact with Na atoms in the DSE pre- and post-test for Q2-Q5.

	Short Wavelength	Long Wavelength
Full Intensity	Q2	Q3
Half Intensity	Q4	Q5

Students' responses to Q1 through Q5 were categorized based on the most common types of responses in order to identify the students' specific difficulties. (Detailed analysis of student responses is included in Section 3.6.) Between 20-30% of the students were independently categorized by a second rater for each question/question pair, and an inter-rater agreement of greater than 90% was obtained in all cases.

3.5.1 Concept-based Rubric

Student performance on the pre- and post-tests was evaluated using a concept-based rubric which often used "holistic" scoring designed to assess student understanding of relevant concepts across multiple questions (as discussed below) in order to determine whether students had developed a coherent knowledge structure of the relevant foundational issues in quantum mechanics and had met the learning objectives outlined in Section 3.4. For example, Learning Objective 3 focuses on helping students learn that changing the wavelength of the photons may alter the interference pattern formed by the particles incident on the slits and why that would be the case under certain conditions. Students' responses to Q2 and Q3 were scored together in order to determine whether the students recognize and explain why (1) changing the wavelength of the photons that interact with the particles incident on the slits alters the interference pattern, and (2) a short wavelength photon (compared to the distance between the slits) localizes the particles (e.g., Na atoms) close to one slit or the other and therefore provides WPI, whereas a long wavelength photon does not. Similarly, Q4 and Q5 were scored together using the same criteria used to score Q2 and Q3. Thus, the concept-based rubric was aligned with the learning objectives outlined in Section 3.4. A summary of the grading rubric is shown in Table 3-II.

Table 3-II Summary of the rubric used to evaluate student responses to Q1, Q2-3, and Q4-5, with a total of two points possible for Q1 and eight points possible for each question pair (Q2-3 and Q4-5).

Q1	Possible Scores
1. Mention that scattering a photon localizes the particle and may provide WPI and destroy the interference pattern.	1,0
2. Mention that the wavelength of the photons must be smaller than the distance between the slits ($\lambda < d$) in order to provide WPI.	1,0
Total points possible	2
Q2-3 or Q4-5	
1. Mention that the photon wavelength is an important consideration in determining the pattern that forms on the screen.	1,0
2. Correctly interpret the effect of wavelength on the interference pattern. (1 point possible for each question.)	2,1,0
3. Find different number densities for the two questions (whether or not they are correct).	1,0
4. Number densities are correct. (1 point possible for each question.)	2,1,0
5. Number densities are consistent with patterns. (1 point possible for each question.)	2,1,0
Total points possible	8

Table 3-III Transcribed solutions of Student A and Student B to Q2 and Q3.

	Student A
Q2	$\frac{N}{2}(\lvert\psi_1\rvert^2 + \lvert\psi_2\rvert^2)$ No interference, even distribution of photons.
Q3	$\frac{N}{2}(\lvert\psi_1\rvert^2 + \lvert\psi_2\rvert^2)$ Still no interference pattern since photons give path info for each electron.
	Student B
Q2	$\frac{N}{2}(\lvert\psi_1\rvert^2 + \lvert\psi_2\rvert^2)$ There will be no interference pattern, the lamp photons give each atom which-path information when scattering.
Q3	$\frac{N}{2} \cdot [\lvert\psi_1(x)\rvert^2 + \lvert\psi_2(x)\rvert^2 + \lvert\psi_1(x)\rvert \cdot \lvert\psi_2(x)\rvert \cdot \cos\Delta\phi]$ There will be an interference pattern. If $\lambda_{photon} >$ slit width, the two slits are indistinguishable (unresolvable) from each other to the photon, so the photon cannot give which-path information upon scattering.

Table 3-IV Scores assigned for responses to Q2 and Q3 written by Student A and Student B (shown in Table 3-III) using the rubric (see Table 3-II), with commentary explaining the scores in italics.

	A	B
1. Mention that the photon wavelength is an important consideration in determining the pattern that forms on the screen. *Student A: Made no mention of wavelength and described the same pattern for both situations.* *Student B: Specifically mentioned wavelength.*	0	1
2. Correctly interpret the effect of wavelength on the interference pattern. (1 pt. for each question.) *Student A: Described the correct pattern for Q2 but not Q3.* *Student B: Described both patterns correctly.*	1	2
3. Find different number densities for the two questions. *Student A: Did not find different number densities for Q2 and Q3.* *Student B: Found two different number densities for Q2 and Q3.*	0	1
4. Number densities are correct. (1 pt. for each question.) *Student A: Wrote the correct number density for Q2 but not for Q3.* *Student B: Wrote the correct number densities for Q2 and Q3.*	1	2
5. Number densities are consistent with patterns. (1 pt. for each question.) *Student A: Number densities were both consistent with the patterns described.* *Student B: Number densities were both consistent with the patterns described.*	2	2
Total Score	4	8

Between 20%-30% of the data collected were independently rated by two different researchers using the rubric for all questions/question pairs, and the inter-rater reliability was excellent (greater than 90% agreement). As an example of how the rubric was applied, Table 3-

III includes examples of responses (transcribed) for Q2 and Q3 written by two students (referred to as Student A and Student B), and Table 3-IV shows how the rubric was applied to score the two students' responses for Q2 and Q3.

Average normalized gain [21] is commonly used to determine how much the students learned and takes into account their initial scores on the pretest. It is defined as

$$\langle g \rangle = \frac{\langle S_f \rangle - \langle S_i \rangle}{100\% - \langle S_i \rangle},$$

where $\langle S_f \rangle$ is the average percent score of the class on the post-test and $\langle S_i \rangle$ is the average percent score of the class on the pre-test [21]. We calculated the average normalized gains for both the upper-level undergraduate and graduate students using this equation.

3.6 RESULTS

In order to determine the extent to which the QuILT was effective in helping students develop a coherent understanding of these concepts and addressing issues discussed in Section 3.3 related to Learning Objectives 1-3, we compared students' performances on the pre-test and post-test and measured their improvement. Below, we discuss our findings.

3.6.1 Reasoning in Terms of "Which-path" Information

Question 1 was an open-ended question and asked students to describe a situation in which the introduction of a lamp would destroy the electron interference pattern on the screen and why that would be the case. Many students struggled with this question on the pre-test and provided a variety of responses. The student responses were categorized into six possible categories, as shown in Table 3-V. A student response can fall in more than one category, which is why the percentages do not necessarily add up to 100%.

Table 3-V Categorization of student responses to Q1 as a percent of total responses for undergraduate (U) and graduate (G) students on the pre- and post- test. (A) is correct and (B) is partially correct.

Q 1	A	B	C	D	E	F
U Pre	**9%**	13%	33%	20%	20%	9%
U Post	**91%**	80%	0%	0%	0%	0%
G Pre	**14%**	32%	36%	5%	14%	5%
G Post	**64%**	22%	9%	4%	16%	2%

The responses in Table 3-V are categorized as follows:

(A) Mention $\lambda < d$: A correct response mentioned that the wavelength of the lamp's photons should be shorter than the separation between the slits (e.g., with the reasoning that the WPI is known for the electrons in this case). The students in this category had demonstrated that they understood the role of photon wavelength in determining whether we have information about which slit the particle passed through to reach the screen. Credit was also given to students who described how scattering via a photon localizes the particles and alters their momenta.

(B) Mention "Which-path" Information: At least half credit was given to any students who mentioned that if WPI is known from the scattered photons, then the interference pattern vanishes even if they did not explicitly describe the connection between WPI and the wavelength of the lamp's photons. Learning Objective 1 of the DSE QuILT was that students learn to reason in terms of WPI in order to make predictions about the patterns that form on the screen. Any response that mentioned WPI (or used reasoning related to knowing which slit the particle went through to reach the screen) is counted in category B, even if the response was included in another category, which is why the rows of Table 3-V do not necessarily add up to 100%.

(C) Scattering: The most common response on the pre-test described any type of physical scattering of the electrons due to collisions with the photons destroying the interference pattern without mentioning the constraints on photon wavelength. For example, one student stated the following: *"If scattering occurs enough between the lamp photons & the particles, they will completely convolute the interference pattern so it will no longer be visible. The screen will simply appear completely lit up"* Another student stated: *"The interference pattern will be destroyed if the lamp has high enough intensity to scatter off the electrons."* The question specifically mentions that the photons scatter off the electrons, so the responses in this category were mostly restating the information provided in the question without providing any additional details about the scattering process and how it would impact the interference on the screen when the particles arrive there. The responses of students in this category do not provide any evidence that students understand the mechanisms involved in destroying the interference pattern in this situation.

(D) <u>Photon-electron Interference</u>: Several students (mostly undergraduates) described situations in which the wavelengths and phases of the photon and electron were aligned in such a way that the two would destructively interfere. For example, one student noted, *"for destructive interference to occur the phase (scattering angle) between the photon and the electron must be such that maxima of the photon's wavelength correspond to minima of the electron's wavelength and vice versa."* It is interesting that students are treating the incident particles and the photons from the lamp as "waves" that can interfere with each other and annihilate each other. Students with these types of responses are potentially invoking the principle of superposition as though the photon and electron are identical particles and the crest of one particle's wave will cancel the trough of the other particle's wave. This hypothesis is confirmed from interviews with students who invoked such a notion.

(E) Other Responses: Many responses in this category were too simplistic and did not fall into other categories. These students often claimed that whenever a lamp is present, the interference pattern on the screen will vanish (without mentioning anything about the scattering of the particles off the photons from the lamp). For example, one student stated, "*there will be an interference pattern when the light bulb is off. When the light bulb is on, there will not be interference.*"

(F) Incomplete or No Response: This category also includes those who wrote "I don't know." We note that all the students were given sufficient time to complete both the pre-test and the post-test and nearly all the students submitted their tests voluntarily. So if a student left a question blank, it is very likely that he/she did not know how to answer that question. Also, occurrences in which a particular question was left blank, but a subsequent question was answered were also fairly common, especially in the pre-test, thus indicating that students most likely did not know how to answer the questions they left blank.

Table 3-V shows that on the pre-test 9% of undergraduates and 14% of graduate students were able to correctly identify the photon wavelength condition for whether an interference pattern will form on the screen. On the post-test, 91% of undergraduates and 64% of graduate students received full credit for their responses. As shown in Table 3-V, 80% of undergraduate students explicitly used reasoning involving WPI to answer Q1 on the post-test, compared to 13% on the pre-test. These results demonstrate that the QuILT was effective in achieving Learning Objective 1 for a majority of students by addressing their initial difficulties with reasoning in terms of WPI. One possible explanation for the discrepancy between undergraduate and graduate students' post-test scores in this regard is that the graduate students may be less motivated to engage with the QuILT due to the fact that (unlike the undergraduates) the graduate

students were not graded for correctness on the post-test and this material was not part of their other exams since there was no letter grade in the TA training course. We note however, that these first year physics graduate students were also simultaneously enrolled in their first semester of a two semester core quantum mechanics course simultaneously although this material was not part of that course and that course was very traditional and did not focus on conceptual understanding of foundational concepts as in the QuILT.

Also, as shown in category D of Table 3-V, on the pre-test, about 20% of undergraduate students and 5% of graduate students described how the interference pattern on the screen will disappear if destructive interference occurs between the electrons and the photons from the lamp. None of the undergraduate students and only 4% of the graduate students used this reasoning on the post-test.

3.6.2 Difficulty Recognizing the Effect of Lamp Wavelength on the Interference Pattern

Student responses to Q2 and Q3 were considered together, as were Q2 and Q4, and Q3 and Q5. The responses for these pairs were divided into the following six categories:

(A) Patterns and number densities are both correct.

(B) Patterns are correct, but not the number densities.

(C) Patterns are different and incorrect.

(D) Patterns are the same and incorrect.

(E) Other responses.

(F) Incomplete or no response.

Student responses to Q2 and Q3 were scored together to determine the extent to which Learning Objective 2 was achieved and students understood what will happen in the experiment if the wavelength of the photons emitted by the lamp is altered. For Q2, the wavelength of the photon is significantly smaller than the distance between the two slits (which localizes the particles incident on the slits sufficiently and impacts the interference pattern), while for Q3, the wavelength is significantly larger than the distance between the two slits (so the localization due to scattering does not give WPI for the particles incident on the slits in this case and interference is observed on the screen). The breakdown of the student responses to this question pair is shown in Table 3-VI.

Table 3-VI Categorization of undergraduate (U) and graduate (G) student responses to Q2 and Q3 as a percent of total responses. Responses which received full credit are marked in bold, and responses which received at least partial credit are underlined.

Q 2,3	(A)	(B)	(C)	(D)	(E)	(F)
U Pre	<u>**2%**</u>	<u>30%</u>	26%	20%	0%	22%
U Post	<u>**91%**</u>	<u>9%</u>	0%	0%	0%	0%
G Pre	<u>**25%**</u>	<u>5%</u>	30%	20%	5%	16%
G Post	<u>**71%**</u>	<u>2%</u>	9%	13%	4%	0%

(A) <u>Patterns & Number Densities Correct</u>: Table 3-VI shows that graduate students were more likely than undergraduates to respond correctly to question pair 2-3 on the pre-test (25% versus 2%, respectively). On the post-test, however, 91% of undergraduates answered correctly compared to only 71% of the graduate students.

(B) <u>Only Patterns Correct</u>: Table 3-VI shows that about 30% of undergraduate students on the pre-test had a correct qualitative understanding of the role of photon wavelength in question pairs Q2-3 but did not know how to correctly represent the number densities in different situations (depending upon whether the interaction with the photons localized the particles sufficiently and there was WPI for the particles that arrived at the screen).

(C) Patterns Different, Incorrect: Students in this category understood (or correctly guessed) that changing the wavelength of the photons should change the pattern observed on the screen, but were not sure what that change should be.

(D) Patterns the Same, Incorrect: Table 3-VI shows that in the pre-test, 20% of undergraduate and graduate students did not realize that changing the photon wavelength from significantly smaller to significantly larger than the distance between the slits will alter the pattern observed on the screen. Interestingly, 13% of graduate students on the post-test maintained that the two patterns should be the same. They either did not think that changing the photon wavelength should affect the interference pattern, or did not make an effort to distinguish between the two situations.

(E) Other Responses: Some students, particularly graduate students, drew pictures that may or may not have represented interference patterns in the researchers' view, and a few of them wrote "Yes" or "No" for their responses without any elaboration. Since researchers did not understand what those responses meant even though there was an attempt to answer the questions, they were classified in this category.

(F) Incomplete or No Response: About 22% of undergraduates and 16% of graduate students did not fully respond on the pre-test, or simply wrote "I don't know."

3.6.3 Difficulty Recognizing the Effect of Lamp Intensity on the Interference Pattern

The pre-test responses to question pair Q2 and Q4 and question pair Q3 and Q5 were assessed using the same categories as for Q2 and Q3 in the previous subsection to investigate Learning Objective 3, which is to understand the role of lamp intensity in determining the interference pattern on the screen. The categorization of responses to Q2 and Q4 is shown in Table 3-VII.

The students whose responses were placed in category (D) either failed to recognize that the intensity of the lamp would affect the pattern on the screen or did not make an effort to distinguish between the two situations in Q2 and Q4. About 26% of undergraduate and 18% of graduate students claimed that the patterns on the screen would be the same for both of these questions on the pre-test.

Table 3-VII Categorization of undergraduate (U) and graduate (G) student responses to Q2 and Q4 as a percent of total responses. Responses which received full credit are marked in bold, and responses which received at least partial credit are underlined.

Q 2,4	(A)	(B)	(C)	(D)	(E)	(F)
U Pre	**9%**	20%	24%	26%	0%	22%
U Post	**88%**	7%	5%	0%	0%	0%
G Pre	**25%**	7%	23%	18%	5%	23%
G Post	**56%**	16%	9%	13%	4%	2%

As shown in category (D) of Table 3-VII, about 13% of graduate students on the post-test incorrectly maintained that the patterns should be the same in Q2 and Q4. None of the undergraduate responses manifest this mistake on the post-test, even though about one fourth of the undergraduate students had made this mistake on the pre-test. This type of dichotomy in the performance of the undergraduate and graduate students demonstrates that the QuILT was more effective in helping undergraduate students learn to account for lamp intensity than the graduate students.

Question pair Q3 and Q5 present a situation in which the intensity of the lamp is altered while the wavelength of the photons is significantly larger than the distance between the slits such that scattering between the photons and atoms (the incident particles) will not affect the pattern on the screen. Student responses to these questions were compared and categorized, as shown in Table 3-VIII. Correct responses are again in bold and partially correct are only underlined.

Table 3-VIII Categorization of undergraduate (U) and graduate (G) student responses to Q3 and Q5 as a percent of total responses. Responses which received full credit are marked in bold, and responses which received at least partial credit are underlined.

Q 3,5	(A)	(B)	(C)	(D)	(E)	(F)
U Pre	**4%**	24%	35%	9%	0%	28%
U Post	**80%**	14%	7%	0%	0%	0%
G Pre	**14%**	5%	34%	18%	5%	25%
G Post	**49%**	2%	40%	4%	4%	0%

In Table 3-VIII, responses in categories (A) and (B) indicate that many students understood or correctly guessed that the intensity of the lamp does not matter in this situation since the wavelength of the photons is not small enough to localize particles sufficiently to provide WPI. While about 94% of undergraduates recognized this fact on the post-test, only about 51% of the graduate students did so.

As shown in category (C) of Table 3-VIII, about one third of undergraduates on the pre-test did not realize that photons with wavelengths longer than the distance between the slits cannot alter the interference pattern, regardless of the intensity of the lamp. However, Table 3-VIII shows that the percentage of undergraduates whose responses fell in category (C) was significantly lower in post-test. Interestingly, the percentage of graduate students who made this mistake and thought that the patterns should be different in Q3 and Q5 on the post-test was actually slightly higher than the percentage on the pre-test. The persistence of this difficulty with question pair Q3 and Q5 especially among graduate students on the post-test illustrates a powerful phenomenological primitive, i.e., if you change something in the input, it should change *something* in the output [20]. However, in this case, changing the intensity of the lamp has no effect on the pattern. Prior research suggests that when students do not have a robust knowledge structure in a particular domain, it is common for students to use phenomenological primitives such as this [20] due to their prior conceptions. For example, Newton's 3rd law of motion is a difficult concept for introductory students, and in the context of a small car and a large truck colliding head-on, many students claim that the truck exerts a larger force on the car than the car exerts on the truck. This is often due to the phenomenological primitive that "bigger means more," and since the truck has the larger mass it must therefore exert a larger force. The students' difficulty with the role of lamp intensity is specifically addressed in the QuILT to help them reason that while in some cases changing the intensity may impact the interference pattern, in other cases it has no impact. The fact that only 7% of undergraduate students made this error in the post-test but a comparable number of graduate students used this primitive both on the pre-test and post-test suggests that many graduate students may not have engaged with the QuILT as effectively as the undergraduates.

3.7 OVERALL STUDENT PERFORMANCE ON PRE-TEST/POST-TEST

The average scores on the pre-/post-tests for the undergraduate and graduate students are shown in Fig. 3-5. We also calculate average normalized gains [21], *p*-values, and effect sizes in the form of Cohen's $\mathrm{d} = \frac{\mu_1 - \mu_2}{\sigma_{\text{pooled}}}$ (where μ_1 and μ_2 are the averages of the two groups being compared and $\sigma_{\text{pooled}} = \sqrt{\frac{{\sigma_1}^2 + {\sigma_2}^2}{2}}$, where σ_1 and σ_2 are the standard deviations of the two groups), using individual group means and standard deviations. While the graduate students on average performed significantly better than the undergraduate students on the pre-test (44% vs. 23%, respectively, $p = 0.005$, $\mathrm{d} = 0.43$), they performed significantly worse than the undergraduate students on the post-test (73% vs. 95%, respectively, $p < 0.001$, $\mathrm{d} = 0.67$). Undergraduate students' average normalized gains were near the ceiling ($g = 0.94$), while the graduate students' corresponding gains were much lower ($g = 0.51$).

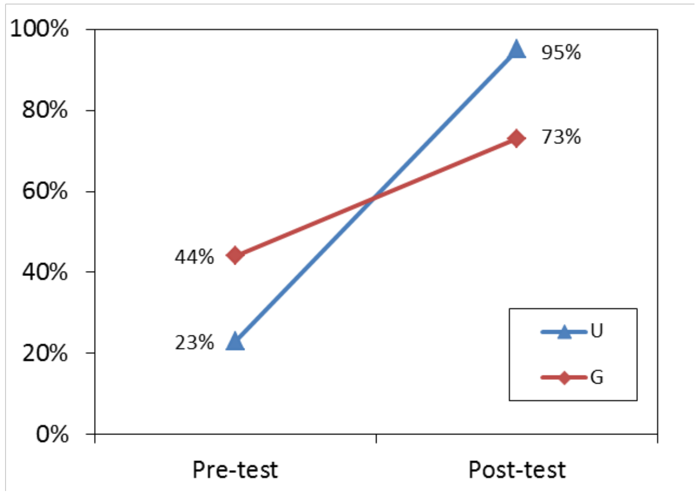


Figure 3-5 Average pre-test and post-test scores for undergraduate (U) and graduate (G) students.

Figure 3-6 shows the distribution of the pre-test and post-test scores for each of the 45 undergraduate students (represented by blue triangles) and 46 graduate students (represented by red diamonds). The solid diagonal line through the middle of the plot represents the same score on the pre-test and post-test, so that all data points located above that line represent students who performed better on the post-test than the pre-test. The dotted lines located above and below the solid line represent the range of post-test and pre-test scores that were within 20% of each other. While nearly half of the graduate students had scores within this range (20 out of 45 students), only three of the undergraduate students had post-test scores that were within 20% of their pretest scores. Moreover, those three undergraduate students already had pre-test scores that were greater than 70% to begin with.

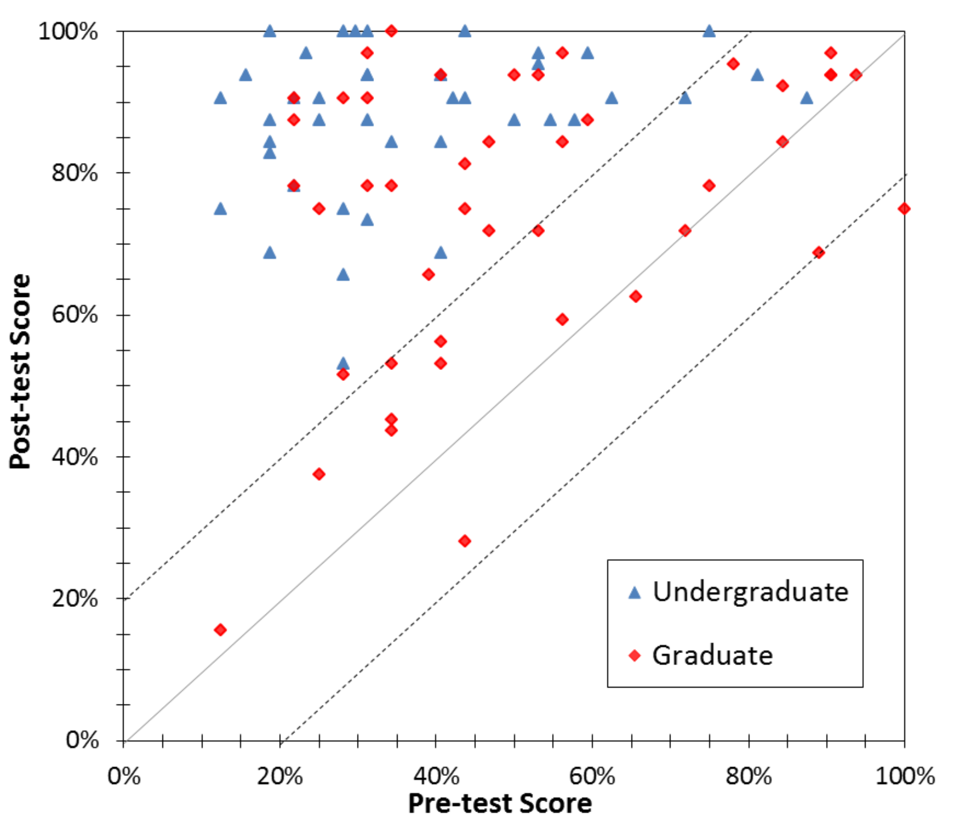


Figure 3-6 Individual student post-test scores versus pre-test scores for undergraduate and graduate students. The solid diagonal line represents the cutoff for students whose post-test scores were higher than their pre-test scores. The dotted diagonal lines located above and below the solid diagonal line indicate cutoffs for students whose post-test scores were within ±20% of the corresponding pre-test score.

Figure 3-7 shows a histogram of the individual normalized gains for the undergraduate and graduate students, with dashed lines representing the average normalized gains for each group. Most undergraduate students had normalized gains greater than 0.7, and only two of them had normalized gains below 0.4. However, those two students scored very high on both the pre-test and post-test. Compared to the undergraduate students, the graduate students had more variation in their normalized gains. For example, 13 of the graduate students had normalized gains of 0.4 or less, compared to only two of the undergraduate students. (Note: four graduate students with negative normalized gains are not included in the histogram.)

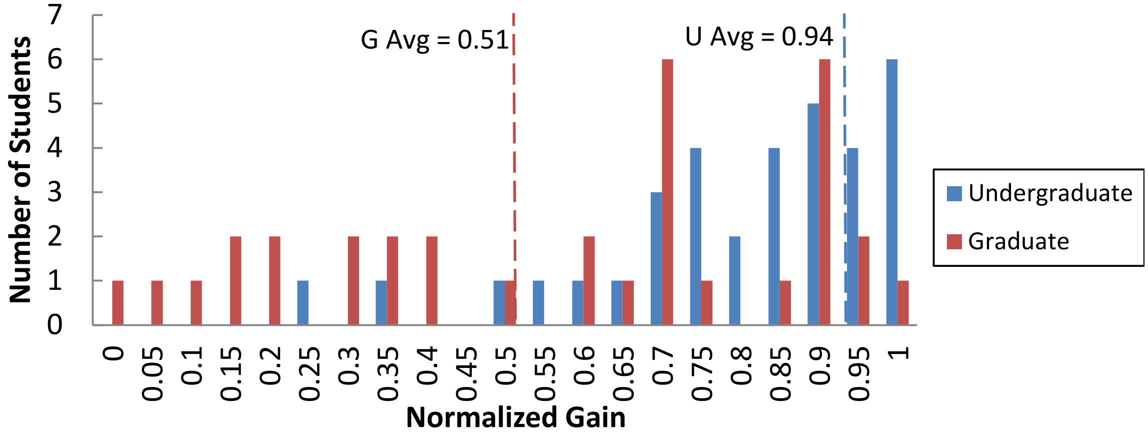


Figure 3-7 Histogram of individual normalized gains for undergraduate students (blue bars) and graduate students (red bars), with average undergraduate (U) and graduate (G) student normalized gains represented with blue and red dashed lines, respectively.

The average undergraduate and graduate student scores for questions Q1, Q2-3, and Q4-5 are shown in Table 3-IX, with *p*-values and effect sizes for various comparisons. Note that Q2 and Q3 were graded together according to the rubric described in Section 3.5, as were Q4 and Q5. On average, graduate students performed better than the undergraduates in the pre-test on all questions/question pairs, and the reverse was true for the comparison of the post-test scores for these two groups (as seen from the *p*-values and effect sizes d in the last two rows in Table 3-IX for each vertical comparison). A t-test comparison also indicated that the difference between the means of the pre-test and post-test for each question/question pair is significant for each group (undergraduate and graduate students) but the effect sizes are significantly higher for the undergraduate students (d = 2.29 for Q1, 2.78 for Q2-3 and 2.52 for Q4-5 for undergraduates). We note that in educational interventions large effects are considered to occur for Cohen's d of 0.8 or more; the effect sizes for undergraduate students in this study are three times larger than that.

Table 3-IX Average pre-test and post-test percentages on Q1, Q2-Q3, and Q4-Q5 for undergraduate (U) and graduate (G) students, with *p*-values and effect size Cohen's d for comparison of undergraduates and graduate students (the *p*-values and effect size are in the last two rows for each vertical comparison). Also listed are the *p*-values and effect sizes for the difference between the means of the pre-test and post-test for each question (or question pair) for each group.

	Q1				Q2-Q3				Q4-Q5			
	Pre	Post	*p*	d	Pre	Post	*p*	d	Pre	Post	*p*	d
U	16	94	< 0.001	2.29	34	97	< 0.001	2.78	19	95	< 0.001	2.52
G	47	68	0.016	0.37	49	83	< 0.001	0.71	35	69	< 0.001	0.71
p	< 0.001	< 0.001			0.018	0.005			0.023	< 0.001		
d	0.60	0.57			0.37	0.44			0.35	0.73		

The QuILT was administered to both groups (undergraduate and graduate students) over a short time frame (the pre-test and post-test for each group were separated by one week) without any additional in-class instructions on these topics. While there are other possible frameworks through which the differences between undergraduate and graduate student performances from pre-test to post-test may be interpreted, the impact of grade incentive is one of them. In particular, since other aspects of implementation were similar in both courses, one possible reason for the post-test score discrepancy is that, as noted earlier, the undergraduates had grade incentives to learn from the QuILT while the graduate students worked on the QuILT in a TA training course with no final exam on which these types of questions could show up and a pass/fail grading scheme. Some graduate students may have been less cognitively engaged in learning from the QuILT since it was graded only for completeness. We hypothesize that many students are not intrinsically motivated to learn even in advanced physics courses, and grade incentives for learning may provide the needed external motivation.

3.8 SUMMARY

We investigated student difficulties with quantum mechanics concepts pertaining to the double-slit experiment in various situations that appear to be counter-intuitive and contradict classical notions of particles and waves. We developed and carried out a preliminary evaluation of a research-validated QuILT which makes use of an interactive simulation to improve student understanding of the double-slit experiment and to help them develop a better grasp of foundational issues in quantum mechanics.

Preliminary data comparing the pre- and post-test scores of upper-level undergraduate and graduate students indicate that the DSE QuILT was effective in improving students' understanding of these concepts that defy classical intuition such as wave-particle duality, effect of quantum measurement on the wavefunction, and explanation of whether interference should be observed after a large number of single particles pass through the slits. The QuILT strives to help students develop a coherent understanding of foundational concepts in various situations involving the DSE and helps students reason about whether or not interference of single particles is observed at the screen in the DSE in various situations. For example, when the photons from the lamp scatter off the particles at the slits, many students initially had difficulty understanding the effects of wavelength of the photons and intensity of the lamp on the interference pattern at the screen formed by single particles incident on the slits. For example, about one-fifth of undergraduate students noted on the pre-test that the photon wave and electron wave would somehow destructively interfere with each other during the scattering if their wavelengths were comparable, but none of the undergraduate students used this reasoning in their responses on the post-test.

However, upper-level undergraduates outperformed physics graduate students in the post-test, although the reverse was true in the pre-test. One possible reason for this difference may be the level of engagement with the QuILT due to the grade incentive. In the undergraduate course the post-test was graded for correctness, while in the graduate course it was graded for completeness.

3.9 ACKNOWLEDGEMENTS

We thank the National Science Foundation for award PHY-1202909, and Klaus Muthsam for the simulation that was adapted for the QuILT [24]. We also are thankful to various members of the department of physics and astronomy at the University of Pittsburgh (especially R. P. Devaty and Emily Marshman) for helpful conversations during the development of the tutorial.

3.10 CHAPTER REFERENCES

1. R. Crease, The most beautiful experiment, Phys. World **15**, 19 (2002).
2. R. Feynman, R. Leighton and M. Sands, *The Feynman Lectures on Physics, Vol. III*, Chapter 1 (Addison-Wesley, Reading, MA, 1963).
3. J. Wheeler, in *Mathematical Foundations of Quantum Theory*, edited by A. Marlow (Academic Press, New York, 1979).
4. P. Jolly, D. Zollman, N. Rebello, and A. Dimitrova, Visualizing motion in potential wells, Am. J. Phys. **66**, 57 (1998).
5. 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, Physics Today **8**, 43 (2006); C. Singh, Assessing and improving student

understanding of quantum mechanics, *Proceedings of the 2005 Phys. Ed. Res. Conference, Salt Lake City, UT* (AIP Conf. Proc., Melville, NY, 2006), p. 69; Helping Students Learn Quantum Mechanics for Quantum Computing, *Proceedings of the 2006 Phys. Ed. Res. Conference, Syracuse, NY* (AIP Conf. Proc., Melville, NY, 2007), p. 42; Student difficulties with quantum mechanics formalism, *Proceedings of the 2006 Phys. Ed. Res. Conference, Syracuse, NY* (AIP Conf. Proc., Melville, NY, 2007), p. 185; S. Lin and C. Singh, Categorization of quantum mechanics problems by professors and students, Eur. J. Phys. **31**, 57 (2010); Assessing expertise in quantum mechanics using categorization task, *Proceedings of the 2008 Phys. Ed. Res. Conference, Ann Arbor, MI* (AIP Conf. Proc., Melville, NY, 2009), p. 185; A. Mason and C. Singh, Do advanced physics students learn from their mistakes without explicit intervention?, Am. J. Phys. **78**, 760 (2010); A. Mason and C. Singh, Reflection and self-monitoring in quantum mechanics, *Proceedings of the 2008 Phys. Ed. Res. Conference, Ann Arbor, MI* (AIP Conf. Proc., Melville, NY, 2009), p. 197; B. Brown, A. Mason, and C. Singh, Improving performance in quantum mechanics with explicit incentives to correct mistakes, Phys. Rev. PER **12**, 010121 (2016); B. Brown, C. Singh, and A. Mason, The effect of giving explicit incentives to correct mistakes on subsequent problem solving in quantum mechanics, *Proceedings of the 2015 Phys. Ed. Res. Conference, College Park, MA* (AIP Conf. Proc., Melville, NY, 2015), p. 67.

6. M. Wittmann, R. Steinberg, and E. Redish, Investigating student understanding of quantum physics: Spontaneous models of conductivity, Am. J. Phys. **70**, 218 (2002); J. Morgan and M. Wittmann, Examining the evolution of student ideas about quantum tunneling, *Proceedings of the 2005 Phys. Ed. Res. Conference, Salt Lake City, UT* (AIP Conf. Proc., Melville, NY, 2006), p. 73.

7. C. Manogue, and E. Gire, Representations for a spins first approach to quantum mechanics, *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE* (AIP Conf. Proc., Melville, NY, 2012), p. 55.

8. E. Marshman and C. Singh, Developing an interactive tutorial on a quantum eraser, *Proceedings of the 2013 Phys. Ed. Res. Conference, Portland, OR* (AIP Conf. Proc., Melville, NY, 2014), p. 175; C. Singh and E. Marshman, Developing an interactive tutorial on a Mach-Zehnder Interferometer with single photons, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN* (AIP Conf. Proc., Melville, NY, 2015), p. 239.

9. G. Passante, P. Emigh, and P. Shaffer, Student ability to distinguish between superposition states and mixed states in quantum mechanics, Phys. Rev. ST PER **11**, 020135 (2015).

10. 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 (2013).

11. C. Singh, Interactive learning tutorials on quantum mechanics, Am. J. Phys. **76**, 400 (2008); G. Zhu and C. Singh, Improving students' understanding of quantum mechanics via the Stern–Gerlach experiment, Am. J. Phys. **79**, 499 (2011).

12. G. Zhu and C. Singh, Improving students’ understanding of quantum measurement: I. Investigation of difficulties, Phys. Rev. ST PER **8**, 010117 (2012).

13. G. Zhu and C. Singh, Improving students’ understanding of quantum measurement: II. Development of research-based learning tools, Phys. Rev. ST PER **8**, 010118 (2012); Surveying students' understanding of quantum mechanics in one spatial dimension, Am. J. Phys. **80**, 252 (2012).

14. G. Zhu and C. Singh, Improving student understanding of addition of angular momentum in quantum mechanics, Phys. Rev. ST PER **9**, 010101 (2013).

15. A. Collins, J. Brown, and S. Newman, Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics, in *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, edited by L. Resnick (Lawrence Erlbaum, Hillsdale, NJ, 1989), pp. 453-494.

16. For example, see J. Docktor and J. Mestre, Synthesis of discipline-based education research in physics, Phys. Rev. ST PER **10**, 020119 (2014).

17. C. Singh, What can we learn from PER: Physics education research?, The Phys. Teacher **52**, 568 (2015).

18. D. Schwartz, J. Bransford, and D. Sears, Efficiency and innovation in transfer, in *Transfer of Learning From a Modern Multidisciplinary Perspective*, edited by J. Mestre (Information Age, 2005), pp. 1-51.

19. M. Chi, Thinking aloud, in *The Think Aloud Method: A Practical Guide to Modeling Cognitive Processes* (Academic Press, London, 1994).

20. A. diSessa, in *Constructivism in the Computer Age*, edited by G. Forman and P. Pufall (Lawrence Erlbaum, Hillsdale, NJ, 1988).

21. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66**, 64 (1998).

22. B. Brown and C. Singh, Development and evaluation of a quantum interactive learning tutorial on Larmor precession of spin, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN* (AIP Conf. Proc., Melville, NY, 2015), p. 47.

23. S. DeVore and C. Singh, Development of an interactive tutorial on quantum key distribution, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN* (AIP Conf. Proc., Melville, NY, 2014), p. 59.

24. The double-slit simulation was developed by Klaus Muthsam, muthsam@habmalnefrage.de.

3.11 APPENDIX B

This is the full text of questions Q1 through Q5 on the DSE QuILT pre-test and post-test (which were identical).

In questions 1-5, assume that particles are sent one at a time from the particle source. The figure below shows a double-slit experiment which was modified by adding a lamp (light bulb) between the double slit and the screen. The lamp is slightly off to the side so it does not block the slits. Assume that when the lamp is turned on, if scattering occurs between a particle used in the double-slit experiment and a photon from the lamp, this scattering occurs at the slits only. (An illustration of the double slit setup with the addition of a lamp is shown in Fig. 3-1 [24].)

- Assume that <u>ALL</u> the particles scattered by photons still reach the screen.
- Assume that a particle only scatters a single photon, i.e., multiple scattering is neglected.

1. Suppose you perform a double slit experiment with electrons while the lamp is turned off and observe an interference pattern on the screen. You then repeat the experiment with the lamp turned on (assume that the intensity of the lamp is such that every particle used in the experiment scatters off a photon).

(i) Describe a situation in which this addition of the lamp between the double slit and the screen destroys the interference pattern observed on the screen (in the situation you describe, assume that all particles reach the screen even if scattering occurs between the particles and the photons emitted by the lamp).

(ii) Explain your reasoning for your answer in 1(i).

Questions 2-5 refer to the following setup:

You perform a double-slit experiment using Na atoms and observe an interference pattern on the screen. You then change the experiment by adding a lamp as discussed earlier.

- If slit 2 is closed, the wavefunction of a Na atom that goes through slit 1 and arrives at a point *x* on the screen is $\Psi_1(x)$. If instead, slit 1 is closed, the wavefunction of a Na atom that goes through slit 2 and arrives at a point *x* on the screen is $\Psi_2(x)$.
- For this example, if slit 2 is closed, and a total number *N* of particles arrives at the screen, the number density of the particles at a point *x* on the screen is $N|\Psi_1(x)|^2$.
- For questions 2-5, both slits are open.

2. For (i) and (ii) below, suppose that the wavelength of the photons is significantly smaller than the distance between the slits and the intensity of the lamp is such that each Na atom scatters off a photon. Also, assume that all the scattered atoms still reach the screen.

   (i) Write down an expression for the number density of Na atoms at a point *x* on the screen in terms of $\Psi_1(x)$ and $\Psi_2(x)$ after a large number *N* of Na atoms arrive at the screen.

   (ii) Describe the pattern you expect to observe on the screen after a large number *N* of Na atoms have arrived at the screen. Explain your reasoning.

3. For (i) and (ii) below, suppose that the wavelength of the photons is significantly larger than the distance between the slits and the intensity of the lamp is such that each Na atom scatters off a photon. Also, assume that all scattered atoms still reach the screen.

   (i) Write down an expression for the number density of Na atoms at a point *x* on the screen in terms of $\Psi_1(x)$ and $\Psi_2(x)$ after a large number *N* of Na atoms arrive at the screen.

   (ii) Describe the pattern you expect to observe on the screen after a large number *N* of Na atoms have arrived at the screen. How, if at all, is this pattern different from the pattern in 2(ii)? Explain your reasoning.

4. For (i) and (ii) below, suppose that the wavelength of the photons is significantly smaller than the distance between the slits and the intensity of the lamp is such that about half of the Na atoms scatter off a photon. Also, both slits are open and all the atoms reach the screen, including the ones that scatter.

   (i) Write down an expression for the number density of Na atoms at a point x on the screen in terms of $\Psi_1(x)$ and $\Psi_2(x)$ after a large number N of Na atoms arrive at the screen.

   (ii) Describe the pattern you expect to observe on the screen after a large number N of Na atoms have arrived at the screen. How, if at all, is this pattern different from the pattern in 2(ii)? Explain your reasoning.

5. For (i) and (ii) below, suppose that the wavelength of the photons is significantly larger than the distance between the slits and the intensity of the lamp is such that about half of the Na atoms scatter off a photon. Also, both slits are open and all the atoms reach the screen, including the ones that scatter.

   (i) Write down an expression for the number density of Na atoms at a point x on the screen in terms of $\Psi_1(x)$ and $\Psi_2(x)$ after a large number N of Na atoms arrive at the screen.

   (ii) Describe the pattern you expect to observe on the screen after a large number N of Na atoms have arrived at the screen. How, if at all, is this pattern different from the pattern in 3(ii)? Explain your reasoning.

4.0 INVESTIGATING TRANSFER OF LEARNING IN ADVANCED QUANTUM MECHANICS

4.1 INTRODUCTION

Transfer of learning from one context to another context is a hallmark of expertise. Despite the beauty and simplicity of physics, it is particularly difficult for students to apply physics concepts from the contexts in which they learned them to new contexts. Learning theory suggests that transferring learning from one context to another context can be difficult especially if the "source" (from which transfer is intended) and the "target" (to which transfer is intended) do not share surface features. This difficulty arises because knowledge is encoded in memory with the context in which it was learned and solving the source problem does not automatically manifest its "deep" similarity with the target problem [1-4].

Transfer of learning between different contexts requires that students engage in problem solving in a deep meaningful way and use it as an opportunity for extending and organizing their knowledge structure. It is therefore not surprising that developing expertise in problem solving constitutes a major goal of most physics courses [5-12]. Problem solving can be defined as any purposeful activity in which one is presented with a novel situation and devises and performs a sequence of steps to achieve a set goal [15] in a limited amount of time. Both knowledge and experience are required to solve the problem efficiently and effectively. Genuine problem solving is not algorithmic, but rather it is heuristic. There are several stages involved in effective problem solving, including initial qualitative analysis, planning, assessment, and reflection upon the problem-solving process in addition to the implementation stage [16-20]. The problem solver must make judicious decisions in order to reach the goal in a reasonable amount of time. Given a problem, the range of potential solution trajectories that different people may follow to achieve the goal can be called the problem space [21]. For each problem, the problem space is very large (essentially infinite) and, based upon one's expertise, people may traverse very different paths in this space which can analogically be visualized as a maze-like structure [21,22].

Simon and Hayes defined two problems as isomorphic if they have the same structure in their problem space [23-25]. They were among the first to analyze why one problem in an isomorphic problem pair may be more difficult than the other using their model of problem solving [23-25]. Cognitive theory suggests that the context in which something is learned and the way it is stored in memory have important implications for whether cues in a problem statement will trigger a recall of the relevant concepts in order to be able to solve the problem successfully [26-29]. Depending upon the context, the problem space for the isomorphic problems may be such that one problem may trigger the recall of relevant concepts from memory while another problem may not. The famous “Tower of Hanoi problem” is isomorphic to the “cannibal and the missionary problem” [23-25, 26]. Research shows that the Tower of Hanoi problem in this pair is more difficult than the latter [25]. Despite the same underlying features of these problems, the problem solvers, in general, traverse very different trajectories in the problem space and use different knowledge resources while solving the two isomorphic problems [23-25].

The isomorphic problem pairs chosen by Simon and Hayes shared “deep” features but had very different surface features involving pegs and disks of varying radii in the Tower of

Hanoi problem, and cannibals, missionaries, river and boats in the other. Here, we will define problems to be isomorphic if they require the same physics principle to solve them. The similarity of the isomorphic problems can span a broad spectrum. Isomorphism between problems has been observed in studies about students' conceptions, e.g., in the context of changes of reference frames [30]. Very closely related isomorphic problems may include those in which the situation presented is the same but some parameters are varied, e.g., two similar projectile problems with different initial speed and/or angle of launch. One level of difficulty with regard to discerning their similarity can be introduced by changing the context of one of the problems slightly. For example, two isomorphic problems about projectiles can involve a person kicking a football or throwing stones from a cliff. Depending upon an individual's level of expertise, the person may or may not discern the similarity between these problems completely and be able to transfer his/her learning from one context to another. Another level of difficulty can be introduced, e.g., by making one problem in the isomorphic problem pair quantitative and one qualitative [31]. A high level of complexity can be introduced by making the surface features of the problems very different as in the problem pair chosen by Simon and Hayes or by introducing distracting features into one of the problems. These complexities can make the transfer of learning from one problem to another isomorphic problem more difficult.

Several studies have focused on investigating the differences between the problem-solving strategies employed by experts and novices in physics [32-38]. These studies suggest that a crucial difference between the problem solving capabilities of experts and beginners lies in both the level and complexity with which knowledge is represented and rules are applied. Expert knowledge can be thought to be organized hierarchically in pyramid-like knowledge structures where the most fundamental concepts are at the top of the hierarchy followed by the ancillary concepts [33]. Experts view physical situations at a much more abstract level than novices. Prior studies have often found that students, unlike physics experts, have difficulty in transferring learning appropriately from one isomorphic problem to another which has a different context but involves identical physics principles [31,36]. For example, experts in physics consider a problem involving angular speed of a spinning skater moving her arms close or far from her body isomorphic to a problem related to the change in angular speed of a neutron star collapsing under its own gravitational force. Rather than focusing on the “surface” features of the two problems: a spinning skater in one case and rotating neutron star in the other case, which appear very different, experts focus on “deep” features based upon abstract physics principles: the fact that there are no external torques on the relevant system in each case implies that angular momentum is conserved. For experts, angular momentum conservation immediately implies that both a spinning skater and slowly spinning neutron star would speed up when their moment of inertia decreases. On the other hand, introductory students may not discern the isomorphism and transfer their learning from the skater problem to correctly answer questions about the neutron star, even if the two problems are posed back to back as part of the same quiz [31].

This dichotomy in expert/novice problem solving and ability to transfer learning from one context to another may arise because novices focus on surface features, may get distracted by irrelevant details, and may not see the inherent similarity of the two problems. Two classic studies about problem categorization of introductory mechanics problems indicate that novices categorize problems according to the objects of the problems, regardless of the physical principles required for solving them [32,33]. For example, novices deemed problems similar if they involved inclined planes, or pulleys, or springs, as opposed to whether they could be solved by applying Newton's laws or conservation of energy. In contrast, physics experts categorize problems based on physics principles, not the problems' surface similarity [32,33]. Experts' knowledge representation and organization along with their superior problem-solving strategies help them narrow the problem space without cognitive overload and retrieve relevant knowledge efficiently from memory [39-42]. Although expertise studies usually classify individuals either as an expert or a novice, people's expertise in a particular domain spans a large spectrum in which novices and "adaptive" experts are at the two extremes [43].

Research on transfer involving analogical reasoning [44-50] can also be valuable for understanding how individuals with different levels of expertise transfer their learning from one context to another. Studies have shown that using analogy can improve students' learning and reasoning in many domains. [48, 51-54]. Although the surface similarity may help people recall the analogy better, understanding the underlying similarity at the "deep" level is important in order to apply the analogy to the new situation appropriately [46]. Research on learning from solved examples also sheds light on how students transfer their learning to solve new problems by first looking for similar problems that they already know how to solve and applying similar strategies from one problem to another [55-59].

Here, we investigate transfer of learning from one context to another in advanced quantum mechanics. Prior research suggests that in quantum mechanics, students have many common difficulties due to the unintuitive and abstract nature of the subject [60-78], and introductory and advanced students often show analogous patterns of reasoning difficulties [73]. Other investigations have focused on instructional approaches to help students learn quantum mechanics [75-90].

Research suggests that the ability to transfer learning improves with expertise because as individuals develop expertise, their knowledge is better organized and represented at a more abstract level in memory, which facilitates categorization and recognition based upon deep features [6,7,91-93]. Additionally, as students transition towards adaptive expertise in a particular domain, they also develop metacognitive skills [94-98], which are not constrained to that domain. In particular, once the level of expertise of an individual reaches a certain threshold, the individual may be able to exploit his or her metacognitive skills to transfer his or her learning to a new context even if those contexts only share deep similarity [94]. The ability to transfer learning from one context to another is closely related to metacognition because in order for transfer to occur, one must be able to recognize the deep features of a problem while engaged in problem solving. Therefore, it will be particularly useful to investigate the extent to which students in advanced quantum mechanics are able to transfer their learning from one context to another because these students are higher on the physics expertise spectrum than students in introductory physics and the majority of transfer studies in physics have focused on introductory physics students [81].

We begin by discussing the motivation for the research and describe the isomorphism between the MZI and the DSE contexts, after which we describe the methodology used and the research questions investigated. We then present the results, discuss some possible reasons for the observed transfer, and present results from a discussion with a subset of graduate students enrolled in a course for physics teaching assistants who participated in this study about why they thought they were able to transfer their learning from one context to another. We conclude with a summary of our findings. For those interested, we have included an in-depth analysis of the common student difficulties with the questions posed in this investigation in Appendix C.

4.2 MOTIVATION AND ISOMORPHISM BETWEEN MZI AND DSE

In this study, we first investigate the extent to which upper-level physics undergraduate students and physics graduate students are able to transfer their learning about the concept of "which-path" information (WPI) [99] from a research-based tutorial on the Mach-Zehnder interferometer (MZI) with single photons and polarizers in one or both paths [100] to answer questions about interference of single photons in the context of the double-slit experiment (DSE) [101]. The concept of WPI at a detector may be useful when the state of the system is a superposition of two different spatial path states (e.g., MZI, DSE with single photons). In general, when a detector can project both components of the path state, then WPI is unknown. On the other hand, when a detector can project only one component of the path state, then we have complete which-path information, i.e., WPI is known.

The DSE and MZI are experiments that can be used to illustrate fundamental principles of quantum mechanics using concrete contexts, and the underlying principles used to predict interference in both experiments are the same. In the MZI tutorial, students were guided to apply WPI reasoning to answer questions on various MZI setups. These students could use WPI reasoning to answer analogous questions in the DSE experiment, and we investigated the extent to which they were able to transfer learning about WPI from the MZI tutorial to answer DSE questions (more detail on the study design is presented in Section 4.3).

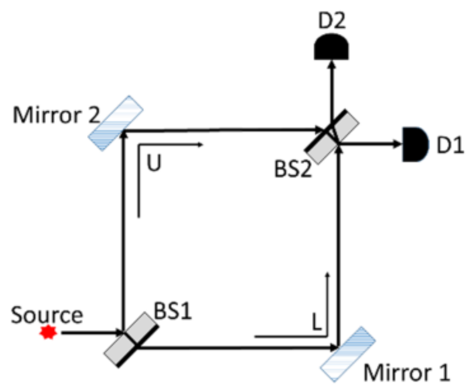


Figure 4-1 Basic Mach-Zehnder interferometer setup.

To understand the isomorphism between the MZI and DSE, we first consider the most basic MZI setup (shown in Fig. 4-1). BS1 and BS2 are beam splitters. BS1 is oriented such that it puts the single photon emitted from the source into an equal superposition of the upper (U) and lower (L) path states shown (which we represent as $|U\rangle$ and $|L\rangle$, respectively). Mirrors are for proper alignment, and BS2 ensures that the components of the single photon state from both the U and L paths can be projected into each (photo) detector D1 and D2 after BS2 so that constructive or destructive interference (or anything in between) can be observed at the two detectors D1 and D2 in Fig. 4-1 (depending on the path length difference between the U and L paths). If an *additional* detector is placed anywhere in the lower path L between BS1 and BS2, after encountering the detector, the superposition of the U and L path states of a photon collapses and if the photon does not get absorbed by the detector, the state of the photon inside the MZI is the upper path state $|U\rangle$. Conversely, if an additional detector is placed in the upper path U, after encountering the detector, if the photon is not absorbed by that detector, the state of the photon inside the MZI collapses to the lower path state $|L\rangle$. In these situations (additional detector in the U or L path of the MZI), if a photon arrives at the detector D1 or D2 after BS2, we have WPI because either detector can only project the component of the photon state along the U or L path and no interference is observed at D1 or D2. However, if no detector is placed in either the U or

L path of the MZI (as in Fig. 4-1), the state of a photon inside the MZI remains an equal superposition of the U and L path states, WPI is unknown (because the detectors can project both the $|U\rangle$ and $|L\rangle$ components of the photon state), and therefore interference is observed at D1 and D2.

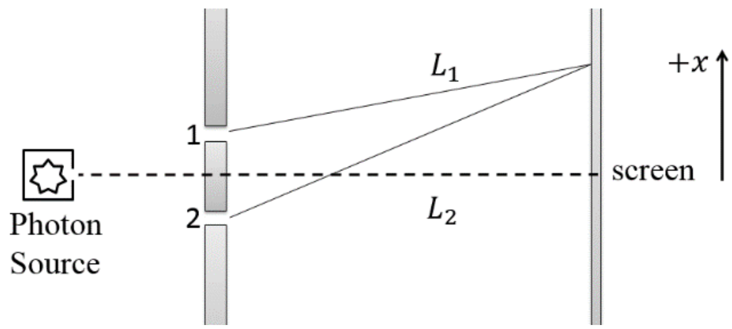


Figure 4-2 Basic double-slit experiment setup with single photons.

Now consider the DSE setup shown in Fig. 4-2, which consists of a photon source that sends photons one at a time towards a plate with two parallel slits (which we'll refer to as "slit 1" and "slit 2"). If slit 2 is blocked, the state of a photon inside the DSE (after passing through the slits) collapses to $|\Psi_1\rangle$, and if slit 1 is blocked, the state of a photon collapses to $|\Psi_2\rangle$. If one of the slits is blocked and a photon arrives at the screen in Fig. 4-2 (the screen is the detection device in the DSE equivalent to detectors D1 and D2 in the MZI), we have WPI because the screen can only project one component of the photon's path state (either $|\Psi_1\rangle$ or $|\Psi_2\rangle$) and, therefore, no interference is observed. If neither slit is blocked, the photon state remains an equal superposition of $|\Psi_1\rangle$ and $|\Psi_2\rangle$. In other words, $|U\rangle$ and $|L\rangle$ in the MZI are analogous to $|\Psi_1\rangle$ and $|\Psi_2\rangle$ in the DSE. In the situations in which there is no detector in either path of the MZI and neither slit is blocked for the DSE, we do not have WPI and each photon interferes with itself.

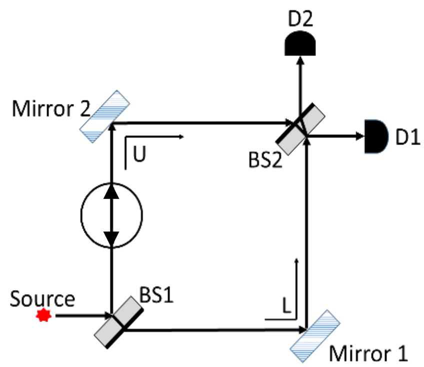


Figure 4-3 Mach-Zehnder interferometer setup with a vertical polarizer placed in the upper path.

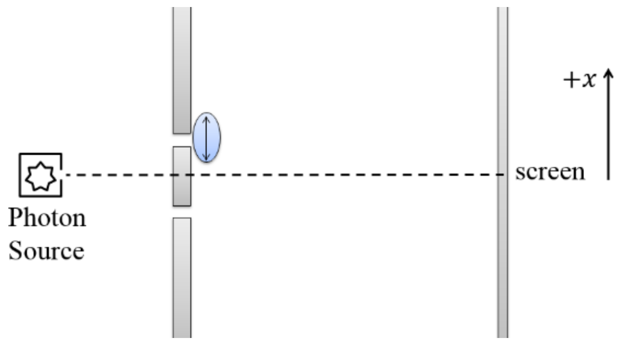


Figure 4-4 Double-slit experiment setup with a vertical polarizer placed after slit 1.

Now consider the situation shown in Fig. 4-3 in which we place a vertical polarizer in the upper path of the MZI and the source emits $+45°$ polarized single photons. This situation is analogous to the situation shown in Fig. 4-4 in the DSE in which a vertical polarizer is placed after slit 1 (and the source emits $+45°$ polarized single photons). We now must use a four-dimensional Hilbert space: two dimensions account for the allowed path or slit states ($\{|U\rangle, |L\rangle\}$ or $\{|\Psi_1\rangle, |\Psi_2\rangle\}$, respectively), and two dimensions account for polarization states, for which a convenient basis for the situations described in Fig. 4-3 and Fig. 4-4 is $\{|V\rangle, |H\rangle\}$ (vertical, horizontal polarization states, respectively). If a vertical polarizer is placed in the upper path of the MZI, the $|U\rangle$ state will be associated with only the vertical polarization state, $|U\rangle|V\rangle$, but the $|L\rangle$ state will be associated with both the vertical and horizontal polarization states, $|L\rangle|V\rangle + |L\rangle|H\rangle$. In both experiments we will assume that the detectors are sensitive to polarization (they are covered with polarizers with a particular orientation, e.g., vertical or horizontal), which means that the collapse of the photon state after it is measured by the detectors D1 or D2 provides information about the polarization of the photon. Therefore, in the situation depicted in Fig. 4-3, we have WPI for horizontally polarized photons arriving at D1 and D2 because the horizontal polarization is associated with the lower path state only—each detector can only project the $|L\rangle$ component of the state of a horizontally polarized photon. We do not have WPI for the vertically polarized photons because the vertical polarization is associated both with the upper and the lower path states—each detector can project both the $|L\rangle$ and $|U\rangle$ components of the state of a vertically polarized photon. The fact that we have WPI for horizontally polarized photons and we do not have WPI for vertically polarized photons implies that the photons that arrive at the detectors in the $|V\rangle$ polarization state interfere and those that arrive in the $|H\rangle$ state do not. The situation is analogous in the DSE (Fig. 4-4): If a vertical polarizer is placed after slit 1, horizontally polarized photons arriving at the screen will not interfere, while vertically polarized photons arriving at the screen will show interference.

Throughout this study, we will refer to questions focusing on interference in these types of situations ($+45°$ polarized single photons emitted by the source and polarizers of various orientations placed in one or both paths of the MZI or in front of one or both slits of the DSE) as "MZI polarizer questions" or "DSE polarizer questions" depending on the context in which they are asked (MZI or DSE).

It is important to note that while the DSE and MZI contexts are isomorphic, the "surface" features of these two experiments are rather different. In the MZI, the paths are restricted and the photons arrive at point detectors D1 and D2, while in the DSE the photons are delocalized in the space between the slits and the screen and can be detected anywhere on the extended screen. In addition, in the DSE there is no explicit optical element corresponding to BS2 in the MZI which mixes the components of the photon state from the two paths: The screen itself does the mixing of the components of the single photon state from the two slits before the projective quantum measurement of the photon at the screen. These differences suggest that the surface features of these problems are quite different, which can make it challenging for novices to recognize the isomorphism [1,31]. In order to recognize the isomorphism between the MZI and DSE, students must be able to reason about the deep features of the contexts and recognize the utility of the concept of WPI and its relation to whether or not interference will take place in both contexts. Thus, transfer of learning from the MZI to the DSE context is not guaranteed a priori even if students understand the underlying physics principles in the MZI context.

Also, it is worthwhile to keep in mind that while it is very likely that graduate students have some knowledge of the DSE, it is unlikely that more than a small percentage of them have been introduced to the concept of WPI and learned how to reason using WPI to answer questions similar to the ones discussed above. With regards to the MZI, very few graduate students are likely to have any knowledge other than perhaps the fact that interference is observed in this experiment (the pre-test results confirm this). The physics undergraduate students in this study were almost all nearly at the end of the undergraduate curriculum (more than 80% were seniors) and the physics graduate students were all in their first year. Therefore, for the purposes of this study, the two populations are not very different in terms of background knowledge on the DSE and MZI.

4.3 METHODOLOGY

4.3.1 Participants

The participants in this study were 46 undergraduate students enrolled in an upper-level quantum mechanics course and 59 physics graduate students enrolled in a mandatory semester-long TA professional development course which met once a week for two hours. For the undergraduate students, the MZI and DSE were part of the course material, the tutorials and post-tests (described in detail below) were graded for correctness, and the post-tests were counted as regular quizzes. In addition, the undergraduate students were aware that topics discussed in these tutorials may appear in future exams. For the graduate students, one of the topics of the TA professional development course was the benefits of using the tutorial approach to teaching physics. The TAs in the course were required to engage with two research-based tutorials on topics which they are expected to be somewhat familiar with but do not fully understand (MZI and DSE) as opposed to engaging with tutorials on introductory physics topics for which many graduate students are likely to be experts (although there was brief discussion of the introductory physics tutorials in the class). If graduate students engage with tutorials on topics they do not fully understand, they can learn the topics discussed and understand the value of utilizing these tools as supplements to instruction. For the graduate students, the pre-/post-tests and the tutorials were graded for completeness instead of correctness since the course performance was graded as satisfactory or unsatisfactory. The DSE and MZI polarizer questions were part of the DSE and MZI pre-/post-tests, respectively.

4.3.2 Materials

The materials used in this study are research-based Quantum Interactive Learning Tutorials (or "QuILTs") involving either the MZI or DSE. Both the MZI and the DSE QuILTs include pre-tests and post-tests. The DSE pre-/post-tests also included the DSE polarizer questions (a topic which was not covered in the DSE tutorial). These polarizer questions were designed specifically for the transfer study reported here and will be discussed in detail later in this section. Both the MZI and DSE tutorials focus on helping students learn about topics in quantum mechanics such as wave-particle duality (in the context of single photons in the MZI and in the context of particles with mass in the DSE), self-interference of a single photon (MZI) or particle with mass (DSE), the probabilistic nature of quantum measurements, and collapse of a quantum state upon measurement. Both tutorials make use of interactive simulations in which students can manipulate the MZI or DSE setups to predict and observe what happens at the photo-detectors (MZI) or screen (DSE) for various setups.

The development of both tutorials included think-aloud interviews with both graduate and undergraduate students in which students worked on the tutorials while articulating their thought processes. Students were not disturbed while they worked on the tutorials, though after they had finished they were asked for clarification on points they had not made clear earlier while thinking out loud. Approximately 85 hours of individual think-aloud interviews were conducted with students while developing the DSE tutorial, and similar interviews were conducted while developing the MZI tutorial as well. In addition, five physics faculty members were consulted several times during the development of each of these tutorials to ensure that the wording of the questions was unambiguous and that the topics covered in the tutorials were addressed appropriately and unambiguously.

In the MZI tutorial, students learn how photo-detectors and optical elements such as beam-splitters in the path of the MZI with single photons affect the measurement outcomes. In addition, the MZI tutorial discusses setups in which polarizers of various orientations are placed in one or both paths. It guides students to reason in terms of WPI to predict the outcome at the detectors. Thus, the MZI tutorial provides explicit help for answering the MZI polarizer questions which describe situations that are isomorphic to situations in the DSE polarizer questions. We hypothesized that if students learn how to reason in terms of WPI to answer the MZI polarizer questions, they may be able to transfer their learning about this reasoning to correctly answer the DSE polarizer questions if they realize that the two contexts are isomorphic. An investigation of the extent to which this transfer occurs was one of the main goals of our study. (More details on the study design are provided in Section 4.3.3.)

In the DSE tutorial, students learn the basics of single particle interference in the context of the DSE and how different parameters (e.g., mass and kinetic energy of the particles, widths and separation distance of the two slits, etc.) affect the interference pattern observed on the screen. In addition, students learn how placing a monochromatic lamp between the slits and the screen which emits photons that scatter with the particles sent through the slits can alter, and in some situations destroy, the interference pattern on the screen. The reasoning used to help students make sense of the photon-particle scattering in the DSE is based on WPI. However, this WPI based reasoning in the context of the DSE is for a completely different task and in a very different context than in the context of the MZI with single photons. We hypothesized that students learning about WPI and its connection to interference (even though the learning was in a different context) may recognize the isomorphism between the DSE and MZI if they had developed a deep functional understanding of the underlying physics and reasoning in terms of

WPI in the first context. It is important to emphasize that students do not learn about interference of single photons or polarizers in front of slits at all in the DSE tutorial. Thus, the DSE tutorial provides no explicit support for answering the DSE polarizer questions included in the DSE pre-test and post-test involving situations in which single photons are emitted by a monochromatic lamp and polarizers of various orientations are placed after one or both slits.

Each of the MZI and DSE pre-/post-tests included some questions focused on transfer of learning. The DSE pre-/post-test had two major parts:

(1) Questions related to the impact on the interference pattern of single particles (with mass) due to the addition of a monochromatic lamp close to the slits so that single particles passing through the slits scatter off photons emitted by the lamp. We refer to these questions as the "DSE lamp questions." These questions were explicitly discussed in the DSE tutorial which helped students make sense of them by using reasoning related to WPI.

(2) DSE polarizer questions related to interference of single photons passing through the slits and the effect on the interference pattern of placing polarizers of various orientations after one or both slits. These topics were not discussed in the DSE tutorial.

The DSE polarizer questions are summarized as follows:

"You perform a DSE in which photons that are polarized at $+45°$ are sent one at a time towards the double slit. The wavelength of the photons is comparable to the slit width and the separation between the slits is more than twice the slit width. In all questions, assume that the same large number N of photons reaches the screen. In each situation, describe the pattern you expect to observe on the screen. Explain your reasoning.

1. The situation described above.
2. A vertical polarizer is placed in front of one slit.

3. A vertical polarizer is placed in front of both slits.

4. A vertical/horizontal polarizer is placed in front of slit 1/slit 2, respectively.

5. A vertical/horizontal polarizer is placed in front of slit 1/slit 2, respectively. Additionally, a polarizer which makes an angle of +45° with the horizontal is placed after both slits, between the slits and the screen."

Although these types of questions involving single photons and polarizers were not included in the DSE tutorial, these questions are analogous to those students considered in the context of the MZI. For example, situation 2 above is analogous to the MZI setup shown in Fig. 4-3. We also note that the DSE polarizer questions above were part of a larger quiz (pre-/post-test) about the DSE which had other questions related to the DSE with single particles with mass. In particular, in the DSE pre-/post-tests, students answered 13 more questions in addition to the five DSE polarizer questions outlined above. The MZI pre-/post-tests were comprised of the MZI polarizer questions and many other questions in other situations (e.g., effect of removing BS2 on interference at the detectors D1 or D2, the percentages of photons of a given polarization arriving at D1 and D2 in different situations, etc.) which are very different from the polarizer questions.

4.3.3 Research Questions and Study Design

We hypothesized that at least some students who learned how to reason about the MZI polarizer questions in terms of WPI from the MZI tutorial may be able to transfer their learning and correctly reason about the DSE polarizer questions even though the DSE tutorial did not discuss situations involving single photons and polarizers at all. The extent to which this occurs without an explicit instructional intervention designed to help them recognize the isomorphism is the focus of the first research question.

RQ1. To what extent are graduate and undergraduate students able to transfer their learning about WPI from the context of the MZI to the context of the DSE without an instructional intervention designed to help them make the connection between the different contexts? This research question was investigated using a study design with two parts, A and B, as described below.

RQ1.A: The graduate students and upper-level undergraduates in two different years who participated in this study formed Cohort 1 and did the following in the given order:

1) Answered the MZI polarizer questions as part of the MZI pre-test.

2) Worked on the MZI tutorial and learned how to reason in terms of WPI to answer the MZI polarizer questions.

3) Answered the MZI polarizer questions as part of the MZI post-test.

4) Answered the DSE polarizer questions which were not discussed in the DSE tutorial as part of the DSE pre-test.

5) Worked on the DSE tutorial and learned how to reason in terms of WPI in a context other than single photons and polarizers (recall that single photons and polarizers were not included in the DSE tutorial at all).

6) Answered the DSE polarizer questions which were not discussed in the DSE tutorial as part of the DSE post-test.

We emphasize that the DSE tutorial does not include anything about single photons and polarizers, but the MZI and DSE polarizer questions are analogous. Therefore, if students transfer WPI learning from the MZI tutorial to the context of the DSE, they would exhibit improved performance on the DSE polarizer questions in the DSE pre-test compared to the MZI polarizer questions in the MZI pre-test. We therefore compared students' performance on these two sets of questions. We note that when answering the DSE polarizer questions in the pre-test, students had only been exposed to the concept of WPI in the context of the MZI.

We also investigated how often students explicitly used WPI reasoning to answer the DSE polarizer questions in the DSE pre-test. When students took the DSE pre-test, they had not learned how to reason using WPI to determine whether an interference pattern is observed on the screen in the DSE. However, students who worked on the MZI tutorial learned how to reason using WPI to determine whether interference is observed at detectors D1 and D2 in the MZI. Therefore, if these students use reasoning related to WPI to motivate their answer to the DSE polarizer questions, they are likely transferring their learning about WPI reasoning from the MZI context to the DSE context.

RQ1.B: As mentioned earlier, the DSE pre-test had two major parts: DSE lamp questions which are very different from the questions that the MZI tutorial was designed to help students with and the DSE polarizer questions which are very similar to questions discussed in the MZI tutorial (but that are not mentioned in the DSE tutorial). Therefore, if students are transferring their learning about WPI from the MZI context to the DSE context, when we compare the performance of students who have had the opportunity to work on the MZI tutorial and learn about WPI before answering the DSE pre-test with the performance of students who have not had this opportunity, we should observe improved performance on the DSE polarizer questions for those who used the MZI tutorial. For the DSE lamp questions for which the MZI tutorial was not designed to help, we should observe no difference in performance. In order to investigate if this is indeed the case, we switched the order of the MZI and DSE tutorials for a subsequent cohort of students – Cohort 2, but all the materials were kept exactly the same. The students in Cohort 2 did the following in the order given:

1) Answered the DSE polarizer questions as part of the DSE pre-test.

2) Worked on the DSE tutorial.

3) Answered the DSE polarizer questions as part of the DSE post-test.

4) Answered the MZI polarizer questions as part of the MZI pre-test.

5) Worked on the MZI tutorial and learned how to reason in terms of WPI to answer the MZI polarizer questions.

6) Answered the MZI polarizer questions as part of the MZI post-test.

We then compared the performance on the DSE polarizer questions and the DSE lamp questions on the pre-test of these two cohorts of students: Cohort 1 which engaged with the MZI tutorial first before the DSE pre-test, and Cohort 2 which did not engage with the MZI tutorial before the DSE pre-test. The study design to investigate RQ1 is summarized in Fig. 4-5.

Due to lack of participation from the faculty member teaching undergraduate quantum mechanics, only graduate students participated as Cohort 2 in the investigation of RQ1.B and thus we only performed this data analysis with graduate students. However, as we will discuss later in this study, the undergraduate students generally appeared to be more motivated than the graduate students to learn from the tutorials (as evidenced by their larger learning gains from the pre-test to post-test compared to the graduate students). This dichotomy between graduate and undergraduate students may at least partly be due to grade incentives: The post-tests were counted as regular quizzes and students were aware that questions related to the DSE and MZI can appear on the exams (and some of them they did) since the DSE and MZI were part of the course material. (As noted earlier, the graduate students had no grade incentive to perform well in the TA professional development course.) Therefore, if our data suggest that transfer of learning does occur for the graduate students based upon comparing the pre-test performance on the polarizer questions of Cohorts 1 and 2 (only Cohort 1 worked on the MZI tutorial before taking the DSE pre-test), had the investigation been carried out with the undergraduate students (consisting of advanced students in a quantum mechanics course who had almost finished the entire physics undergraduate curriculum), we would likely also observe transfer from the MZI to the DSE context, perhaps even more pronounced transfer when compared to the graduate students. For clarity, for the remainder of this paper, we refer to Cohort 1 as the "MZI→DSE cohort" and Cohort 2 as the "DSE→MZI cohort."

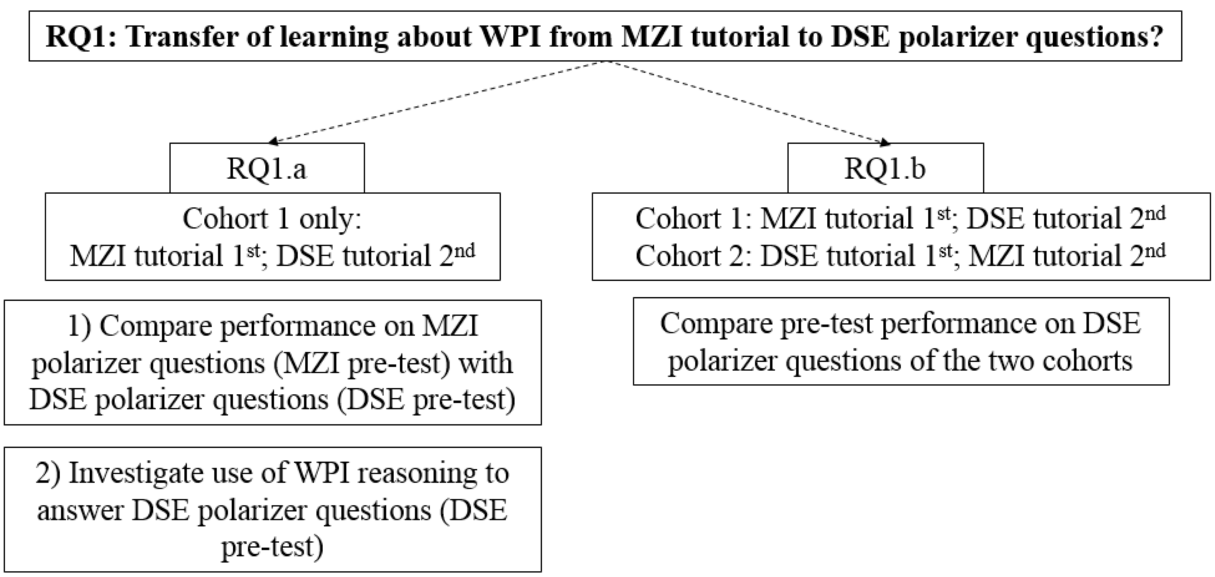


Figure 4-5 Schematic description of the research design to investigate RQ1. The MZI pre-test was given immediately before the MZI tutorial and the MZI post-test was given immediately after the MZI tutorial. Likewise for the DSE pre- and post-tests.

Before moving on to RQ2, we note the following: Students in both cohorts completed the DSE pre-test, worked on the DSE tutorial, then completed the DSE post-test. The DSE tutorial did ***not*** discuss interference of single photons with polarizers in front of the slits at all. However, the DSE tutorial focused on helping students learn about interference on the distant screen when the source emits single particles with mass and a monochromatic lamp is placed between the slits and the screen so that the photons from the lamp may scatter off the particles sent through the slits. Students learn how to determine whether the scattering process can provide WPI for the particles and erase the interference pattern on the screen. Thus, via the DSE tutorial, students are exposed to the concept of WPI in the DSE setup, but in a context that is quite different from the context of the DSE polarizer questions (in which single photons are used and polarizers of various orientations are placed after one or both slits). It is possible that this exposure to the concept of WPI in the same DSE setup but in a different context may also result in transfer of WPI reasoning and lead to increased performance on the DSE polarizer questions after students work on the DSE tutorial (in the post-test) compared to before working on the DSE tutorial (in the pre-test). Investigating the extent to which this occurs is the focus of the second research question:

RQ2. To what extent are students able to transfer their learning about WPI from one context of the DSE (single particles and a monochromatic lamp placed between the slits and the screen) to a different context of the DSE (single photons and polarizers placed in front of one or both slits) without an instructional intervention designed to help them make the connection between these different contexts in which WPI reasoning can be used?

To investigate RQ2 we compared students' performance on the DSE polarizer questions in the pre-test (before they worked on the DSE tutorial) with their performance in the post-test (after working on the DSE tutorial). This was done both for the students who worked on the DSE tutorial after working on the MZI tutorial (undergraduate and graduate students from Cohort 1) and the students who worked on the DSE tutorial before working on the MZI tutorial (graduate students from Cohort 2).

We also compared the percentages of students who used WPI reasoning to justify their responses to the DSE polarizer questions from before to after working on the DSE tutorial (i.e., from the pre-test to the post-test). If students transfer their learning about WPI reasoning from one context of the DSE (single particles with mass and a monochromatic lamp placed between slits and screen, which they learned about in the DSE tutorial) to another context of the DSE involving a polarizer (even though, as noted earlier, polarizers are never discussed in the DSE tutorial), they may be more likely to use WPI reasoning in the polarizer context after working on the DSE tutorial in the DSE post-test as opposed to in the DSE pre-test. The design of this investigation to answer RQ2 is summarized in Fig. 4-6.

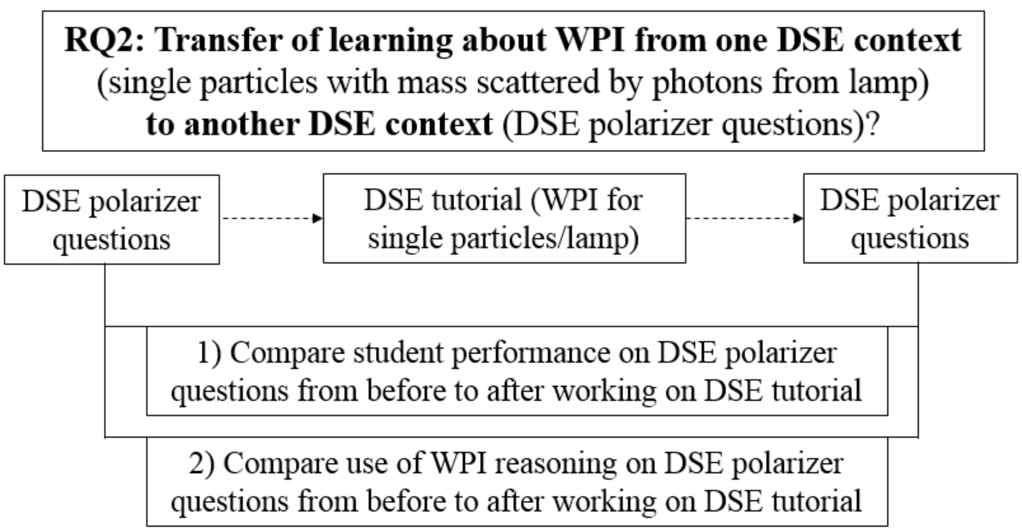


Figure 4-6 Schematic description of the design used to investigate RQ2 which involved the same setup (DSE) but different tasks for the same cohort of students.

4.4 TRANSFER RESULTS

4.4.1 RQ1. Transfer of Learning About "Which-path" Information From MZI Tutorial to DSE Polarizer Questions

4.4.1.1 RQ1.A: MZI tutorial first, DSE tutorial second.

(1.) Comparison of performance on MZI polarizer questions in MZI pre-test to DSE polarizer questions in DSE pre-test: Table 4-I shows undergraduate and graduate students' average performance on the MZI and DSE polarizer questions described in Section 4.3.2 for the MZI and DSE pre-tests. Situations 1 and 3 in Table 4-I were not posed to students in this form in the context of the MZI (i.e., these questions on the MZI pre- and post-tests did not ask about the interference pattern, but rather asked about the fraction of photons that would be detected by D1 and D2, which are somewhat different questions although closely related to the presence or absence of interference) so responses to those in the MZI context are not included (the corresponding cells in Table 4-I are shaded gray). The *p*-values for the analysis of variance (ANOVA) [102] comparing mean student performance on the analogous questions in the two contexts show that students performed significantly better on the DSE polarizer questions in the pre-test than on the MZI polarizer questions in the pre-test. On average, undergraduate and graduate students' performance on these questions is statistically significantly higher (by 45% and 34%, respectively) in the DSE context than in the MZI context. This level of transfer of learning is quite significant and, to our knowledge, has never been observed in empirical studies in physics for contexts whose surface features are so different.

Table 4-I Average performance (in percentage) of undergraduate (US) and graduate students (GS) on questions related to the effect of polarizers on the interference pattern in the MZI and DSE contexts in the pre-test, with *p*-values for comparison of the performance in the two contexts.

		1. No Polarizers	2. One Polarizer	3. Two parallel polarizers	4. Two Orthogonal Polarizers	5. Quantum Eraser
US	MZI		16%		27%	20%
	DSE	91%	50%	71%	81%	67%
	p		0.004		<0.001	<0.001
GS	MZI		24%		42%	41%
	DSE	88%	46%	78%	81%	71%
	p		0.029		<0.001	0.001

The data in Table 4-I include only the students who responded to the questions. However, students were given more than enough time to complete each pre-test, and nearly all students handed in their pre-tests voluntarily. In a subsequent section we will provide statistics for how many students did not provide a response on each of these questions. While the performance of the graduate students is comparable to that of the undergraduate students on the MZI and DSE polarizer questions both on the pre-test and on the post-test, on all of these questions, undergraduate students show more transfer than the graduate students by roughly 11%. It is possible that the lack of grade incentive for the graduate students is partly responsible for their lower level of transfer. In general, the graduate students appear to have learned less from these tutorials than the undergraduate students as evidenced by their lower normalized gains [103].

(2.) <u>Use of WPI reasoning to answer DSE polarizer questions on DSE pre-test</u>: Table 4-II shows, for DSE polarizer questions 2-5, the percentage of both undergraduate and graduate students who reasoned using WPI out of the students who provided any reasoning for their answers. These students all worked on the MZI tutorial before answering the DSE polarizer questions. Table 4-II shows that students who worked on the MZI tutorial before answering the DSE polarizer questions on the pre-test often used WPI reasoning (which they learned in the context of the MZI) to answer these questions, especially on the last two questions. In addition, a majority of the students who used reasoning related to WPI used it correctly to answer the DSE polarizer questions, thus indicating appropriate transfer from the MZI context to the DSE context. In contrast, students seldom used such reasoning on the MZI pre-test. Also, among the graduate students in the DSE→MZI cohort, only one used WPI reasoning in only one DSE polarizer question (question 4) on the DSE pre-test which indicates that most graduate students are unlikely to know how to reason using WPI to answer the DSE polarizer questions. For the undergraduate students, use of WPI reasoning would be even less likely since they are less likely to have exposure to the WPI concept before the course. This suggests that students' use of WPI reasoning on the DSE polarizer questions in the DSE pre-test was primarily due to the transfer of learning about the WPI concept from the MZI tutorial to the DSE context. We provide examples of correct and incorrect reasoning in Appendix C, where we discuss student difficulties on the DSE polarizer questions.

Table 4-II Percentage of undergraduate (US) and graduate students (GS) who used reasoning related to WPI out of those who provided reasoning on DSE polarizer questions 2-5 (Q2-Q5) in the pre-test.

	Q2	Q3	Q4	Q5
US	37%	37%	57%	62%
GS	33%	20%	60%	44%

4.4.1.2 RQ1.B: Comparing two cohorts of students: (1) Students who worked on the MZI tutorial before answering DSE polarizer questions. (2) Students who did not work on the MZI tutorial before answering DSE polarizer questions.

Table 4-III shows the performance of the two graduate student cohorts on the DSE polarizer questions (which probed single photon and polarizer contexts analogous to some MZI questions that were explicitly addressed in the MZI tutorial) and on the DSE lamp questions (which did not

have analogous MZI questions). Table 4-III shows that the MZI→DSE cohort (two years, $N = 45$) in which students worked on the MZI tutorial before the DSE pre-test significantly outperformed the DSE→MZI cohort (1 year, $N = 14$) only on the DSE polarizer questions (65% compared to 38%, $p = 0.006$, Cohen's d $= 0.831$), thus providing strong evidence for transfer of learning of WPI reasoning from the MZI to the DSE context. On the other hand, the two cohorts exhibit identical performance on the DSE lamp questions (42%). This suggests that some students (from the MZI→DSE cohort) who had worked on the MZI tutorial before the DSE pre-test were able to apply what they learned from the MZI tutorial to correctly answer the DSE polarizer questions in the DSE pre-test.

Table 4-III Average performances (Avg.) and standard deviations (Std. Dev.) of two graduate student cohorts (depending on the order in which they worked on the MZI and DSE tutorials) on the DSE transfer questions and on the DSE lamp questions in the DSE pre-test, with *p*-values and Cohen's d effect sizes to compare the performance of the two different cohorts.

	MZI→DSE Cohort		DSE→MZI Cohort			
	Avg.	Std. dev.	Avg.	Std. dev.	p	d
DSE Polarizer Questions	65%	13%	38%	26%	0.006	0.831
DSE Lamp Questions	42%	34%	42%	27%	0.955	0.016

Table 4-IV shows the percentage of students from the MZI→DSE cohort who answered each of the MZI polarizer questions and the DSE polarizer questions correctly (first and third row) and the performance of the graduate students from the DSE→MZI cohort on the DSE polarizer questions on the pre-test. Although the numbers are too small to perform meaningful chi-square tests to compare these percentages (only 14 students in the DSE→MZI cohort, a few students did not provide answers to all the questions, hence they were excluded from the analysis), we note the following:

1) Students appear to show comparable performance on the MZI and DSE polarizer questions in the pre-test when they have had no opportunity to learn from a tutorial. One may expect this to be the case given that the questions are analogous and students need to apply similar reasoning to answer them (though question 4 is an exception – more on that below).

2) Students who worked on the MZI tutorial before answering the DSE polarizer questions tend to perform better on the DSE polarizer questions than the students who did not work on the MZI tutorial before answering the DSE polarizer questions (once again, question 4 is an exception).

Our data therefore suggest that some of the graduate students were able to apply what they learned in the MZI tutorial, in particular, reasoning related to WPI, to answer the DSE polarizer questions in the DSE pre-test correctly.

Table 4-IV Graduate student performance on 1) the MZI transfer questions on the pre-test (before working on the MZI tutorial), 2) the DSE transfer questions on the pre-test (before working on the DSE tutorial) for the cohort which did not work on the MZI tutorial beforehand, and 3) the DSE transfer questions on the pre-test for the cohort which did work on the MZI tutorial beforehand.

		1. No Polarizers	2. One Polarizer	3. Two Parallel Polarizers	4. Two Orthogonal Polarizers	5. Quantum Eraser
1) MZI Transfer Questions			24%		42%	41%
DSE Transfer Questions	2) DSE→MZI Cohort	77%	8%	58%	80%	50%
	3) MZI→DSE Cohort	88%	46%	78%	81%	71%

Question 4, which deals with two orthogonal polarizers, seems to be an exception to the trends mentioned above. It appears that graduate students are better able to predict that no interference pattern forms on the distant screen in this situation in the DSE context than in the MZI context. One may wonder why this question is special, and it may be that students are answering it correctly in the DSE context for the wrong reason. Interviews conducted with graduate students shed some light on this issue and suggest that they sometimes reasoned that interference is observed at the screen because two different photons which go through different slits recombine at the screen and can interfere with each other (this is despite the fact that the questions make it clear that a source which emits single photons one at a time is used). Thus, a graduate student could reason that one photon emerging from one slit is horizontally polarized and another photon emerging from the other slit is vertically polarized, and since the polarizations are orthogonal, these photons do not interfere.

4.4.2 RQ2. Comparison of Performance and Use of WPI Reasoning on the DSE Polarizer Questions Before and After Working on the DSE Tutorial

(1.) <u>Comparison of the performance on the DSE polarizer questions before and after working on the DSE tutorial</u>: Table 4-V shows undergraduate and graduate students' performance on the DSE polarizer questions both before and after working on the DSE tutorial (these students had worked on the MZI tutorial before the DSE pre-test). As shown in Table 4-V, undergraduate students' performance improved significantly from the pre-test to the post-test for three of the five questions, indicating that the undergraduate students benefited from the DSE tutorial with regards to the DSE polarizer questions. This improvement may seem surprising because the DSE tutorial did not address any of the situations in the DSE polarizer questions at all and did not even mention interference of photons in the DSE. We discuss some possible reasons for this improvement in great detail in Section 4.5.2. In contrast, graduate students' performance is similar in the post-test to the pre-test, thus indicating that graduate students gained little from the DSE tutorial with regards to the DSE polarizer questions. This was true for both the graduate students who worked on the MZI tutorial before the DSE tutorial and the ones who worked on it after, so their results were combined in the data shown in Table 4-V. As noted earlier, at least one possible reason is that, unlike the undergraduate students, the graduate students were not given a grade incentive to engage with the tutorial, so they may not have engaged with it as deeply as the undergraduates did.

Table 4-V Percentages of undergraduate (US) and graduate students (GS) who answered the DSE transfer questions correctly before and after working on the DSE tutorial, with *p*-values for comparison.

	1. No Polarizers	2. One Polarizer	3. Two Parallel Polarizers	4. Two Orthogonal Polarizers	5. Quantum Eraser
US-Before Tutorial	88%	50%	73%	79%	66%
US-After Tutorial	98%	70%	95%	100%	93%
p-value	0.304*	0.066	0.008*	0.003*	0.004
GS-Before Tutorial	85%	37%	73%	80%	67%
GS-After Tutorial	86%	52%	86%	81%	76%
p-value	0.901	0.120	0.096	0.905	0.279

* Fisher's exact test was used instead of the chi-square test due to one or more expected cell frequencies being less than 5 [102].

(2.) <u>Comparison of the usage of WPI reasoning to answer the DSE polarizer questions from before to after working on the DSE tutorial</u>: In order to answer DSE polarizer questions 2-5 correctly, students were very likely to use reasoning related to WPI because they did not learn any other way of answering them. Table 4-VI shows, for DSE questions 2-5, the percentages of both undergraduate and graduate students who provided reasoning related to WPI among those who provided any reasoning for their answers both before and after working on the DSE tutorial (similarly to the previous table, all the graduate students are included in these data). The *p*-values for ANOVA listed in Table 4-VI show that both undergraduate and graduate students were statistically significantly more likely to provide reasoning related to WPI after working on the DSE tutorial on three out of the four questions. Given that the vast majority of students who used WPI reasoning used it correctly, it appears that increased usage of WPI reasoning may be responsible for the improvement observed in Table 4-VI for undergraduate students. Also, the graduate students were statistically more likely to use WPI reasoning on all questions except for question 2 after working on the DSE tutorial, and thus they too may have learned some things from the DSE tutorial, even if this learning did not necessarily result in significantly improved performance similar to the undergraduates.

Table 4-VI Percentages of undergraduate (US) and graduate (GS) students who used WPI reasoning among those who provided reasoning on DSE transfer questions 2-5, with *p*-values for comparisons.

	2. One Polarizer	3. Two Parallel Polarizers	4. Two Orthogonal Polarizers	5. Quantum Eraser
US-Before Tutorial	37%	37%	57%	62%
US-After Tutorial	88%	52%	87%	88%
p-value	<0.001	0.118	0.012	0.045
GS-Before Tutorial	27%	14%	47%	31%
GS-After Tutorial	48%	52%	77%	70%
p-value	0.155	0.020	0.049	0.018

4.5 DISCUSSION

4.5.1 Possible Reasons for Transfer from MZI to DSE Context

While it is difficult to identify the exact causes of the substantial transfer of learning from the MZI to the DSE context observed in this investigation, we hypothesize that the following may play a role:

1) Upper-level undergraduate and graduate students have developed sufficient abstract reasoning skills which allow them to recognize the isomorphism between these situations and the usefulness of reasoning about WPI in both contexts. This is supported by the finding that many students provide WPI reasoning for the DSE polarizer questions. Reasoning in terms of WPI to

answer questions related to interference of single photons was only discussed in the context of the MZI and students had to recognize the similarity between the MZI and DSE contexts in terms of underlying physics principles in order to answer the DSE polarizer questions.

It is possible that both of these groups of students have developed sufficient metacognitive skills to transfer their learning between these contexts whose surface features were sufficiently different. As discussed by Schraw [94], if the metacognitive skills developed by a learner in a particular domain reach a certain threshold, these skills become more readily transferable to a new domain. Furthermore, if the domains share similar characteristics, transfer of metacognitive skills is more likely to occur. In this study, the domains are not different but the contexts are sufficiently different. We can therefore interpret the transfer results found in this investigation to be partly due to many advanced students' ability to utilize metacognitive skills to transfer their learning from one context to another.

2) While the isomorphism between the MZI and DSE is in underlying physics and the contexts are different, both use single photons and polarizers of various orientations placed after one or both slits or paths. This type of similarity may have prompted students to utilize analogous reasoning when answering the DSE polarizer questions. We note, however, that in the MZI post-test, student average scores were near the ceiling (~ 90%), while the averages on the DSE polarizer questions were around 70% for both undergraduates and graduate students, implying that the transfer from the MZI to the DSE context is not perfect. In addition, the questions on both the MZI and DSE discussed here were part of longer pre-/post-test/quizzes on these experiments which asked about other situations and included other types of questions.

3) After students worked on the MZI tutorial and took the related post-test, the DSE polarizer questions were given in the following class. This proximity in timing may make it more

likely for students to be able to discern the similarity between the two contexts and transfer their learning from the MZI context to the DSE context. However, as mentioned earlier, in introductory physics, even if two questions which require use of the same underlying physics principles are asked back to back as part of the same quiz, a majority of students may not discern the similarity between the questions and therefore answer them using different reasoning [31].

4.5.2 Possible Reasons for Transfer of WPI Reasoning from DSE Tutorial to DSE Polarizer Questions

As evidenced in Tables 4-V and 4-VI, while the graduate students did not exhibit improved performance in predicting whether interference is observed in the DSE polarizer questions after working on the DSE tutorial, they were more likely to make use of WPI reasoning to motivate their answers (and most students who used WPI reasoning, did so correctly). It is possible that their intuition about whether or not interference is observed in the DSE was fairly good and going through the DSE tutorial helped some of them understand how to reason correctly.

On the other hand, undergraduate students (who were more motivated to learn from the tutorials mainly due to grade incentives) performed significantly better and were more likely to use WPI reasoning on the majority of the DSE polarizer questions after working on the DSE tutorial. Since the DSE tutorial was not designed to address student difficulties with interference of single photons in the DSE (it did not even mention anything about single photons and polarizers in front of the slits) and only focused on single particles with mass such as electrons or sodium atoms, this improvement in student performance on DSE transfer questions may seem surprising. However, the DSE tutorial did guide students through the concept of WPI and how it can be used to determine whether interference is observed in the DSE with single particles when a monochromatic lamp which emits photons that scatter with the particles (with mass) is placed between the slits and the screen. In some of these situations, scattering between the particles emitted by the source and the photons emitted by the lamp can provide WPI for the particles and destroy the interference pattern. It is possible that students who engaged with the DSE tutorial deeply can transfer their learning and recognize on their own how this type of WPI reasoning can be applied to answer the DSE polarizer questions.

To test this hypothesis we conducted think-aloud interviews with students who had completed the study of Modern Physics 1, which typically discusses the DSE. In an interview, students answered the DSE pre-test questions, worked on the DSE tutorial, and then answered the DSE post-test questions while thinking aloud. These students had not worked on the MZI tutorial so there was no possibility of transfer of the WPI concept and its relation to interference from the MZI context to the DSE context. Students were not disturbed during the interviews except when they became quiet for a long time, in which case the interviewer prompted the student to keep talking. After working on each part (e.g., pre-test), students were asked for clarification on points they had not made clear earlier while thinking aloud.

The interviews suggested that the DSE tutorial helped students reason using WPI to determine the pattern observed on the screen for a given DSE setup. In many cases, they were able to transfer this reasoning correctly to the DSE polarizer questions. For example, one interviewed student, Andrew, when answering DSE polarizer question 3 (a vertical polarizer placed in front of each slit) before completing the DSE tutorial, noted that a full interference pattern will form. However, he was not sure why. When the interviewer probed further (after the student had answered all pre-test questions) it appeared that the student was primarily guessing on this question and he did not have a very good reason for his answer. On the other hand, after working the DSE tutorial, when answering the same question he said: "*There will be interference. If the photon is vertical [vertically polarized], there is no which path knowledge, so there is interference. If [the photon is] horizontal, it doesn't go through*."

Thus, Andrew reasoned correctly using the concept of WPI, which was discussed in the DSE tutorial in completely different situations which involve placing a monochromatic lamp between the slits and the screen for a DSE with single particles with mass instead of using single photons and placing polarizers of various orientations in front of one or both slits (transfer questions). After working on the DSE tutorial, Andrew used WPI reasoning to answer the other DSE polarizer questions, and for the most part, used this reasoning correctly. For example, on DSE polarizer question 4 (two orthogonal polarizers) he recognized that WPI is known for all photons and therefore no interference is observed on the screen.

John, another interviewed student, while working on DSE polarizer question 4 before completing the DSE tutorial, understood that the vertically polarized photons will go through one slit and the horizontally polarized photons will go through the other. However, he thought that both will create an interference pattern. (Andrew's answer on the pre-test was very similar.) He stated: "*So there are two cases to consider: one where there's […] a horizontal photon coming in and the other is when there's a vertical photon coming in. So if it's a horizontal photon coming in, it only goes through the right one [slit with horizontal polarizer] and you get an [interference] pattern, and if the vertical one [photon] comes in, it only goes through the left one and you get an [interference] pattern. I don't know if those patterns are going to overlap […] If they overlap you'd just get a normal [interference] pattern, but if they don't overlap, you'd get a continuum [random background]*."

Discussions suggest that initially John thought that both the horizontally and the vertically polarized photons will create an interference pattern, and depending on where the two patterns form, they can either overlap perfectly, or are offset by a half of a wavelength so that the highs of one pattern overlap over the lows of the other pattern to produce an overall random distribution.

On the other hand, after working on the DSE tutorial, John correctly reasoned that both a horizontally and a vertically polarized photon goes through only one slit, and therefore neither interferes with itself because WPI is known. In all the questions with polarizers, he reasoned by thinking about WPI, which is a concept he learned in the DSE tutorial in a different context. Interestingly, when reading the first DSE polarizer question in the post-test, he stated, "*Hmm… So I don't think this was in the tutorial, but I assume something in the tutorial should help me answer these [questions]*." It appeared that he was able to use what he learned about how gaining WPI affects the pattern observed on the screen to reason about the DSE polarizer questions. It is possible that similar reasoning applies to other students like John who improved on the DSE polarizer questions after working on the DSE tutorial, which did not discuss the setups in the DSE polarizer questions.

It is important to keep in mind that these students only worked on the DSE tutorial and were not exposed to the MZI tutorial at all. Apparently they were all able to make connections between what they learned in the DSE tutorial, in particular how to reason in terms of WPI to determine whether an interference pattern is formed, to answer the DSE polarizer questions. It is possible that if they had also worked on the MZI tutorial earlier, they would have been able to transfer their learning from that context and make some connections between the type of WPI reasoning used in the MZI context and similar reasoning used in the DSE context. In that case, working on both tutorials is likely to consolidate their knowledge of WPI further and can lead to even better performance on the post-test, similar to the undergraduate students for whom both tutorials were a part of their course.

4.5.3 Class Discussion with Graduate Students About Transfer of WPI Reasoning

For the DSE→MZI cohort of graduate students (who worked on the DSE tutorial and the corresponding pre-test and post-test before the MZI tutorial and the corresponding pre-test and post-test), the instructor had a class discussion about students' ability to transfer their learning about WPI from the DSE tutorial in one context (particles with mass scattered by photons from a lamp) to answer the DSE polarizer questions (which were not mentioned in the DSE tutorial at all) over a one hour period during the TA professional development class. It was integrated in the class as a pedagogical discussion about transfer of learning. These graduate students performed quite well on the DSE polarizer questions on the post-test compared to the DSE pre-test (i.e., many appropriately transferred WPI learning from the DSE tutorial from a different context of particles with mass scattered by photons from a lamp to the new DSE polarizer context), but not so well on the MZI polarizer questions on the pre-test which were administered immediately after the DSE post-test. The discussion focused on the reasons students thought were responsible for 1) the effective transfer of WPI reasoning from one context in DSE to another DSE context involving polarizers, and 2) the difficulty of transferring from the DSE context to the MZI context as evidenced by worse performance on the polarizer questions on the MZI pre-test (which was administered immediately after the DSE post-test) compared to the DSE post-test. The following themes emerged:

1) Regarding the good performance on the DSE polarizer questions (a topic that was not touched upon in the DSE tutorial) in the DSE post-test, many graduate students noted that familiarity with both the DSE and polarizers in different contexts helped them make connections with the concept of WPI learned in a different context of particles with mass being scattered by photons from a lamp and interference. In particular, several students noted that learning about the WPI concept and its connection with whether one should observe interference in another context (DSE lamp questions involving particles with mass scattered by photons emitted by a lamp) helped them realize that this concept should be used when answering questions regarding whether interference should be observed in the DSE context involving polarizers. In other words, the concept of WPI became so integral to how they answered questions about interference observed on a distant screen in the DSE that even though the DSE polarizer questions were not mentioned in the DSE tutorial, the fact that these questions asked whether one would observe an interference pattern in different contexts with polarizers prompted them to use their knowledge of polarizers, double slit and WPI to conclude that vertical and horizontal polarizers in front of the two slits will give WPI (and destroy the interference pattern) and an additional 45° polarizer may erase it (and restore the interference pattern).

2) Many graduate students noted that the reason they did not perform as well on the polarizer questions in the MZI pre-test compared to the DSE post-test (even though the MZI pre-test was administered immediately after the DSE post-test) is that the MZI context is not nearly as familiar to them as the DSE context. They noted that transferring what they learned from the DSE context to the MZI context was therefore not as easy in this new context of the MZI. Some of them noted that they realized that the polarizers must have some effect but they were not sure how it would influence the interference pattern in this unfamiliar context of the MZI.

3) Discussions suggest that the graduate students were quite proud of the fact that they were able to transfer their learning about WPI and polarizers in other contexts to answer the DSE polarizer questions correctly in the DSE post-test. There was a discussion contrasting their ability to transfer their learning in this situation with the struggles of introductory students, whose learning may be significantly more context dependent resulting in significant difficulties in the transfer of learning from one context to another (e.g., transferring learning that angular momentum conservation for a spinning ballerina putting her arms closer to her body implies that she will start spinning faster to a spinning neutron star collapsing on itself is quite difficult for them [31]). They mentioned that unlike introductory students, the fact that they have learned so many different physics concepts in so many contexts makes them more likely to be able to put disparate pieces of information together (polarizers, DSE, WPI) to reason about new, less familiar situations. The discussion with the graduate students appears to corroborate our discussion in Section 4.1 regarding development of metacognitive skills assisting transfer of learning. Once the level of expertise surpasses a certain threshold (which graduate students and many advanced undergraduate students may have reached), when they learn in a new context, they may be able to use their metacognitive skills to transfer prior learning to the new context better than introductory physics students [94].

These discussions with the graduate students support the possible reasons we discussed earlier, and also highlight the fact that transfer from an unfamiliar context (MZI) to a familiar context (DSE) may be more facile than transfer in the opposite direction. While the number of graduate students in the DSE→MZI cohort was small (14), we did find that they were not able to transfer their learning about WPI from the DSE to the MZI context as well as they did from the MZI to the DSE context. This suggests that even advanced students who have developed superior metacognitive skills can encounter significant difficulties in transferring their learning to an unfamiliar context. In this study, both the undergraduate and graduate students had some familiarity with the DSE context and most understood the effect of polarizers in classical physics even if they didn't have experience reasoning about the novel context of using polarizers with various orientations in the DSE context. This may have facilitated transfer observed in this investigation. On the other hand, while they did have familiarity with polarizers, despite their good metacognitive skills, the context of the MZI which was significantly less familiar to students may have hindered transfer of learning from the DSE context to the MZI context compared to transfer from the MZI context to the DSE context. In other words, developing good metacognitive skills may be a necessary but not sufficient condition for transfer of learning to occur – the familiarity with the contexts may also play an important role.

4.6 SUMMARY

In this study, we find evidence that many upper-level undergraduate students and graduate students can transfer learning reasonably well from a tutorial on the MZI to an isomorphic context in the DSE without an explicit intervention to aid them in this regard. The MZI tutorial introduced students to the concept of WPI and guided them to use this concept to reason about whether or not interference is observed at the detectors in a particular MZI setup. When the DSE polarizer questions were administered, students who had worked on the MZI tutorial first performed significantly better on the DSE polarizer questions on the pre-test (average above 70%) than on the analogous MZI pre-test questions (average ~35%). Additionally, the graduate students, who worked on the MZI tutorial before answering the DSE polarizer questions on the pre-test, performed significantly better on these questions than the graduate students who did not work on the MZI tutorial. These two cohorts of graduate students showed identical performance on the other DSE questions which did not have analogous situations discussed in the MZI tutorial, which suggests that the improved performance on the DSE polarizer questions is likely due to transfer of learning rather than a difference in population. Another indication of transfer is that students often explicitly used reasoning learned in the context of the MZI to answer the DSE transfer questions, e.g., they used reasoning related to WPI, and most students who used this type of reasoning did so correctly, indicating appropriate transfer from the MZI to the DSE context. Finally, students sometimes explicitly drew the parallel between the DSE and the MZI contexts themselves without any prompting.

It is possible that the observed transfer is partly due to the fact that advanced undergraduate and graduate students have developed sufficient abstract reasoning skills which allow them to recognize the isomorphism between the two contexts. This would in turn make it likely that they are able to apply analogical reasoning between the two contexts. Given that many students (both undergraduate and graduate) reasoned in terms of WPI to answer the DSE polarizer questions (pre-test) and that this reasoning was only discussed in the MZI context it appears that students may be recognizing underlying similarity of the physics. Additionally, the familiar context of the DSE and students' knowledge of polarizers may contribute to their ability to transfer their learning about WPI to answer the DSE polarizer questions. Discussions with graduate students suggested that they generally agreed with these two possible reasons. Other possible reasons include the close temporal proximity of the MZI tutorial to the DSE polarizer questions and the fact that both the MZI and DSE questions relate to single photons and polarizers placed after various paths/slits. However, as noted earlier, introductory students often have difficulty discerning the similarity between isomorphic problems even if they are placed back to back [31]. In addition, the differences between the setups suggest that the surface features of these problems are quite different which can make it challenging to recognize the isomorphism between the MZI and the DSE [104]. Therefore, it is encouraging that advanced students have developed sufficient reasoning skills to be able to transfer their learning at least in the context discussed.

Furthermore, we found that after working on the DSE tutorial, undergraduate students improved significantly on the DSE polarizer questions despite the fact that the DSE tutorial did not mention anything similar to the polarizer situations in the transfer questions. Interviews with students who worked only on the DSE tutorial suggested that this improved performance is partly due to students correctly transferring learning of relevant concepts from the DSE tutorial, in particular, WPI, to correctly reason about the situations described in the transfer questions. It is likely that students who work on the MZI tutorial before working on the DSE tutorial and engage with both tutorials well (e.g., the undergraduates who worked on both tutorials as part of their quantum mechanics course) consolidate their knowledge of WPI further by making connections between the DSE and MZI contexts.

4.7 ACKNOWLEDGEMENTS

We thank the National Science Foundation for awards PHY-1505460 and PHY-1202909, Albert Huber for creating the simulation used in the MZI tutorial, and Klaus Muthsam for creating the simulation used in the DSE tutorial. We also thank F. Reif, R. D. Devaty, and E. Marshman for helpful discussions about this project.

4.8 CHAPTER REFERENCES

1. M. Gick and K. Holyoak, The cognitive basis of knowledge transfer, in *Transfer of Learning: Contemporary Research and Applications*, edited by Cornier and Hagman (Academic Press, New York, NY, 1987), pp. 9-46.

2. M. Gick and K. Holyoak, Schema induction and analogical transfer, Cog. Psych. **15**, 1 (1983).

3. K. Holyoak, The pragmatics of analogical transfer, in *The Psychology of Learning and Motivation*, edited by G. Bower (Academic Press, New York, NY 1985), pp. 59-87.

4. K. Holyoak and P. Thagard, *Mental Leaps: Analogy in Creative Thought* (MIT Press, Cambridge, MA, 1995).

5. A. Newell, *Unified Theories of Cognition* (Harvard University Press, Cambridge, MA, 1990).

6. F. Reif, Teaching problem solving: A scientific approach, Phys. Teach. **33**, 310, (1981).

7. J. Larkin and F. Reif, Understanding and teaching problem solving in physics, European Journal of Science Education **1**, 191 (1979).

8. K. Harper, R. Freuler, and J. Demel, Cultivating problem solving skills via a new problem categorization scheme, *Proceedings of the 2006 Phys. Ed. Res. Conference, Syracuse, NY*, edited by L. McCullough, P. Heron, and L. Hsu (AIP Conf. Proc., Melville, NY, 2007), p. 42.

9. M. Scott, T. Stelzer, and G. Gladding, Explicit reflection in an introductory physics course, *Proceedings of the 2006 Phys. Ed. Res. Conference, Greensboro, NC*, edited by L. Hsu, C. Henderson, and L. McCullough (AIP Conf. Proc., Melville, NY, 2007), p. 188.

10. Q. Ryan, E. Frodermann, K. Heller, L. Hsu, and A. Mason, Computer problem-solving coaches for introductory physics: Design and usability studies, Phys. Rev. ST PER **12**, 010105 (2016).

11. J. Touger, R. Dufresne, W. Gerace, P. Hardiman, and J. Mestre, How novice physics students deal with explanations, Int. J. Sci. Educ. **17**, 255 (1995).

12. J. Mestre, R. Dufresne, W. Gerace, P. Hardiman, J. Touger, Promoting skilled problem solving behavior among beginning physics students, J. Res. Sci. Teach. **30**, 303 (1993).

13. A. Van Heuvelen, Learning to think like a physicist: A review of research based instructional strategies, Am. J. Phys. **59**, 891 (1991).

14. A. Van Heuvelen, Overview: Case study physics, Am. J. Phys. **59**, 898 (1991).

15. M. Chi, R. Glaser, and E. Rees, Expertise in problem solving, in *Advances in the Psychology of Human Intelligence*, Vol. 1, edited by R. Sternberg (Lawrence Erlbaum, Hillsdale, NJ, 1982), pp. 7-75.

16. J. Heller and F. Reif, Prescribing effective human problem-solving processes: Problem description in physics, Cognition Instruct. **1**, 177 (1984).

17. A. Schoenfeld, *Mathematical Problem Solving* (Academic Press, New York, NY, 1985); A. Schoenfeld, Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics, in *Handbook for Research on Mathematics Teaching and Learning*, edited by D. Grouws (McMillan, New York, NY, 1992), pp. 334-370.

18. J. Heller and F. Reif, Prescribing effective human problem-solving processes: Problem description in physics, Cognition Instruct. **1**, 177 (1984).

19. A. Schoenfeld, Teaching mathematical thinking and problem solving, in *Toward the Thinking Curriculum: Current Cognitive Research*, edited by L. Resnick and B. Klopfer (ACSD, Washington, DC, 1989), pp. 83-103.

20. A. Schoenfeld and D. Herrmann, Problem perception and knowledge structure in expert novice mathematical problem solvers, J. Exp. Psychol. Learn. **8**, 484 (1982).

21. A. Newell and H. Simon, *Human Problem Solving* (Prentice Hall, Englewood Cliffs, NJ, 1972).

22. G. Polya, *Mathematical Discovery*, (Wiley, New York, NY, 1962).

23. H. Simon and J. Hayes, The understanding process: Problem isomorphs, Cog. Psych. **8**, 165 (1976).

24. J. Hayes and H. Simon, Psychological differences among problem isomorphs, in *Cognitive Theory*, edited by N. Castellan, D. Pisoni, and G. Potts (Lawrence Erlbaum, Hillsdale, NJ, 1977), pp. 21-40.

25. K. Kotovsky, J. Hayes, and H. Simon, Why are some problems hard? Evidence from the Tower of Hanoi, Cog. Psych. **17**, 284 (1985).

26. J. Bransford, A. Brown, and R. Cocking, *How People Learn: Brain, Mind, Experience, and School* (National Academy Press, Washington, DC, 1999).

27. R. Bjork and A. Richardson-Klavhen, On the puzzling relationship between environment, context and human memory, in *Current Issues in Cognitive Processes: The Tulane Flowerree Symposium on Cognition*, edited by C. Izawa, (Lawrence Erlbaum, Hillsdale, NJ, 1989), pp. 313-344.

28. D. Godden and A. Braddeley, Context-dependent memory in two natural environments: On land and under water, Brit. J. of Psychol. **66**, 325 (1975).

29. S. Reed, G. Ernst, and R. Bannerji, The role of analogy in transfer between similar problem states, Cog. Psych. **6,** 436 (1974).

30. E. Saltiel and J. Malgrange, Spontaneous' ways of reasoning in elementary kinematics, Eur. J. Phys. **1**, 73 (1980).

31. C. Singh, Assessing student expertise in introductory physics with isomorphic problems. II. Effect of some potential factors on problem solving and transfer, Phys. Rev. ST PER **4**, 010105 (2008).

32. P. Hardiman, R. Dufresne, and J. Mestre, The relation between problem categorization and problem solving among novices and experts, Mem. Cognition **17**, 627 (1989).

33. M. Chi, P. Feltovich, and R. Glaser, Categorization and representation of physics knowledge by experts and novices, Cog. Sci. **5**, 121 (1981).

34. T. de Jong and M. Ferguson-Hessler, Cognitive structure of good and poor problem solvers in physics, J. Educ. Psych. **78**, 279 (1986).

35. J. Larkin, J. McDermott, D. Simon, and H. Simon, Expert and novice performance in solving physics problems, Science **208**, 1335 (1980).

36. J. Larkin, Understanding, problem representations, and skill in physics, in *Thinking and Learning Skills*, edited by S. Chipman, J. Segal, and R. Glaser (Lawrence Erlbaum, Hillsdale, NJ, 1985), pp. 141-159.

37. J. Larkin, Cognition of learning physics, Am. J. Phys. **49**, 534 (1981).

38. C. Singh, When physical intuition fails, Am. J. Phys. **70**, 1103 (2002).

39. P. Cheng and K. Holyoak, Pragmatic reasoning schema, Cog. Psych. **17**, 391 (1985).

40. P. Johnson-Laird, *Psychology of Reasoning, Structure, and Content* (Cambridge, MA, 1972).

41. J. Kaput, Representation and problem solving: Methodological issues related to modeling, in *Teaching and Learning Mathematical Problem Solving: Multiple Research Perspectives* (Lawrence Erlbaum Associates, Hillsdale, NJ, 1985), pp. 381-398.

42. J. Zhang, The nature of external representations in problem solving, Cog. Sci. **21**, 179 (1997).

43. G. Hatano and K. Inagaki, Two courses of expertise, in *Child Development and Education in Japan*, edited by H. Stevenson, J. Azuma, and K. Hakuta (W. H. Freeman and Co., NY, 1986), pp. 262-272.

44. D. Gentner, Structure-mapping: A theoretical framework for analogy, Cog. Sci. **7**, 155 (1983).

45. K. Holyoak, The pragmatics of analogical transfer, in *The Psychology of Learning and Motivation*, edited by G. Bower (Academic Press, New York, NY, 1985).

46. K. Holyoak and K. Koh, Surface and structural similarity in analogical transfer, Mem. Cognition **15**, 332 (1987).

47. B. Ross, This is like that: The use of earlier problems and the separation of similarity effects, J. Exp. Psychol. Learn. **13**, 629 (1987).

48. B. Ross, Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems, J. Exp. Psychol. Learn. **15**, 456 (1985).

49. D. Gentner, Mechanisms of analogical learning, in *Similarity and Analogical Reasoning*, edited by S. Vosniadou and A. Ortony (Cambridge University Press, London, 1985), pp. 199-241.

50. L. Reeves and R. Weisberg, The role of content and abstract information in analogical transfer, Psychol. Bull. **115**, 381 (1994).

51. S. Reed, G. Ernst, and R. Bannerji, The role of analogy in transfer between similar problem states, Cog. Psych. **6**, 436 (1974).

52. L. Novick, Analogical transfer, problem similarity and expertise, J. Exp. Psychol. Learn. **14**, 510 (1988).

53. M. Shapiro, Analogies, visualization, and mental processing of science stories, in *Communication Yearbook*, vol. 9 (M. McLaughlin, Beverly Hills, CA, 1988), pp. 339-359.

54. R. Duit, On the role of analogies and metaphors in learning science, Sci. Educ. **75**, 649 (1991).

55. 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).

56. M. Chi, M. Bassok, M. Lewis, P. Reimann, and R. Glaser, Self-explanations: How students study and use examples in learning to solve problems. Cog. Sci. **13**, 145 (1989).

57. G. Cooper and J. Sweller, Effects of schema acquisition and rule automation on mathematical problem-solving transfer, J. Educ. Psych. **79**(4), 347 (1989).

58. T. Gog, L. Kester, and F. Paas, Effects of worked examples, example-problem, and problem-example pairs on novices' learning, Contemp. Educ. Psych. **36**, 212 (2011).

59. S. Kalyuga, P. Chandler, J. Tuovinen, and J. Sweller, When problem solving is superior to studying worked examples. J. Educ. Psych. **93**, 579 (2001).

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

61. D. Zollman and S. Rebello, Quantum mechanics for everyone: Hands-on activities integrated with technology, Am. J. Phys**. 70**, 252 (2002).

62. C. Singh, M. Belloni, and W. Christian, Approaches for improving students' understanding of quantum mechanics: Response to letters, Phys. Today **8**, 12 (2007).

63. G. Zhu and C. Singh, Improving students' understanding of quantum mechanics via the Stern-Gerlach experiment, Am. J. Phys. **79**(5), 499 (2011).

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

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

66. J. Morgan and M. Wittmann, Examining the evolution of student ideas about quantum tunneling, *Proceedings of the 2005 Phys. Ed. Res. Conference, Salt Lake City, UT*, edited by P. Heron, L. McCullough, and J. Marx (AIP Conf. Proc., Melville, NY, 2006), pp. 73-76.

67. C. Manogue and E. Gire, Representations for a spins first approach to quantum mechanics, *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE*, edited by N. Rebello, P. Engelhardt, and C. Singh (AIP Conf. Proc., Melville, NY, 2012), pp. 55-58.

68. E. Gire and C. Manogue, Making sense of operators, eigenstates, and quantum measurements, *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE*, edited by N. Rebello, P. Engelhardt, and C. Singh (AIP Conf. Proc., Melville, NY, 2012), pp. 195-198.

69. C. Singh, M. Belloni, and W. Christian, Improving students' understanding of quantum mechanics, Phys. Today **8**, 43 (2006).

70. C. Singh, Assessing and improving student understanding of quantum mechanics, *Proceedings of the 2005 Phys. Ed. Res. Conference, Salt Lake City, UT*, edited by P. Heron, J. Marx, and L. McCullough (AIP Conf. Proc., Melville, NY, 2006), p. 69.

71. C. Singh, Student difficulties with quantum mechanics formalism, *Proceedings of the 2006 Phys. Ed. Res. Conference, Syracuse, NY*, edited by L. McCullough, P. Heron, and L. Hsu (AIP Conf. Proc., Melville, NY, 2007), p. 185.

72. C. Singh, Helping students learn quantum mechanics for quantum computing, *Proceedings of the 2006 Phys. Ed. Res. Conference, Syracuse, NY*, edited by L. McCullough, P. Heron, and L. Hsu (AIP Conf. Proc., Melville, NY, 2007), p. 42.

73. E. Marshman and C. Singh, Framework for understanding the patterns of student difficulties in quantum mechanics, Phys. Rev. ST PER **11**, 020119 (2015).

74. C. Singh and E. Marshman, Review of student difficulties in upper-level quantum mechanics, Phys. Rev. ST PER **11**, 020117 (2015).

75. M. Belloni, W. Christian, and D. Brown, Open source physics curricular material for quantum mechanics, Comput. Sci. Eng. **9**, 24 (2007).

76. 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).

77. G. Passante, P. Emigh, and P. Shaffer, Investigating student understanding of basic quantum mechanics in the context of time-dependent perturbation theory, *Proceedings of the 2013 Phys. Ed. Res. Conference, Portland, OR*, edited by A. Churukian, P. Engelhardt, and D. Jones (AIP Conf. Proc., Melville, NY, 2014), p. 269.

78. G. Passante, P. Emigh, and P. Shaffer, Student ability to distinguish between superposition states and mixed states in quantum mechanics, Phys. Rev. ST PER **11**, 020135 (2015).

79. G. Passante, P. Emigh, and P. Shaffer, Examining student ideas about energy measurements on quantum states across undergraduate and graduate level, Phys. Rev. ST PER **11**, 020111 (2015).

80. P. Emigh, G. Passante, and P. Shaffer, Student understanding of time dependence in quantum mechanics, Phys. Rev. ST PER **11**, 020112 (2015).

81. C. Singh, Transfer of learning in quantum mechanics, *Proceedings of the 2004 Phys. Ed. Res. Conference, Sacramento, CA*, edited by P. Heron, S. Franklin, and J. Marx (AIP Conf. Proc., Melville, NY, 2005), p. 23.

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

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

84. C. Singh and G. Zhu, Cognitive issues in learning advanced physics: An example from quantum mechanics, *Proceedings of the 2008 Phys. Ed. Res. Conference, Ann Arbor, MI*, edited by C. Henderson, M. Sabella, and C. Singh (AIP Conf. Proc., Melville, NY, 2009), p. 63.

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

86. G. Zhu and C. Singh, Improving students' understanding of quantum measurement I: Investigation of difficulties, Phys. Rev. ST PER **8**, 010117 (2012).

87. G. Zhu and C. Singh, Improving students' understanding of quantum measurement II: Development of Research-based learning tools, Phys. Rev. ST PER **8**, 010118 (2012).

88. G. Zhu and C. Singh, Improving student understanding of addition of angular momentum in quantum mechanics, Phys. Rev. ST PER **9**, 010101 (2013).

89. S. Lin and C. Singh, Assessing expertise in quantum mechanics using categorization task, *Proceedings of the 2008 Phys. Ed. Res. Conference, Ann Arbor, MI*, edited by M. Sabella, C. Henderson, and C. Singh (AIP Conf. Proc., Melville, NY, 2009), p. 185.

90. S. Lin and C. Singh, Categorization of quantum mechanics problems by professors and students, Euro. J. Phys. **31**, 57 (2010).

91. J. Lobato, How design experiments can inform a rethinking of transfer and vice versa, Educ. Res. **32**, 17 (2003).

92. J. Lobato, Alternative perspectives on the transfer of learning: History, issues and challenges for future research, J. Learn. Sci. **15**, 431 (2006).

93. D. Ozimek, P. Engelhardt, A. Bennett, and N. Rebello, Retention and transfer from trigonometry to physics, *Proceedings of the 2004 Phys. Ed. Res. Conference, Sacramento, CA*, edited by J. Marx, P. Heron, and S. Franklin (AIP Conf. Proc., Melville, NY, 2005), p. 173.

94. G. Schraw, Promoting general metacognitive awareness, Instructional Science **26**, 113 (1998).

95. J. Flavell, Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry, Am. Psycho. **34**, 906 (1979).

96. A. Schoenfeld, What's all the fuss about metacognition?, in *Cognitive Science and Mathematics Education*, edited by A. Schoenfeld (Lawrence Erlbaum Associates, Hillsdale, NJ, 1987), pp. 189-215.

97. M. Lawson, Being executive about metacognition, in *Cognitive Strategies and Educational Performance*, edited by J. Kirby (Academic Press, New York, 1984), pp. 89-109.

98. B. White and J. Frederiksen, A theoretical framework and approach for fostering metacognitive development, Educ. Psychol. **40**, 211 (2005).

99. J. Wheeler, in *Mathematical Foundations of Quantum Theory*, edited by A. Marlow (Academic Press, New York, NY, 1979).

100. M. Schneider and I. LaPuma, A simple experiment for discussion of quantum interference and which-way measurement, Am. J. Phys. **70**, 266 (2001).

101. R. Feynman, R. Leighton, and M. Sands, *Feynman Lectures on Physics* (Addison-Wesley, Reading, MA, 1989).

102. G. Glass and K. Hopkins, *Statistical Methods in Education & Psychology* (Allyn & Bacon, Boston, MA, 1996).

103. R. Sayer, A. Maries, and C. Singh, Development and evaluation of a quantum interactive learning tutorial on the double-slit experiment (in press).

104. M. Bassok and K. Holyoak, Interdomain transfer between isomorphic topics in algebra and physics, J. Exp. Psych. **15**, 153 (1989).

4.9 APPENDIX C

Here, we discuss common student difficulties on the DSE transfer questions both before and after students worked on the DSE tutorial. Since the data were qualitatively similar for the graduate students regardless of whether they had completed the MZI tutorial before taking the DSE pre-test, the graduate students from all cohorts are combined. We also carried out think-aloud interviews with undergraduate and graduate students to further understand the common types of incorrect reasoning they used to answer these questions, which often provided further insight into their difficulties.

4.9.1 Difficulties with Interference of Single Photons – No Polarizers

Among the students who answered question 1, the vast majority of both undergraduate and graduate students answered it correctly (clear interference pattern shown) as shown in Table 4-VII. A small percentage of students selected answers which indicated that no interference pattern is observed, but none provided reasoning for their answers. Roughly one quarter of the undergraduate students and one sixth of the graduate students either did not respond or indicated that they did not know whether photons will exhibit interference in this case. These percentages drop to nearly zero in the post-test.

Table 4-VII Percentages of undergraduate (US) and graduate students (GS) with different answers on question 1. Bold italics indicates the correct response.

	Interference	No Interference	Other	No Response / ”I don’t know”
US-Before Tutorial	***63%***	7%	7%	23%
US-After Tutorial	***98%***	2%	0%	0%
GS-Before Tutorial	***71%***	11%	2%	16%
GS-After Tutorial	***83%***	5%	8%	3%

4.9.2 Difficulties with Effect of One Polarizer on Interference Pattern

Question 2 involves a DSE in which a vertical polarizer is placed in front of only one of the slits. In this situation, WPI will be known for horizontally polarized photons and will not be known for vertically polarized photons (as explained in the section discussing the isomorphism between the DSE and MZI). Therefore, the pattern observed on the screen will consist of an interference pattern provided by the vertically polarized photons (which do interfere) on top of a uniform background provided by the horizontally polarized photons which do not interfere. This was the most challenging question for both student populations. As shown in Table 4-VIII, for both populations, the most common incorrect answer choice is that no interference is observed in this situation. Students with this answer typically reasoned that WPI is known for all photons because the polarizer "tags" the photons that go through it by polarizing them (this reasoning did not always mention WPI explicitly). For example, one student stated: "*No interference because you are essentially 'tagging' half the photons*." Another student stated: "*No interference since the polarizer tells us which slit the photon went through*." This difficulty is also common in the MZI context when a vertical polarizer is placed in one of the paths: Many students thought that no interference is observed at either detector because the polarizer provides WPI for the photons that take that path by 'tagging' them. Interestingly, more graduate students use this type of reasoning after working on the DSE tutorial than before. This may be because before working on the DSE tutorial, some students (21%) provided responses that were difficult to categorize, and some (16%) did not provide a response, but after working on the DSE tutorial, the majority of these students provided responses that could be categorized, some of which used the incorrect reasoning that the vertical polarizer provides WPI for vertically polarized photons detected at the screen.

Table 4-VIII Percentages of undergraduate (US) and graduate students (GS) with different answers on question 2 (Full Interference, Partial Interference, No Interference, Other, and No Response/"I don't know"). Bold italics indicates the correct response.

	Full Interf.	***Partial Interf.***	No Interf.	Other	No Response/ "I don't know"
US-Before Tutorial	5%	***38%***	17%	16%	24%
US-After Tutorial	2%	***70%***	16%	12%	0%
GS-Before Tutorial	12%	***31%***	19%	21%	16%
GS-After Tutorial	10%	***51%***	32%	5%	2%

For students who attempted to transfer their learning from the MZI to the DSE context and explicitly reasoned in terms of WPI, 67% of them (including both undergraduate and graduate students) reasoned correctly (note that this is the most challenging question for both undergraduate and graduate students). For example, one student wrote: "*The interference pattern will be fuzzier because we do have which-path data for any photons that are not vertically polarized*" (common correct reasoning). Another student wrote: "*I only see two lines on the screen because we have which-path information about one of the slits*." The second student is using WPI reasoning incorrectly, but at the very least, he is recognizing that this reasoning may be useful in the DSE context and is attempting to transfer learning from the MZI to the DSE context.

4.9.3 Difficulties with Effect of Two Polarizers on Interference Pattern

Questions 3 and 4 evaluate student understanding of the effect of two polarizers on the interference pattern. Students showed significant transfer on these two questions, as shown in Tables 4-IX and 4-X. Among the students who answered these questions before working on the DSE tutorial, the majority of them answered them correctly. Also, on these questions, the performance of undergraduate students after working on the DSE tutorial is close to 100%. It appears that the undergraduate students were able to transfer learning from the MZI to the DSE after going through the MZI tutorial and consolidate their learning while going through the DSE tutorial to develop a solid understanding of the effect of two polarizers on the interference pattern in the DSE. On the other hand, graduate students showed a lesser improvement.

When a vertical polarizer is placed after each slit (question 3), there will be no horizontally polarized photons that reach the screen. For the vertically polarized photons that reach the screen, WPI is not known and therefore these photons will show an interference pattern at the screen. Since the same number N of photons reach the screen, this interference pattern is no different from the pattern observed when no polarizers are placed after either slit. As shown in Table 4-IX, the most common incorrect answer for both the undergraduate and graduate students is that there will be no interference. A common incorrect reasoning, especially before students worked on the DSE tutorial, is that in this situation, WPI will be known for all photons.

Table 4-IX Percentages of undergraduate (US) and graduate students (GS) with different answers on question 3 (Full Interference, Partial Interference, No Interference, Other, and No Response/"I don't know"). Bold italics indicates the correct response.

	Full Interf.	Partial Interf.	No Interf.	Other	No Response/ "I don't know"
US-Before Tutorial	***52%***	2%	14%	6%	26%
US-After Tutorial	***93%***	2%	2%	0%	2%
GS-Before Tutorial	***60%***	3%	14%	5%	17%
GS-After Tutorial	***83%***	7%	5%	2%	3%

If a vertical polarizer is placed after one slit (e.g., the top slit) and a horizontal polarizer is placed after the other slit (e.g., the bottom slit) as in question 4, then WPI is known for all photons because a horizontally polarized photon detected at the screen must have gone through the bottom slit and a vertically polarized photon detected at the screen must have gone through the top slit. On this question, the most common incorrect answer was that a full interference pattern should form, as shown in Table 4-X. Students who provided responses of this type may have had difficulty recognizing that the polarizers provide WPI for all photons, or may believe that even though WPI is known for all photons, an interference pattern is still observed. For example, one graduate student recognized that WPI can be obtained both for a vertically and a horizontally polarized photon detected at the screen, and concluded that neither horizontally nor vertically polarized photons interfere with themselves. However, she thought that they can interfere with each other and said: "*I don't know… would they [photons coming from one slit] be able to interfere with the ones [photons] coming from the other slit…?*" When probed further, she said: "*If it [photon] can only go through one slit or the other it can't interfere with itself, but once it goes through it, there would still be wave propagation […] would it [a vertically* *polarized photon] be able to interfere with the horizontally polarized photons or not… I don't know.*" When the interviewer asked, "*So what you're saying is that a single photon can only go through one slit or the other but you're not sure if that implies that there's no interference because that photon might interfere with another photon that's coming through the other slit, is that right?*", she responded, "*Yeah.*"

Table 4-X Percentages of undergraduate (UG) and graduate students (GS) with different answers on question 4 (Full Interference, Partial Interference, No Interference, Other, and No Response/"I don't know"). Bold italics indicates the correct response.

	Full Interf.	Partial Interf.	***No Interf.***	Other	No Response/ "I don't know"
US-Before Tutorial	10%	0%	***52%***	2%	36%
US-After Tutorial	2%	0%	***93%***	0%	5%
GS-Before Tutorial	9%	0%	***64%***	7%	21%
GS-After Tutorial	8%	7%	***81%***	3%	0%

4.9.4 Difficulties with Quantum Eraser

The last situation (vertical polarizer after one slit, horizontal polarizer after the other, 45° polarizer in front of the screen) is known as a "quantum eraser" because the last polarizer erases WPI that could be obtained due to the effect of the other two polarizers. Table 4-XI shows that the most common incorrect answer for both undergraduate and graduate students was that there will be no interference in this situation. Many students who provided these types of responses ignored the third polarizer. For example, one student stated: "*I don't think interference is possible because you are still identifying the path of one side of photons as different from the other.*" Another student stated: "*See no interference since one is horizontally and the other vertically polarized.*" These types of reasoning indicate that students essentially ignored the effect of the third polarizer, which erases WPI. As further evidence of transfer for this question, many students, especially in the DSE post-test, specifically mentioned the similarity to the MZI, wrote down "quantum eraser," or reasoned in a manner which could have been learned only in the context of the MZI (e.g., the third polarizer erases the WPI obtained from the other two polarizers) since this situation was not mentioned at all in the DSE tutorial.

Table 4-XI Percentages of undergraduate (US) and graduate students (GS) with different answers to question 5 (Full Interference, Partial Interference, No Interference, Other, and No Response/"I don't know"), including percentages of students who mention MZI or quantum eraser when responding to question 5. Bold italics indicate the correct response.

	Full Interf.	Partial Interf.	No Interf.	Other	No Response/ "I don't know"	Mention MZI or Quantum Eraser
US-Before Tutorial	***43%***	2%	17%	2%	36%	24%
US-After Tutorial	***86%***	5%	2%	0%	7%	66%
GS-Before Tutorial	***52%***	2%	12%	12%	22%	5%
GS-After Tutorial	***76%***	10%	12%	2%	0%	27%

5.0 THE CHALLENGES OF CHANGING TAS' GRADING PRACTICES: SHIFTING THE BURDEN OF PROOF FROM INSTRUCTOR TO STUDENT

5.1 INTRODUCTION

At large research institutions in the U.S., graduate students in physics play an important role in the education of undergraduate students in physics courses. In particular, it is quite common for physics graduate Teaching Assistants (TAs) to teach introductory physics recitations or lab sections. Common goals of instructors of introductory physics courses are to help students learn disciplinary concepts and principles [1], to help them develop effective problem-solving approaches and to make better use of problem solving as an opportunity for learning [2-10].

TAs are typically involved in grading students' work. Grading can help shape student learning by communicating instructors' goals and expectations to their students [11-15]. However, grading practices are shaped by a vast array of beliefs, goals, and knowledge based upon TAs' and instructors' past experiences as students and various aspects of the immediate classroom context (e.g., students disagreeing with the TAs and instructors about their grades, expectations of peers and administrators, workload, etc.) [16-21], and their grading goals may often be in conflict with their actual grading practices. For example, prior research suggests that the common grading practice of TAs is often to treat grading as summative (feedback to the instructor about what students have learned) and ignore the formative assessment aspect of grading, e.g., the fact that emphasizing and rewarding the explication of the problem-solving process in grading can help students develop problem solving skills and learn physics [22-23]. Moreover, many physics graduate TAs are taking core courses and learning to be researchers and teaching assistants concurrently, and they must meet the expectations of both their research advisors and course instructors. The resources accessible to them for teaching are usually their own experiences as students as well as the requirements of the departments and/or instructors they assist. In addition, TAs usually have very little time and support to clarify their goals for grading and develop grading practices that adequately reflect their instructional values and beliefs. Under these constraints, TAs can benefit from the opportunity to reflect upon their grading goals to make them more formative (instead of just summative) and align their grading practices with these goals for grading that can foster student learning.

To help TAs reflect upon and refine their grading goals and align them with their grading practices, professional development courses can provide valuable opportunities. The case study presented here involved 15 first-year physics graduate TAs participating in a mandatory semester-long TA professional development course at a large research university in the U.S. To help TAs reflect upon their goals and beliefs about grading and provide them with scaffolding support to promote positive changes in their beliefs and practices about grading, we carried out an intervention in which the TAs were asked to 1) clarify their initial goals for grading and their grading practices, 2) consider a physics education research (PER) inspired grading rubric and reflect upon how the rubric supports the goals of helping students develop effective problem solving skills and learn physics, and 3) reflect on and possibly resolve conflicts between their initial grading practices and the rubric criteria. The intervention was designed to promote positive changes in TAs' beliefs and practices related to grading. There was extensive discussion about the value of using such a rubric in the TA professional development course in which the TAs and instructors participated. Moreover, the TAs participating in this research were told that they should assume that they had full control over grading policies and they had distributed the rubric to their students and told them that they would always be graded using it. We hypothesized that a PER inspired grading rubric may prompt TAs to think about specific grading criteria (that they may not have considered on their own) in light of the grading goals that support conceptual understanding of physics and development of effective problem-solving skills. In addition, TAs may also reflect on how grading with a rubric may lead to greater objectivity, consistency, and repeatability when assigning scores to student work.

We investigated how providing opportunities for TAs to grade introductory physics student solutions with and without the rubric and facilitating class discussions about the rubric may help TAs reflect on the advantages of using a rubric that appropriately weights effective problem-solving practices. Our investigation is a part of a series of design experiments [17-19] implemented in the context of a semester long graduate physics TA professional development course. This case study focuses on how asking TAs to grade several introductory physics solutions (with different levels of explication of the problem-solving process) for two isomorphic problems while being provided a PER inspired grading rubric and class discussions (about why grading using such a rubric can foster learning) impacts TAs' beliefs about grading and their grading practices. The findings from this design experiment may inform leaders of professional development courses for TAs and instructors and physics education researchers in contemplating strategies for improving beliefs about grading and grading practices to foster learning. The case study was designed to investigate the following research questions:

1. How do TAs apply the different components of a PER inspired rubric that weights the problem-solving process to grade student solutions of introductory physics problems?
2. Do TAs use the rubric consistently when grading problems involving the same physics principles but having different surface features?
3. Do TAs apply the grading rubric differently than an “expert rater”, e.g., physics education researchers who study problem solving?
4. Do TAs’ grading practices change during a 15 week semester after using the rubric to grade and having discussions about the benefits of using a good rubric in the professional development course and carrying out their assigned teaching responsibilities simultaneously?
5. According to the TAs, what are the pros and cons of using a rubric to grade student solutions in introductory physics?

5.2 BACKGROUND

5.2.1 Effective Problem-solving Approaches

Many prior studies [2-13] have documented differences between experts and novices in a particular domain when approaching problems. Both use heuristics to guide their search process in identifying the gap between the problem goal and the state of the solution and taking action to bridge this gap. However, novices differ from experts in the types of heuristics they use to solve problems. Novices approach problems in a haphazard manner, typically searching for appropriate equations first and then plugging in numbers until they get a numerical answer [3-8]. Furthermore, novices often draw on their naive knowledge base rather than formal physics knowledge [3-8]. Novices also engage in pattern matching, i.e., attempting to solve a problem using another previously solved problem with similar surface features, even if the underlying concepts and principles are different [3-8]. On the other hand, experts devote time and effort to qualitatively describe the problem situation, identify principles and concepts that may be useful in the analysis of the problem, and retrieve effective representations based on their better organized domain knowledge [3-8]. In addition, experts devote time to plan a strategy for constructing a solution by devising a useful set of intermediate goals and means to achieve them, frequently by working in a backward manner [3-8]. Experts also spend more time than novices in using diverse representations to analyze and explore problems (especially when they are not sure how to proceed) [3-8]. Experts also engage more than novices in self-monitoring by evaluating previous steps and revising their choices as needed [9-11]. They utilize problem solving as a learning opportunity more effectively by engaging in self-repair: identifying and attempting to resolve conflicts between their own mental model and the scientific model conveyed by peers' solutions or worked-out examples [9-11].

The major goals of many introductory physics courses include helping students develop expertise and be able to transfer their learning from one context to another in future problem solving. To prepare students for future learning [24], instruction that fosters both the development of problem-solving skills and conceptual understanding can be particularly beneficial. When students solve a large number of similar problems, they may become routine experts – they learn to solve a similar set of problems faster and more accurately, but this process does not necessarily help develop their problem solving and metacognitive skills nor does it help them build a robust knowledge structure [25]. As a result, they may lack the flexibility and adaptability necessary to solve novel problems. Students who are given opportunities to develop both conceptual understanding and effective problem-solving skills may become adaptive experts who possess a well-organized knowledge structure and have developed robust problem solving, reasoning, and meta-cognitive skills [25].

One related issue is how effective grading practices can impact the development of expertise. For example, whether students are graded on explicating the problem-solving process or not can impact whether students use effective approaches to problem solving (e.g., starting problem solving with a conceptual analysis of the problem, doing planning and decision making before implementing the plan, and then doing a reasonability check for the solution obtained and reflecting upon the problem-solving process to learn from it) instead of a plug-and-chug approach (e.g., starting by looking for a formula that matches the quantities in the problem statement). Prior research suggests that if students in traditionally taught introductory physics classes are matched in terms of their prior performance in a physics class (for example, students who have a C grade at the time of the interview are matched with other students with a similar grade) and students in one of the two groups are forced to use effective approaches to problem solving (experimental group) and those in the other group are allowed to use whatever approach they want to use to solve the problem (control group), the performance of the students in the experimental group becomes significantly better than those in the control group as the complexity of the problem increases [4]. This study was conducted in a one-on-one situation outside of the class and many students in the experimental group were themselves surprised at how they were able to solve complex problems when they were forced to use a systematic approach [4]. The researchers of this study noted that the students in the experimental group often had a tendency to start looking at the formula sheet before doing a qualitative analysis of the problem and planning the problem solution and had to be reminded that they could not do so since they were part of a research study [4]. The fact that students had a tendency to start the problem-solving process by looking at a formula sheet is a testament to the fact that traditionally taught courses do not help students learn effective problem solving strategies, even though they can be helpful both for the development of skills and learning physics.

Since students often value what they are graded on, grading their solutions using a rubric that emphasizes the explication of the problem-solving process and puts the BOP for showing their work on students (and not on the instructor for interpreting what the students must have been thinking) can encourage students to use a systematic approach to problem solving. If we have a buy-in from the instructor and TAs, such a rubric can be distributed to all students and students can be informed that they should always follow the rubric when solving problems since they would always be graded on it. The students can also be told about the prior study [4] that shows the benefits of solving problems systematically by explicating the problem-solving process on the ability of students to solve complex problems.

5.2.2 Grading Rubrics

In grading, findings of PER suggest placing the burden of proof (BOP) for explicating the problem-solving process on students. In the spirit of formative assessment [26,27], grading can provide feedback that can improve student learning and communicate to learners what practices are useful for learning the discipline and for developing problem-solving skills [28]. Effective grading practices can also communicate to students what to focus on in future learning activities [29-35]. Such practices can encourage students to explain their reasoning by placing the BOP on the student (i.e., requiring that the students explain the reasoning underlying their solutions) and provide them with an artifact to reflect on and learn from after problem solving (i.e., their own clearly articulated solution in which the problem-solving process is explicated) [36]. Grading should reward the use of effective problem-solving strategies such as drawing a diagram, listing known and unknown quantities, clarifying considerations in setting up sub-problems, and evaluating the reasonability of the problem solution.

Grading rubrics are scoring tools which outline the performance expectations for an assignment. Good rubrics often divide a problem into various parts and provide descriptions of how scores should be allocated for varying levels of mastery. Effective grading rubrics offer many advantages to both students and instructors. A grading rubric that rewards explication of the problem-solving process (instead of focusing mainly on the correctness of the final answer) can give students an incentive to use problem solving strategies which are useful for the development of important skills and learning physics. Prior research [4] suggests that when students are forced to use effective problem-solving approaches, they are significantly more successful in solving complex physics problems compared to matched students with similar course grades at the time. A good grading rubric can provide a consistent grading standard for students with focus on approaches that enhance students' knowledge and skills. Research in various domains has shown that good rubrics can serve as formative assessment tools for students, helping them recognize strengths and weaknesses of their work and monitor their progress toward mastery [37-40]. By knowing ahead of time what is expected of their work (if they are given a rubric and informed that they will always be graded on it), students may be encouraged to practice effective problem-solving strategies (e.g., initial analysis and planning, explication, reasonability check for the solution and other types of reflection on the problem-solving process to learn from solving the problem) that may help them develop problem-solving skills and learn [37]. In addition, students and instructors may have a more consistent judgement of the students' work because the use of rubrics decreases variation in scores between different graders [41]. Rubrics can also benefit students and instructors by allowing them to review a student's score on a problem for each component of the rubric and obtain a clearer picture of what is causing student difficulties [41]. Instructors can then develop an instructional approach that reduces student difficulties, e.g., modeling effective problem solving in class and providing coaching and feedback to help students learn. Also, if students develop a better understanding of their difficulties, they may focus on developing a better knowledge structure and problem solving skills.

An effective rubric can be a means of formative assessment if it includes criteria for effective problem-solving approaches discussed earlier. Docktor and Heller [42] designed grading rubrics to reinforce in students the perception that problem solving is a process and requires both content knowledge and problem solving skills. Their rubrics assess students' proficiency related to both content knowledge and skills and include the processes of organizing problem information into a useful description, selecting appropriate physics principles, applying physics concepts and principles to the specific situations in the problem, using mathematical procedures appropriately, and communicating an organized reasoning pattern [42]. Docktor and Heller state that "it is important to consider only what is written and avoid the tendency to assume missing or unclear thought processes are correct [42]." Such a rubric aligns with the notion that instructors should place the BOP of explicating the problem-solving process on the student and value a logical, coherent solution by always grading on explication of reasoning to improve students' learning.

5.2.3 Physics Graduate TAs and Their Typical Role and Training

In introductory physics courses (both recitations and labs) at large research universities, graduate TAs are often responsible for grading homework and quizzes and at least part of the exams (though part of the exams may be graded by the course instructor). At the Graduate Education in Physics Conference jointly sponsored by the American Physical Society and the American Association of Physics Teachers, which was attended by physics graduate directors and chairs of 66 physics departments in 2008 and representatives of 74 physics departments in 2013, discussions with faculty about teaching assistantships indeed suggest that the majority of physics departments at research institutions in the U.S. employ physics graduate students as TAs for introductory physics course recitations and for introductory laboratory classes [43]. The TAs are expected to do the bulk of grading in these courses. The discussions at the conference suggest that in some physics departments at research universities, one or two semesters of TA work is a mandatory requirement towards their PhD degree.

However, the conference participants noted that even in the departments in which the TA work is not mandatory, a majority of PhD students spend at least one or two semesters as a TA, typically for introductory physics recitation or laboratory courses [43]. A majority of physics departments provide a very short training to the TAs (half day or less) to help them learn how to carry out their teaching responsibilities [43]. However, a handful of departments have provided semester long TA professional development programs similar to the one discussed in this study. Moreover, most conference participants noted that the TAs usually carry out the tasks in their recitations, labs, and grading without significant supervision or guidance from their supervising instructor except for general guidelines about how to carry out the recitation or how to grade (e.g., whether they should solve homework problems on the board in the recitation, give a quiz at the beginning or at the end of the recitation, how easy or strict they should be in grading homework and quizzes, etc.) [43].

5.2.4 Physics Graduate TAs' Instructional Beliefs and Practices

Prior studies have identified common beliefs and practices among physics TAs that have implications for improving learning [20-23,44-49]. For example, research suggests that graduate TAs sometimes struggle to understand the value of thinking about the difficulty of a problem from an introductory student's perspective [47,48] and believe that if they know the material and can explain it to their students in a clear manner, it will be sufficient to help their students learn. Also, while graduate TAs are able to recognize useful solution features and articulate why they are important when looking at sample introductory physics student solutions provided to them, they do not necessarily include those features in their own solutions written for introductory physics courses [20,21,49].

One of the tasks that physics TAs are often responsible for is grading. However, due to their prior experiences as students, time-constraints, and the limited training and feedback offered to new TAs, misalignments between their instructional beliefs and their teaching practices can occur (for example, between their grading goals and their grading practices) [22,23]. Some physics TAs may understand that grading can help students develop problem solving skills and help instructors identify common student difficulties, but their grading practices may not necessarily be conducive to helping students learn expert-like problem solving strategies and develop a coherent understanding of physics [22,23]. Prior research suggests that many instructors place the BOP on themselves for explication of the problem-solving process and are hesitant to take off points if the final answer is correct but the problem-solving process is not shown [1]. Prior research also hints at the fact that many TAs may not have had the opportunity and support to think about their goals for grading in introductory physics and reflect upon how their goals are aligned with their grading practices [22,23].

Appropriate professional development of physics graduate TAs that provides them the support and incentives to help their students learn better is an important task. The purpose of this case study is to investigate whether encouraging and supporting TAs to use a carefully designed grading rubric to grade introductory physics student solutions and discussing with them the benefits of using a good grading rubric in a professional development course can help TAs shift where they place the BOP for the explication of the problem-solving process. In particular, the investigation focuses on the impact of activities in a TA professional development course that focused on helping the TAs reflect upon the purpose of grading and why an effective rubric has the potential to help introductory physics students learn physics and develop effective problem solving skills. The graduate TAs who participated in this study were also simultaneously teaching introductory physics recitations or introductory labs, which could provide synergistic benefits for what the TAs learned and discussed with peers and their instructor in the TA professional development course. As noted earlier, conflicts often exist between TAs' grading goals and grading practices [22,23]. The present case study investigates the impact on TAs' grading goals and practices of efforts to help them align those goals and practices by 1) giving them opportunity to clarify their initial goals for grading and their grading practices, 2) facilitating opportunities for them to think about, practice and discuss (with peers and the professional development course instructor) how a good rubric can support the goals of helping students learn physics and develop effective problem solving skills, 3) and allowing for reflection on possible conflicts between their initial grading practices and the rubric criteria.

5.3 METHODOLOGY AND DATA COLLECTION

5.3.1 Participants

In this investigation, we collected grading data from a mandatory semester-long TA professional development course led by one of the authors. The course met for two hours each week for the entire semester. The TAs in general were expected to do one hour of homework each week pertaining to the professional development course. A total of 15 first-year TAs were enrolled in the course, which was designed to prepare them for their teaching responsibilities. The TAs had also attended a day-long new teaching assistant workshop facilitated by the university, but this workshop was general and did not focus on discipline-specific issues in teaching and learning physics. The majority of the TAs were concurrently teaching recitations for introductory physics courses for the first time. A few TAs were also assigned to facilitate a laboratory section or grade students' work in various physics courses for the first time. A majority of the TAs also served as tutors in a resource room where introductory students are assisted with any help they need with physics including their physics homework and laboratory reports. The participants consisted of a mix of domestic and international students originating from nations such as China, India, Turkey, etc. There were four female TAs and 11 male TAs. The demographics of the TAs in this course are similar to national norms [50].

5.3.2 Data Collection

5.3.2.1 Development and validation of the data collection tool

The data on TAs' goals for grading and grading approaches were collected using a Group Administered Interactive Questionnaire (GAIQ), previously developed and validated by three of the investigators in collaboration with two graduate student researchers in physics education for use with TAs and instructors [20]. This tool consists of a series of activities involving worksheets which are designed to clarify a TA or instructor's ideas about helping students learn physics content and problem solving skills. The GAIQ worksheets and artifacts encourage reflection on various facets of teaching physics problem solving: Designing problems on a particular physics topic with features effective for use in different situations (e.g., questions for clicker and class discussion, homework, quizzes, exams, collaborative learning etc.), designing solutions to homework problems that will help students learn, and grading student solutions. Questionnaires on each facet of teaching problem solving (e.g., problem types, instructors' example solutions, and grading) involve three stages: 1. TAs/Instructors are individually asked to solve a core problem (shown in Fig. 5-1) suitable for distributing to their students. 2. TAs/Instructors work in groups of three to answer the same questions as in the pre-class activity and then a whole class discussion takes place in which groups share their work. 3. TAs/Instructors individually complete another worksheet in which they can modify their previous answers and connect their ideas to a list of pre-defined features about teaching problem solving developed by the researchers. In this investigation, we focus on these stages in the context of grading.

You are whirling a stone tied to the end of a string around in a vertical circle having a radius of 0.65 m. You wish to whirl the stone fast enough so that when it is released at the point where the stone is moving directly upward it will rise to a maximum height of 23 meters above the lowest point in the circle. In order to do this, what force will you have to exert on the string when the stone passes through its lowest point one-quarter turn before release? Assume that by the time you have gotten the stone going and it makes its final turn around the circle, you are holding the end of the string at a fixed position. Assume also that air resistance can be neglected. The stone weighs 18 N.

The correct answer is 1292 N.

Figure 5-1 Core Problem 1.

The specific artifacts involving grading activities included five student solutions (see an example of two student solutions in Fig. 5-2), which were based upon actual students' common answers in the final exam. The artifacts were chosen to reflect differences between expert and novice problem solving from the research literature such as including a diagram describing the problem, explication of sub-problems, justification of solution steps, evaluation of the final answer, explication of the scientific principles used, evidence of reflective practices, etc. [7].

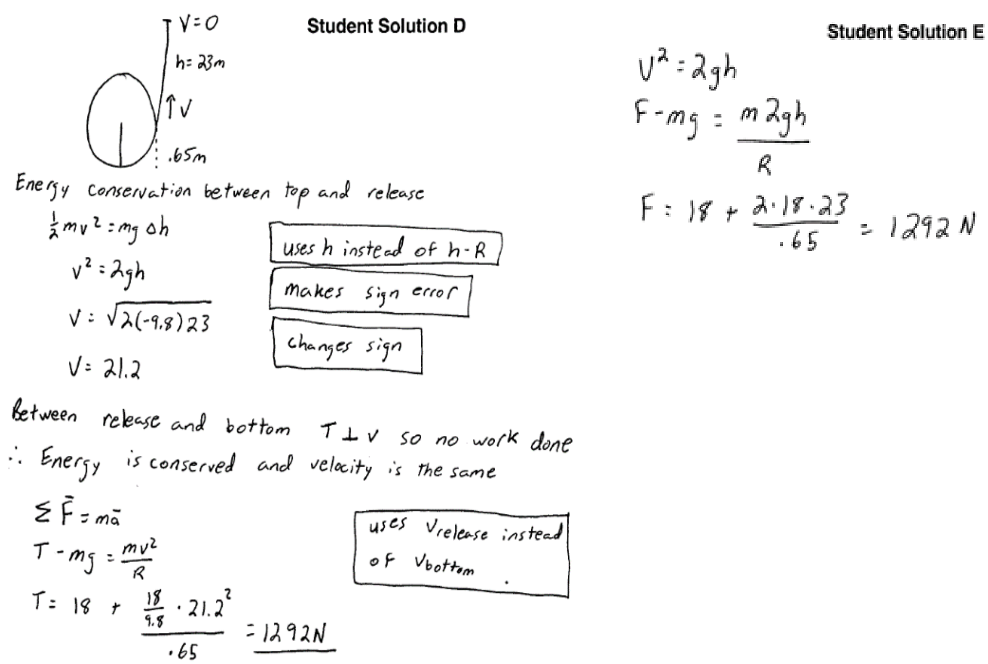
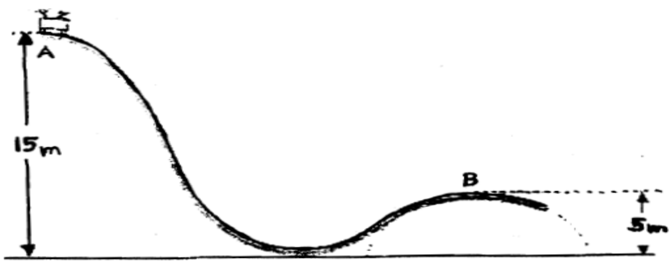


Figure 5-2 Student Solution D (SSD) and Student Solution E (SSE) to Core Problem 1.

Instructors' responses to interview questions about the instructional artifacts revolving around Core Problem 1 and five student solutions were used to create the initial GAIQ worksheets including those the worksheets on grading [20]. The GAIQ is meant to take the place of individual TA/instructor interviews about the teaching and learning of problem solving. While the development and validation of the GAIQ was a very time-consuming process [7], the GAIQ requires significantly less time than interviews for data collection and analysis. Equally important, it avoids researcher intervention in the process of clarifying the interviewees' responses, and the inter-rater agreement on the coding of the data obtained and interpretation of the data is excellent. Thus, the GAIQ worksheets can be used by researchers and professional developers at different institutions to collect and analyze data, and data across different institutions can readily be compared with relative objectivity.

The initial version of the GAIQ was iterated between the researchers and physics instructors and modified to a version which was administered several times in the context of professional development for Israeli pre-service and in-service teachers [51]. After each initial implementation and feedback from the teachers, the GAIQ was refined further until a version was developed that satisfied the researchers. The GAIQ tool was then adapted for a professional development course for physics teaching assistants in the U.S. The GAIQ including the grading activities was implemented in three different semesters in a TA training course in the U.S. After each implementation, the researchers iterated the version several times between them. A graduate student researcher in PER observed the three different semesters of the TA training course when TAs' worked on the GAIQ. The graduate student researcher and two of the authors revised and iterated the GAIQ based upon the TAs' comments and responses. This validation process in the context of the TA training class ensured that TAs interpreted all components of the GAIQ appropriately as the researchers had intended.

A friend told a girl that he had heard that if you sit on a scale while riding a roller coaster, the dial on the scale changes all the time. The girl decides to check the story and takes a bathroom scale to the amusement park. There she receives an illustration (see below), depicting the riding track of a roller coaster car along with information on the track (the illustration scale is not accurate). The operator of the ride informs her that the rail track is smooth, the mass of the car is 120 kg, and that the car sets in motion from a rest position at the height of 15 m. He adds that point B is at 5 m height and that close to point B the track is part of a circle with a radius of 30 m. Before leaving the house, the girl stepped on the scale which indicated 55 kg. In the roller coaster car the girl sits on the scale. Do you think that the story she had heard about the reading of the scale changing on the roller coaster is true? According to your calculation, what will the scale show at point B?



The correct answer is 180 N.

Figure 5-3 Core Problem 2.

To investigate TAs' use of grading rubrics, we added another Core Problem 2 (see Fig. 5-3), two additional student solutions (see Fig. 5-4), and an explicit model for grading in the form of a grading rubric (see Table 5-I) to the GAIQ grading activities in the present study. These additional components were added during the iteration process for the GAIQ worksheets involving rubrics used in the TA training course in this study. Core Problem 2 is similar to Core Problem 1 because it also involves a synthesis of the same important physics concepts and principles, is context-rich, and is difficult enough to require an average student to use an exploratory decision making process as opposed to an algorithmic procedure [7]. Core Problem 2 was designed, validated and approved by four physics instructors who taught introductory physics courses at the University of Minnesota and was used on final exams. Two additional student solutions (Student Solution F (SSF) and Student Solution G (SSG)) were developed for the Core Problem 2 and iterated several times by the researchers based on common student responses to the Core Problem 2 (see Fig. 5-4). SSF for Core Problem 2 is similar to SSD for Core Problem 1 in that SSF includes a diagram, articulation of the principles used to find intermediate variables, and clear justification for the final result. SSG for Core Problem 2 is analogous to SSE for Core Problem 1 in that it is brief with no explication of reasoning, and it does not give away any evidence for mistaken ideas. However, the three lines of work in SSG are also present in SSF.

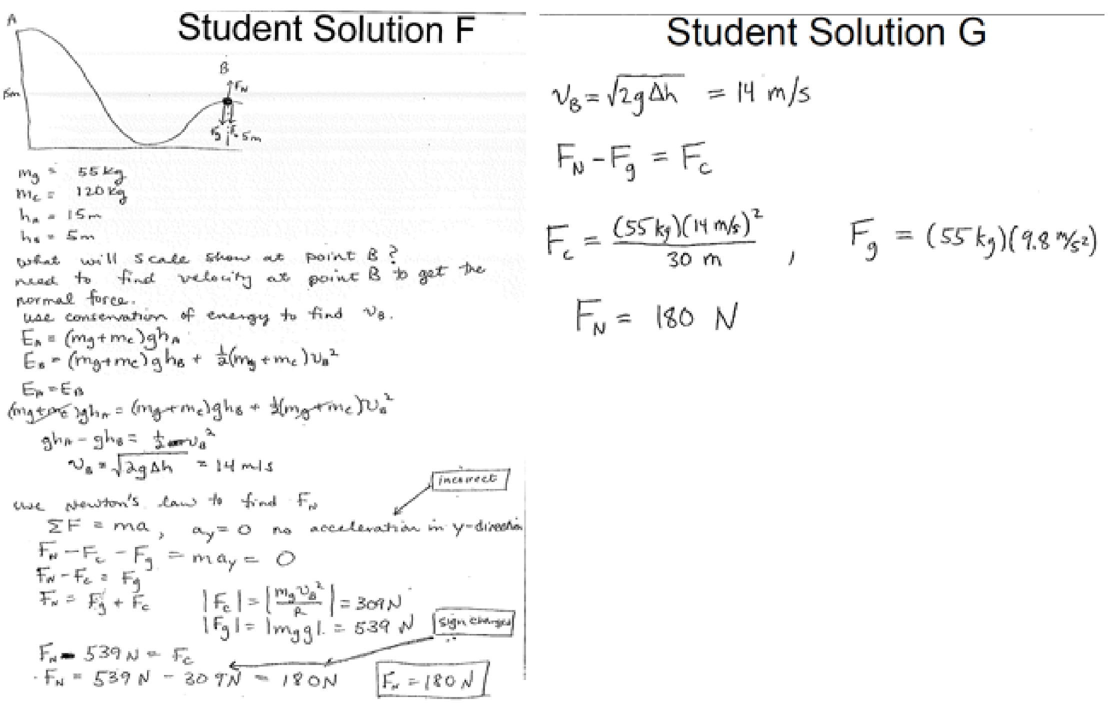


Figure 5-4 Student Solution F (SSF) and Student Solution G (SSG) to Core Problem 2.

A standard grading rubric was developed collaboratively by four physics education researchers and iterated many times before it was implemented in this study (see Table 5-I). The rubric emphasizes critical aspects of problem solving (e.g., invoking and justifying of physics principles, evaluating of final solution, etc.) that have been found in the literature to develop problem solving skills and improve physics content knowledge [37,52]. In addition, it was designed to be general enough that it could be applied to a variety of physics problems. It is similar to the Docktor and Heller rubric [42] in that it divides the grading into five separate categories: our category of problem description is similar to Docktor and Heller's rubric category of "useful description," explication and justification are similar to "specific application of physics," conceptual understanding is similar to "physics approach," mathematical procedures is a rubric category in both our rubric and the Docktor and Heller rubric, and problem evaluation is similar to "logical progression." The rubric in this study was designed to be more concrete in its application by dividing some of the categories into subcategories and by providing more specification of the categories. Table 5-I also includes how an "expert" grader (e.g., an instructor who is aware that effective grading practices can help foster and support the development of problem solving skills and physics learning and has experience in grading using rubrics that weights the process of solving problem) would apply the rubric to grade the four student solutions. The "expert" scores for the four student solutions rubric scores were determined by four authors grading the four student solutions (SSD, SSE, SSF, SSG) using the rubric and comparing their grading until agreement was reached.

Table 5-I Rubric used to grade SSD and SSE (for Core Problem 1) and SSF and SSG (for Core Problem 2), including scores assigned to the student solutions by expert raters.

Sample Grading Rubric			% (points)	Solution			
				D	E	F	G
Problem Description: Evidence that the students tried to translate the problem statement into terms related to appropriate principles (2 points) (20%)	Diagram clarifying parts of the problem (1 point)	Diagram is comprehensive	+10% (+1 point)			1	
		Diagram is partial	+5% (+0.5 point)	0.5			
		Diagram is not present	+0% (+0 points)		0		0
	Knowns and unknowns are listed, providing evidence of an attempt to plan their problem-solving approach (1 point)	List is comprehensive	+10% (+1 point)			1	
		List is partial	+5% (+0.5 point)	0.5			
		List is not present	+0% (+0 points)		0		0
Explication and Justification of the principles and concepts that are relevant to the analysis of the problem (2.5 points) (25%)	Invoking principle(s) (1.5 point)	Principles that are useful to solve the problem are invoked* (e.g., if two principles are involved, then split scoring for each)	+15% (+1.5 points)	1.5		1.5	
		Principles that are NOT useful to solve the problem are invoked*	0% (+0 points)				
	Justifying principle(s) (1 point)	Principles that are useful to solve the problem are justified with respect to the problem scenario*	10% (+1 points)	1		1	
		Principles that are NOT useful to solve the problem are invoked, however they are justified with respect to the problem scenario*	5% (+0.5 points)				
Conceptual Understanding (3 points) (30%)	Applying principle(s), which provide evidence that the student has an adequate understanding of the relevant principles and concepts (3 points)	Principles applied adequately* (e.g., if two principles are involved, then 15% (+1.5 point) each)	+30% (+3 points)			1.5	
		Principles applied are partially correct* (w/ sign errors, missing terms, etc.)	15% (+1.5 point)	1.5	1.5	0.75	1.5
Mathematical Procedures (1 point) (10%)	Executing the solution by selecting appropriate mathematical procedures and following math rules (1 point)	Algebraic procedures applied adequately	+10% (+1 point)		1		1
Problem Evaluation: Evidence of reflection on the problem-solving process (1.5 point) (15%)	Reasonability check of intermediate target variables and answer, e.g., checking consistency of units, limiting cases, realistic numbers, etc. (1.5 point)	Intermediate target variables and answer are *reasonable* and there is evidence of an *attempt to check the reasonability of the solution*	20% (+2 point, extra credit for checking)				
		Intermediate target variables and answer are *reasonable* and there is *no evidence of an attempt to check the reasonability of the solution*	15% (+1.5 point)	1.5	1.5	1.5	1.5
		Intermediate target variables and answer are *unreasonable* and there *is evidence of an attempt to check the reasonability of the solution and/or student acknowledges that the answer is unreasonable*	20% (+2 point, extra credit for checking)				
		Intermediate target variables and answer are *unreasonable* and *no acknowledgement has been made by student*	+0% (+0 points)				
		Total Possible	100% (10 points)	6.5	4.0	8.25	4.0

*If the problem involves only one physics principle, 15% can be given for invoking the appropriate principle correctly, 10% can be given for correct justification of the principle, and 30% can be given for applying the principle adequately. If the problem involves multiple physics principles, the total percentage possible can be divided among the principles, e.g., if two principles are involved then the student could get 15% for each one he or she applied adequately for a total of 30%.

5.3.2.2 Implementation of the data collection tool

The TA professional development course consisted of two-hour meetings held weekly throughout the fall semester. Three consecutive weekly sessions at the outset of the training course revolved around a group administered interactive questionnaire (GAIQ), encouraging reflection on various facets of teaching problem solving: Designing problems, designing example problem solutions, and grading. Table 5-II shows the sequence of grading activities. The activities served as a data collection tool in order to study TAs' grading decisions and considerations in a simulated environment as well as a learning experience within the training program [20]. Grading data were collected twice, at the beginning and end of the semester.

Table 5-II Sequence of TA grading activities.

Time		Activity
Beginning of Semester	Pre-Class Week 1	Individually, TAs wrote an essay regarding the purpose of grading. They then completed a worksheet which asked them to grade student solutions to problem 1 (see Fig. 5-1) in homework (HW) and quiz contexts, list features of each solution, and explain the reasoning underlying the weight assigned to each feature to arrive at a final score.
	In-Class Week 1	In groups of 3-4, TAs graded the student solutions (SSD and SSE) using a group worksheet and then participated in a whole-class discussion in which the groups shared their grading approaches.
	Pre-Class Week 2	Individually, TAs were given a rubric designed to promote effective problem-solving strategies (see Table 5-I) and asked to regrade SSD and SSE and to grade SSF and SSG using the rubric. They were also asked to list the pros and cons of using this rubric and to discuss how the rubric may be improved.
	In-Class Week 2	TAs completed a worksheet in groups of 3-4 which asked them to compare the scores they assigned to SSD, SSE, SSF, and SSG using the rubric, list the categories they agreed on, list the categories they disagreed on, and decide upon a score for each solution.
End of Semester	Pre-Class	Individually, TAs wrote an essay regarding the purpose of grading. They then completed a worksheet which asked them to grade the student solutions in HW and quiz contexts, list features of each solution, and explain the reasoning underlying the weight assigned to each feature to arrive at a final score.
	Reflection	TAs were given copies of their pre-lesson activities from the beginning of the semester and were asked to make comparisons between their responses on the beginning of the semester pre-lesson activities and the end of semester grading activities.

The GAIQ included several stages (see Table 5-II). First, there was a pre-lesson stage in which TAs wrote an essay responding to the following questions:

1) What, in your view, is the purpose of grading students' work?
2) What would you like students to do with the graded solutions returned to them?
3) What do you think most of them actually do?
4) Are there other situations besides final exams and quizzes in which students should be graded?
5) Does grading serve the same purposes for these situations?

The TAs also filled out a worksheet asking them to compare and make judgments about a set of four student solutions to Core Problem 1 (see Fig. 5-1) in a simulated grading context.

In the pre-class Week 1 activity at the beginning of the semester, TAs were given a homework assignment to individually grade the student solutions for both homework (HW) and quiz contexts out of a total score of ten points, list characteristic solution features, and explain the reasoning underlying the weight assigned to each feature to obtain a final score. The TAs were told to assume that 1) they are the instructors of the class and can structure their grading approaches to improve learning, 2) they have authority to make grading decisions, and 3) they have told students how they would grade. An example response is shown in Fig. 5-5.

Features: Solution E	**Score**		**Reasons: explain your reasoning for weighing the different features to result with the score you arrived at.**
	HW	Q	
The answer is correct, the approach is correct. The steps for getting $v^2=2gh$ are not written	**8**	**10**	I gave this student a lower grade on HW because I think that students have enough time to write down all steps, and they should. This answer looks like it has been written just to get a grade, not that the student was learning something while doing the HW. I think that since the approach and the answer are right, this answer gets a full grade on a quiz.

Figure 5-5 One component of a sample TA's worksheet (transcribed) related to SSE which was part of the pre-class grading activity.

During the in-class Week 1 activity of the GAIQ (see Table 5-II), the TAs worked in groups of 3-4 in which they were asked to discuss and try to reach an agreement regarding grading the student solutions. After they had graded the solutions, a representative from each group shared their grading approaches with the entire class. Two of the authors (E.M. and R.S.) were present in the class. E.M. coordinated the class work and the discussion at the end of the class which highlighted grading approaches that promote effective problem solving using a systematic approach and noted the disadvantages of grading which focused exclusively on correctness. The discussion included listing the grading criteria they used to grade the student solutions and then deciding as a class whether they agreed or disagreed on the appropriateness of these criteria. These criteria include listing initial information, drawing a diagram, proof of understanding, errors in physics reasoning, intermediate steps, correct units, admitting mistakes, etc. R.S. observed and documented the TAs' comments during the group and whole-class discussions. TAs were then given a rubric and were explicitly shown how the categories of the rubric aligned with and incorporated many of the solution features and grading criteria mentioned in the class discussion (e.g., "list" and "diagrams" as initial problem description, "proof of understanding" as explication and justification of physics principles, etc.). Each category of the rubric was explained so that TAs would understand how to apply it appropriately. The TAs were told to assume that they had distributed the rubric to their students and told them that they would be graded using the rubric.

In the pre-class Week 2 activity, as a homework, TAs individually graded SSD and SSE to Core Problem 1 (see Fig. 5-1) using the rubric. The TAs also considered an additional introductory physics problem, Core Problem 2 (see Fig. 5-3), and graded Student Solution F (SSF) and Student Solution G (SSG) using the rubric (see Table 5-I). During the in-class Week 2 activity of the GAIQ (see Table 5-II), the TAs completed a worksheet in groups of 3-4 which asked them to compare the scores they assigned to SSD, SSE, SSF, and SSG using the rubric, list the categories they agreed on, list the categories they disagreed on, and decide upon a score for each solution.

The end of semester activities (pre-class and reflection activities, see Table 5-II) examined the effect of the group and class discussion and use of the rubric on TAs' perceptions and attitudes about grading. The pre-lesson stage of the end of semester task included the same essay and grading activity as in the beginning of semester pre-lesson stage. In class, the TAs were given copies of their pre-lesson activities from the beginning of the semester and were asked to reflect on how their grading approaches evolved throughout the semester. They were also asked to consider changes in their consideration of features in re-grading the student solutions.

5.3.2.3 Post-course interviews

After an initial analysis of the collected data, in the following semester, seven of the TAs in the study volunteered to be interviewed to provide further clarification of their stated grading beliefs (which sometimes appeared to contradict their actual grading practices), to investigate whether the grading activities carried out in the TA training class impacted their beliefs about their grading in some manner not captured in their written responses (overall, there were no significant changes in their written reflection and graded solutions at the end of the semester compared to the beginning of the semester), how they graded in actual courses for which they were TAs and what they thought were the pros and cons of using a grading rubric. The interviewer had some pre-determined questions to ask the TAs (e.g.: What in your view are the pros and cons of grading on a rubric? Have your beliefs about grading changed due to the interventions in the TA professional development course? What caused the change in beliefs?). However, the interviewer also asked additional follow-up questions on-the-spot to examine TAs' reasoning and also to give them an opportunity to clarify their written responses on the GAIQ worksheets if there were any ambiguities in their responses (all TAs had the opportunity to take a look at their responses to the GAIQ at the beginning of the semester before and after being provided the rubric and at the end of the semester without the rubric). Some of the queries included questions about why they graded the short solutions highly (if they did so) even when provided with the rubric to grade the solutions.

5.4 RESULTS

5.4.1 Rubric Grading Results

When given the rubric, the TAs were reminded of the grading features they listed during the class discussion and were shown how the five major categories and corresponding subcategories of the rubric corresponded with those grading features. After a discussion of the rubric, including a clarification of its components, TAs were asked to individually grade SSD and SSE corresponding to Core Problem 1 as well as SSF and SSG corresponding to Core Problem 2.

5.4.1.1 Comparison of analogous solutions SSD and SSF to isomorphic problems

To investigate Research Questions 1, 2, and 3 (see end of Section 5.1), we analyzed how the TAs applied the rubric when grading the student solutions. Fig. 5-6 shows the percentage of TAs who selected each category in the rubric when grading the elaborated solution SSD. Bars that are the same color represent categories that are usually graded in an either/or case for one of the categories. There was consensus among the TAs that principles that were useful to solve the problem were invoked (blue bar) and were justified (brown bar), and that the algebraic procedures were applied adequately (yellow bar). There was less consensus among TAs about whether the diagram in SSD was comprehensive or whether it was partial (red bars) and about whether the list of knowns/unknowns was comprehensive, partial, or missing (orange bars). There was also some disagreement about whether the principles used in SSD were applied adequately or whether the application was only partially correct (green bars) and whether there was evidence of a reasonability check (purple bars). Compared to an "expert" grader, the majority of TA graders agreed with the "experts" in the selection of each category for SSD.

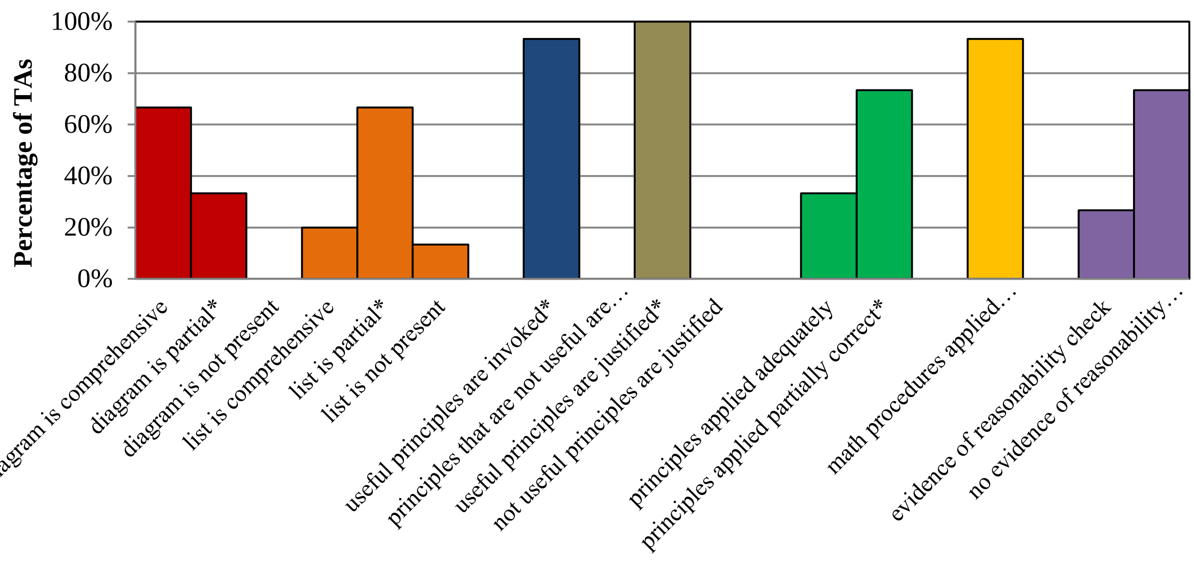


Figure 5-6 Percentage of TAs ($N = 15$) who selected each rubric category when grading SSD. The categories marked with an asterisk represent the category an "expert" grader would choose when grading using the rubric.

Figure 5-7 shows the percentage of TAs who assigned each category in the rubric when grading the elaborated solution SSF (which was analogous to SSD). There was again consensus among the TAs that principles that were useful to solve the problem were invoked (blue bar) and were justified (brown bar), and that the algebraic procedures were applied adequately (yellow bar). There was less consensus among TAs about whether the principles used in SSF were applied adequately or whether the application was only partially correct (green bars), and whether or not there was evidence of a reasonability check (purple bars). Compared with an "expert" grader, the TAs mostly graded SSF as an expert would, though some TAs indicated that there was evidence of a reasonability check in the solution when that was not the case. TAs were generally consistent in applying the rubric for the analogous elaborated student solutions SSD and SSF (i.e., the percentage of TAs grading on components of the rubric for SSD and SSF are similar, as shown by comparing Fig. 5-6 and Fig. 5-7).

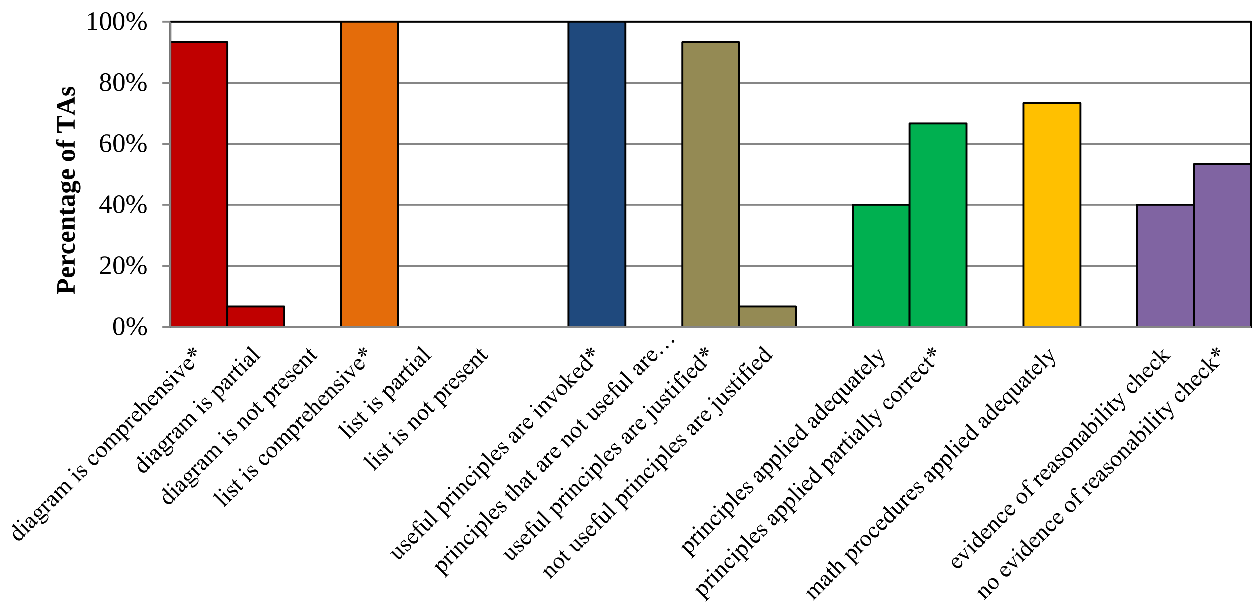


Figure 5-7 Percentage of TAs ($N = 15$) who selected each rubric category when grading SSF. The categories marked with an asterisk represent the category an "expert" grader would choose when grading using the rubric.

5.4.1.2 Comparison of analogous solutions SSE and SSG to isomorphic problems

Figure 5-8 shows the percentage of TAs who assigned each category in the rubric when grading the brief solution SSE (using the same categories and color scheme as Fig. 5-6 and Fig. 5-7). There was consensus among the TAs that the diagram in SSE was not present (red bar) and that the list of knowns/unknowns was also not present (orange bar). TAs were also in agreement that the algebraic procedures were applied adequately (yellow bar), and that there was no evidence of a reasonability check (purple bar). There was less consensus among TAs about whether principles that were useful to solve the problem were invoked (blue bars) and whether the use of those principles was justified (brown bars). As with SSD and SSF, there was also disagreement about whether the principles used in SSE were applied adequately or whether the application was only partially correct (green bars). Compared to an "expert's" use of the rubric, TAs were not in agreement with an "expert" grader when selecting that "useful principles are invoked" and "useful principles are justified." There was no evidence of explicit invoking or justifying of physics in SSE, though the majority of TAs gave this solution credit for those two criteria.

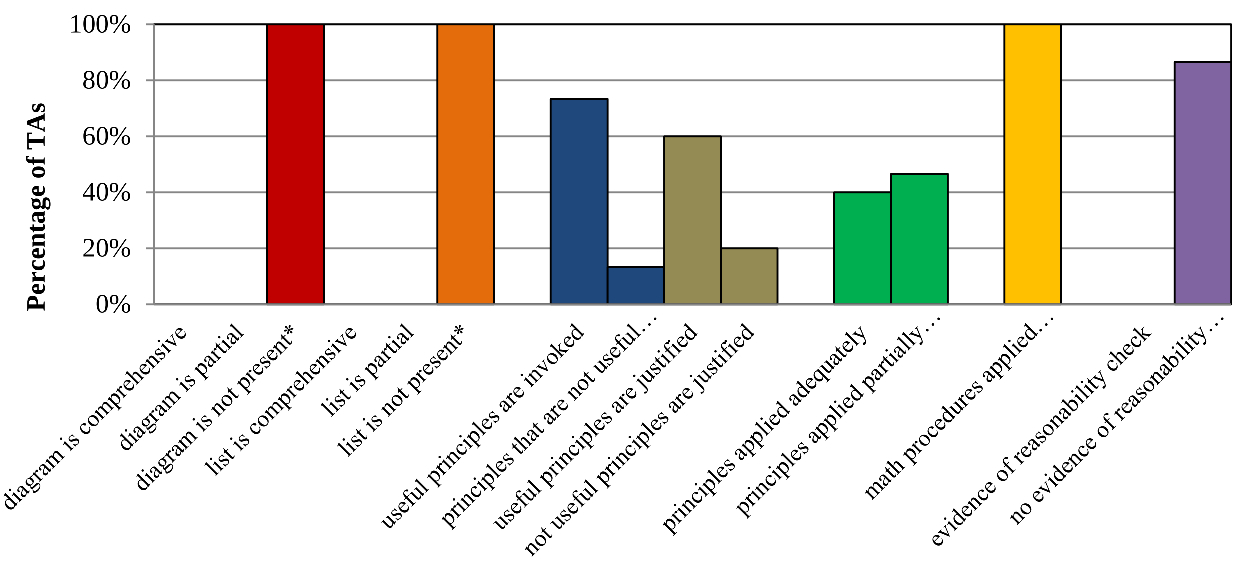


Figure 5-8 Percentage of TAs ($N = 15$) who selected each rubric category when grading SSE. The categories marked with an asterisk represent the category an "expert" grader would choose when grading using the rubric.

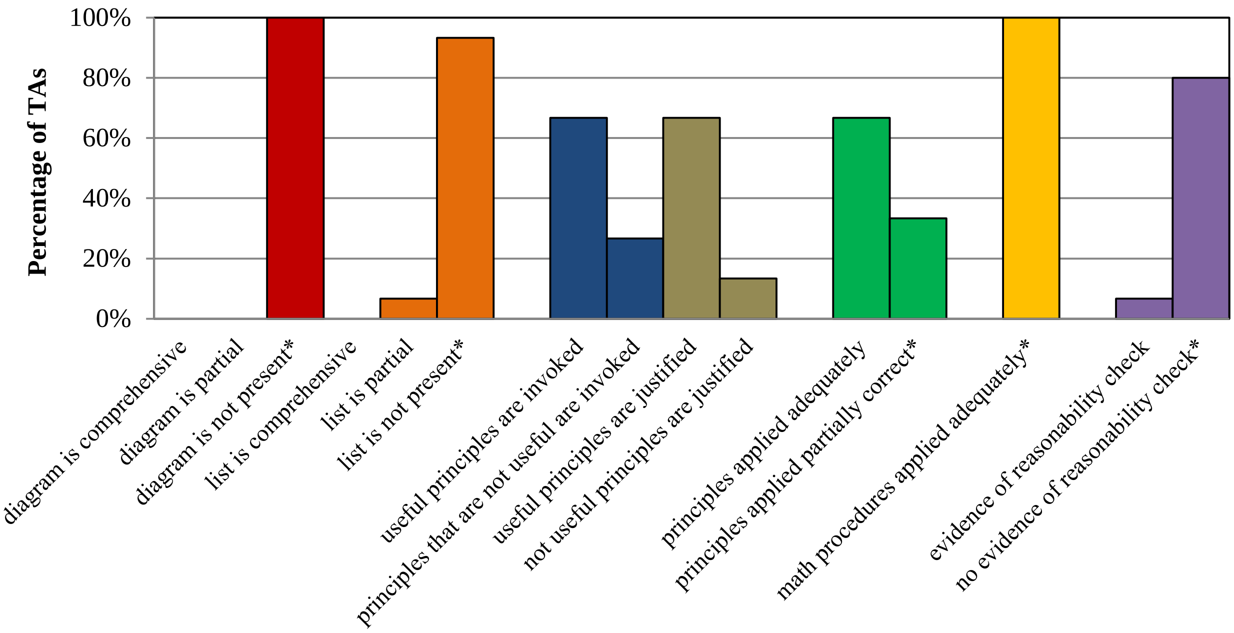


Figure 5-9 Percentage of TAs ($N = 15$) who selected each rubric category when grading SSG. The categories marked with an asterisk represent the category an "expert" grader would choose when grading using the rubric.

Figure 5-9 shows the percentage of TAs who assigned each category in the rubric when grading the brief solution SSG (which was analogous to SSE). There was again consensus among the TAs that the diagram in SSG was not present (red bar), that the algebraic procedures were applied adequately (yellow bar), and that there was no evidence of a reasonability check (purple bar). There was less consensus among TAs about whether principles that were useful to solve the problem were invoked (blue bars) and whether the use of those principles was justified (brown bars). As with all other student solutions, there was disagreement about whether the principles used in SSG were applied adequately or whether the application was only partially correct (green bars). TAs were generally consistent in applying the rubric for the analogous brief student solutions SSE and SSG (i.e., the percentage of TAs grading on components of the rubric for SSE and SSG are similar, as shown by comparing Fig. 5-8 and Fig. 5-9). As with SSE, TAs were again in disagreement with an "expert" grader when selecting that "useful principles are invoked" and "useful principles are justified" in SSG, even though there was no explicit evidence of invoking or justifying in this solution.

5.4.2 Grading Practices After Using Rubric

To investigate Research Question 4 (Do TAs' grading practices change after the intervention?), TAs were given a homework assignment that again asked them to grade SSD and SSE at the end of semester (see Table 5-II). TAs were not given a rubric at this stage, but they were asked to list features of SSD and SSE and explain how they weighed the solution features in grading. In class, TAs were given copies of their pre-lesson activities from the beginning of the semester and were asked to compare their grading from the beginning to the end of the semester and reflect on how their grading approaches evolved throughout the semester. These end of semester grading tasks examined the effect of the group and class discussions and use of rubrics on TAs' approaches to grading.

Table 5-III shows the average score assigned for SSD on the HW and Q contexts before the rubric was introduced, the average score assigned by TAs to SSD using the rubric individually, the score assigned by the authors to SSD using the rubric, and the average score in the HW and Q contexts at the end of the semester after TAs had completed grading activities using a rubric, with standard deviations for each average score and *p*-values for comparison between pre and post scores. The standard deviation of TAs scores for SSD was approximately half as large when grading individually using a rubric compared to grading without a rubric, indicating that the use of the rubric helped TAs achieve greater consistency when assigning scores. A t-test was performed, and the differences in means before and after the grading activities using the rubric were not statistically significant in either the HW or Q context.

Table 5-III Average scores and standard deviations (Std. Dev.) for SSD for the homework (HW) and quiz (Q) context before using the rubric (Pre), when using the rubric to grade (score assigned by experts using the rubric is also shown), and for the HW and Q contexts after using the rubric (Post), with *p*-values for comparison between pre-rubric and post-rubric scores for both contexts.

SSD	Pre-HW	Pre-Q	Rubric	Rubric (Experts)	Post-HW	Post-Q	*p*-HW	*p*-Q
Average	7.40	7.93	7.98	6.50	7.21	8.16	0.845	0.585
Std. Dev.	1.30	1.24	0.70		1.49	1.60		

Table 5-IV shows the average score assigned for SSE on the HW and Q contexts before the rubric was introduced, the average score assigned to SSE using the rubric, the score assigned by the authors to SSE using the rubric, and the average score in the HW and Q contexts at the end of the semester after the rubric intervention, with standard deviations for each score. The standard deviation of TAs scores for SSE was also approximately half as large when grading individually using a rubric compared to grading without a rubric.

Table 5-IV Average scores and standard deviations (Std. Dev.) for SSE for the homework (HW) and quiz (Q) context before using the rubric (Pre), when using the rubric to grade (score assigned by experts using the rubric is also shown), and for the HW and Q contexts after using the rubric (Post), with *p*-values for comparison between pre-rubric and post-rubric scores for both contexts.

SSE	Pre-HW	Pre-Q	Rubric	Rubric (Experts)	Post-HW	Post-Q	*p*-HW	*p*-Q
Average	6.00	7.07	6.07	4.00	6.13	7.65	0.904	0.588
Std. Dev.	3.16	2.71	1.68		2.85	3.10		

Plots of the distribution of TAs' assigned scores to SSD vs. SSE in the quiz context before and after the rubric activities can be found in Appendix D. In particular the distributions were similar before and after using the rubric, which indicates that TAs' scores stayed approximately the same after working on the rubric grading activities. Similarly, the distributions for grading in the homework context before and after the rubric intervention were also very similar.

5.4.3 TAs' Feedback About the Rubric Activity Via Written Responses, Class Discussions, and Interviews

To investigate Research Question 5 (According to the TAs, what are the pros and cons of using a rubric to grade?), part of the assignment to use the rubric to grade SSD and SSE asked TAs to write a short essay in which they listed what they believed to be the pros and cons of using a rubric and identified changes they would make to improve the rubric. The TAs' stated pros and cons for using a rubric were coded to determine if the responses followed any trends. Based upon these trends, categories were created to describe the most common types of responses, as shown in Table 5-V. Two researchers separately coded the responses according to the chosen categories and then compared their individual coding and discussed any discrepancies until an agreement of greater than 90% was reached. In addition, TAs also gave feedback about the pros and cons of using a rubric to grade in class discussions in the professional development course and individual interviews.

Table 5-V also shows the percentage of TAs that mention each category of pros and cons of using a rubric in their written responses (although interviews provided an opportunity for clarification in some cases). The most commonly stated drawback of using a rubric, in TAs' opinions, was that a rubric did not allow for enough flexibility when assigning scores (e.g., to give partial credit in certain cases), or the TAs were uncomfortable in taking off points if the final answer was correct, with 53% of TAs mentioning this negative aspect of using a rubric. In particular, the TAs often felt that they should have the freedom to grade the introductory student solution in a manner they see appropriate based upon their intuition rather than being tied by a rubric. Several TAs mentioned that a rubric is too constraining and they wanted to be able to give a high score to a student whose final answer was correct even if the student did not explicate his

or her problem-solving approach. In an interview, one TA noted that student solutions are too "complicated" to be graded using a rubric. He explained it further with the following statement: "*The answers are not like filling in forms. They're much more interwoven and complicated than that. You cannot really say, 'okay, here we have this, so one point to that.' That's not true in the real case. So I just read it (the rubric) and got some idea out of it, but didn't really follow every instruction.*" Individual interviews and class discussions in the professional development class suggest that this type of feeling was common amongst other TAs as well.

In a different interview, another TA was concerned about the fact that a rubric may restrict creativity, stating, "*I think the rubrics are a little too specific, because I don't think you can categorize everything just by writing a rubric. It's hard to really balance the creativity part of students going to their correct solution. So the rubric kind of is very harsh tool to say, 'okay, these are the correct solutions, and these are not.' Which, personally, I think is against the spirit of education itself.*" This same TA even mentioned that rubrics "*make the whole class boring, make physics boring.*" One TA stated that even if a student has an incorrect solution, "*if any student gives interesting idea in solving problems, we should give them extra points to encourage students to think.*" Further discussions suggest that this TA thought that by following the criteria on a rubric, all students would be forced to solve a problem using the same approach. These types of feelings about a rubric are interesting considering the fact that the rubric the TAs were provided is not constraining in terms of how students approach a physics problem or which physics principles they use to solve the problem so long as they show the problem-solving process used.

Table 5-V Explanation of each category used for coding TAs' stated pros/cons when grading student solutions using the rubric provided, with percentage of TA responses mentioning each category in their written responses.

Code	Definition	Examples	% of TAs
(Pro) Fairness/ consistency	Using the rubric makes the grading process fairer for students. When grading with a rubric, graders are more consistent with their scores.	--"This rubric will give a standard on how to grade, it is very useful to make a just assessment." --"Evaluate the exams and homework fairly." --"reduced the fluctuations of a grader."	40%
(Pro) Easier to grade	The rubric makes the grading process easier for the graders.	--"The pros are that it is easier to grade." --"Easier for partial marking for incomplete answers." --"Easier to point out mistakes."	13%
(Pro) Encourage students to use good practices	The rubric encourages students to use effective problem-solving strategies and practices, such as drawing a diagram and justifying their use of physics principles.	--"Encourage students to follow a procedure for problem solving." --"They learn better strategies for problem solving." --"This rubric favors the solutions that show explication and justification of the principles and concepts, which will help students pay more attention to linking the specific physical scenario with the physical theories."	73%
(Pro) Identify specific difficulties	Grading with the rubric helps the grader/instructor to identify students' specific difficulties with the material.	--"The teacher can understand at what part of the problem most students are making a mistake and he can focus on that more." --"Make it easier for student to get feedback." --"I do think this would be helpful for instructors, since it would be easy for an instructor to look across the grades by rubric and see where students most often lost points."	33%
(Con) Lack of flexibility/discomfort in taking off points if final answer is correct	Using the rubric leads to less flexibility when grading. The graders have less freedom to assign points the way they would like to.	--"The rubric doesn't allow for much nuance. A solution that is really good may not exactly hit the mark on every category, but the student may have still demonstrated their understanding." --"Over formatting/ kill diversities (of student responses to score points, e.g., short and long solutions could both be worthy of high points if they are both correct)." --"A con is that partial credit may be harder to come by (for what I want to give them points for, e.g., more points for the correct final answer)."	53%
(Con) More time-consuming	Use of the rubric would require either students or graders to spend more time on the problem.	--"Forces students to spend more time on (solving each problem) --"Takes more time to evaluate."	20%

Some TAs were also concerned about whether the rubric would be more time-consuming, either for the students, who would be required to include details such as diagrams and justifications for their work, or for the TAs, who would be required to evaluate additional aspects of the student solutions (mentioned by about 20% of the TAs). For example, in an interview one TA said, "*I think in the real world, TAs and graders don't really have much time to look at everything students write, so I think it's important to be concise and write down all that is needed and not more.*" This same TA also mentioned that requiring students to spend more time on the process may be unnecessary, stating: "*The process is one factor, but it's not really that important… I think in most practical cases, the correct answer should be more important than (the process)… that people think (may be important).*" Several TAs explicitly mentioned that they had seldom been penalized for not showing the process in their own courses and did not feel comfortable taking off points if the final answer was correct.

An issue that several TAs mentioned in interviews (but not in their written list of pros/cons) was that they may not use the rubric especially in the quiz which has time constraint if they can infer student understanding from looking at a student's solution. For example, one TA stated: "*When students take a quiz, I know that he's not cheating so he knows the answer, but maybe he's stressed or trying to do it really fast, so he did part of it in his mind. I'm sure that he did the right thing for the quiz so I gave him the full grade for the quiz.*" Another TA mentioned that he identifies with students who write brief solutions, stating: "*In my past I've usually answered questions in that form [of a brief solution like SSE], so I guess I can understand what students are trying to say when they write things like that.*" This TA was among those that gave SSE and SSG (brief solutions) credit for justifying the use of invoked physics principles when grading with the rubric even though there was no explicit evidence of justification in those student solutions.

Some TAs also mentioned in interviews (but not in their written list of pros/cons) that in their opinion, grading should only serve a summative purpose. For example, according to one interviewed TA, "*it is up to the students to get something out of their solution and student learning should not depend upon the TAs' grading practices.*" This TA believed that assigning of points to features such as diagrams and lists of unknown variables was merely "sugar coating" the students' scores, i.e., assigning points that inflated student scores and simply helped the students get a better grade but did not help them learn physics. This TA and several others felt that significantly more points should be given for the correct final answer than what the rubric given to them asked them to give. Some other TAs also had similar views about the "triviality" of grading students on their initial qualitative analysis of the problem such as drawing a diagram. Despite class discussion in the TA professional development course, they were not convinced that any student who drew a diagram and wrote down knowns and unknowns but did not obtain the correct answer should be given any more points than another student who skipped those qualitative analysis and planning stages of problem solving and got the incorrect answer.

Although the BOP seems to be deeply ingrained, individual interviews with the TAs suggest that some TAs' beliefs about grading may have been positively impacted by the grading activities alongside their teaching responsibilities even though this change was not reflected in their grading practices at the end of the semester. Some TAs stated that they initially were grading based completely on their intuition, but that the rubric helped them grade more fairly. For example, in an individual interview, one TA stated, "*in the start when I was asked to grade these (student solutions) it was just my subjective knowledge…but when you give me a rubric I will stick to the rubric and evaluate the performance based on that. Rubrics helped me because when you have a whole class you're doing justice to all of them.*" Another interviewed TA stated, "*(At first) I was going by my basic intuition… this whole semester was a learning curve for me, and as I progressed I learned a lot.*" This TA was happy that he at least knew that he could use a rubric to grade students objectively for any problem (whether he would always use a rubric to grade students for all problems was unclear).

Even though there was little change in TAs' grading of solutions SSE and SSD, individual interviews and class discussions indicate that there may be a ray of hope in that at least some TAs had started to think about the impact of grading students on their problem-solving processes. For example, one TA stated, "*before taking this course I mostly just looked at the answer and if it's right then good, if it's not okay then you don't get anything, but after the course I started to know that you need to look at the process.*" Another TA stated, "*before the rubric I was just paying attention to small details, but after the rubric there's lots of things I have to be careful about when grading … (for example) I wasn't giving any points for diagrams.*" Some of these TAs also mentioned that they are gradually realizing that a brief solution does not necessarily demonstrate that the student understands the concepts. The fact that the TA professional development class in which the TAs did the grading activities was running parallel to their actual teaching helped some of the TAs at least begin to start thinking about the importance of the problem-solving process. For example, one TA stated that the grading activities in the class helped but simultaneous experience with the students in the classes they were teaching also helped: "*When I (initially) see this (short solution), I think, 'he knows what he's doing.' But when I interacted with the students, I saw that sometimes they actually write things and they have NO idea what they're doing, they just know this equation and just go through it. That interaction helped me to understand that the students might sometimes not know what they're doing.*" He noted that after interacting with the students in the class he began to understand why what was discussed in the TA professional development course regarding grading students on the process of problem solving and not just the final answer was important.

In addition, a few TAs stated that the rubric activity affected their grading approaches in actual classroom settings. For example, one interviewed TA noted that he understands the importance of grading for the process and stated, "*if I was given the chance (in my own grading), I would prepare a rubric and I would have my solutions so for each question the scores would be much more distributed (rather than all or nothing).*" Another interviewed TA stated, "*overall, I like this idea of breaking down the marks with a rubric, so when I'm not provided with a rubric I will try to make a reasonable breakdown in my mind and I will try to break them according to that one, so in that sense I would say I like this (rubric).*" Thus, even though the grading activities with a rubric that emphasized the process of problem solving did not necessarily show discernable changes in their grading at the end of the TA professional development course, discussions with the TAs suggest that at least some of them were contemplating the benefits of grading that emphasizes the problem-solving process. The fact that at least some TAs were paying more attention to grading on the process is somewhat encouraging.

5.5 DISCUSSION AND CONCLUSIONS

In this investigation of physics graduate TAs' beliefs and practices regarding introductory physics grading, the TAs were initially asked to grade an elaborated solution which revealed two canceling mistakes (SSD) and a brief solution with no elaboration (SSE) without a rubric. They then completed a grading activity involving the use of a rubric to grade those solutions as well as two analogous solutions to isomorphic problems. They were told to assume that they had distributed the rubric and had full control of how to grade student solutions.

However, despite class discussions about how effective grading practices can promote good problem-solving approaches and aid in the development of physics content knowledge, many TAs gave students the benefit of the doubt in grading solutions in which the problem-solving process was not explicated even when a PER-inspired rubric was given to them to grade student solutions. In particular, the TAs did not use the rubric as intended to grade solutions in which the final answer was correct but the problem-solving process was not explicated. For example, approximately 60% of the TAs who were asked to use the rubric to grade the short solutions in which the problem-solving process was not articulated and justified claimed that the physics principles were invoked and justified appropriately in the solutions (SSE and SSG). Interviews suggest that the TAs felt that they should not take off too many points when the final answer is correct but the problem-solving process is not shown. In other words, the TAs put the BOP for justifying such short solutions on themselves. Interviews also suggest that the TAs were very reluctant to take off too many points for the short solutions (that did not explicate the problem-solving process) and they found the rubric to be constraining and rigid for grading such a short solution that had the correct final answer. They felt the rubric should be "subtractive" in that it should take off points for mistakes that students make but not penalize students if there are no visible mistakes and the final answer is correct.

Furthermore, at the end of the semester TAs were again asked to grade without a rubric a solution in which the problem-solving process was explicated (SSD) and a brief solution with no elaboration (SSE). Comparing the grading of SSD and SSE at the beginning of the semester to the end of the semester, there was little change in the scores given to the solutions and approximately half of the TAs gave the brief solution SSE a score greater than or equal to the elaborated solution SSD. In other words, TAs continued to give benefit of the doubt to the short student solution and did not penalize a student if he/she had not articulated and justified the principles used in the solution. Thus, the TAs had difficulty giving appropriate weight to the process of problem solving and focused more on the correctness of the final answer, even after class discussions and using a rubric that was inspired by PER. This was true even though there was extensive class discussion in the professional development course after the activities at the beginning of the semester about the benefits of using a good rubric to help students learn.

We also found that when using the rubric to grade student solutions, TAs applied the rubric consistently across analogous student solutions for isomorphic problems (i.e., analogous solutions SSD and SSF in which the problem-solving process is explicated and analogous solutions SSE and SSG, in which the problem-solving process is not explicated). This consistency in grading across analogous solutions to isomorphic problems (even though students did not use the rubric as they were instructed to do to grade, e.g., brief solutions SSE and SSG in which the problem-solving process was not explicated) indicates that TAs may hold some prior conceptions about grading (in particular, belief that students who have the correct final answer should not be penalized significantly) and apply these ideas consistently across different student solutions for similar types of responses.

In summary, we find that a one-semester intervention with instructional activities focused on helping graduate TAs discern the value of using a rubric emphasizing the problem-solving process was not sufficient to change where they place the BOP and did not result in measurable changes in TA grading practices. Although the BOP seems to be deeply ingrained, individual interviews with the TAs suggest that at least some TAs' beliefs about grading may have been somewhat positively impacted by the grading activities alongside their teaching responsibilities even though this change was not reflected in their grading practices at the end of the semester.

5.5.1 Possible Reasons for the Lack of Change in TAs' Grading Practices

In this section we discuss some possible reasons for why significant changes in TAs' grading practices from the beginning to the end of the semester were not observed after the rubric activities focusing on the problem-solving process in the TA professional development course. We intend to test these in future research.

1) Some TAs were uncomfortable placing BOP on students partly because they have themselves seldom been penalized for not showing the process if their final answer is correct for the majority of their experiences as students.
2) Some TAs did not like the rubric given to them in the TA professional development course or did not like rubrics in general for various reasons.
3) Some TAs did not internalize that grading can serve as a formative assessment for students even though there was extensive discussion about it in the professional development course.
4) Some TAs may have remained in a state of cognitive conflict in terms of their grading practices in that they realized it may be valuable to grade students on the process of problem solving but they had not fully resolved to grade on the problem-solving process when a student's final answer was correct.

Regarding the burden of proof of understanding, interviews, TAs' written work, and class discussions suggest that even at the end of the 15 week semester, some TAs continued to infer information from introductory physics student solutions which was not explicitly stated. In fact, even when TAs were given a rubric which included criteria for invoking and justifying physics principles, a majority of TAs were willing to give the short solutions, SSE and SSG (for Core Problems 1 and 2, respectively) credit for justifying principles even though those solutions did not contain any form of explicit justification. As noted, some TAs explicitly noted that they were uncomfortable placing the BOP for explicating the problem-solving process and demonstrating understanding on the student because they themselves wrote brief solutions and expected to get full scores if the final answer was correct in their own course work most of the their lives. Since it is unlikely that most TAs have been penalized for not showing proof of understanding in their solutions in their courses, they may empathize with their students for using a similar approach. They may read between the lines and assume that they understand what their students know when their solutions do not show the problem-solving process but have the correct final answer. Individual interviews with some of the TAs confirms this hypothesis.

Some other possible reasons why TAs' grading practices did not change as a result of the activities involving the grading rubric (emphasizing the problem-solving process) was that the TAs may not have liked the rubric, did not engage with it effectively, or did not deeply contemplate the class discussions about why such a rubric is useful. Some TAs claimed that rubrics are too restrictive, either for the graders or for the students. In individual interviews, some TAs did not seem to acknowledge that the rubric they were given can account for many different methods of obtaining a correct answer even though it weighted the problem-solving process much more heavily than the final answer. Some of them explicitly noted that they did not want to penalize students who had not explicated the problem-solving process but had the correct final answer so they did not like the rigidity of the rubric. In their view, those students who had the correct final answer knew how to solve the problem and should not be penalized for not showing their work. Therefore, they often ignored the rubric in such cases.

Other TAs noted that they did not in general like rubrics that weighed the problem-solving process more heavily than the final answer because they felt that such rubrics give extra points to students for things that are unimportant and do not show understanding (e.g., drawing a diagram). Despite extensive class discussions, some TAs were not convinced that writing such detailed solutions, which had diagrams or known and unknown variables written explicitly, help students become better at problem solving. They felt that assigning points to features such as diagrams and lists of unknown variables was merely inflating the students' scores, i.e., assigning points that simply helped the students get a better grade but did not help them learn physics or develop good problem-solving skills. These TAs felt that significantly more points should be given for the correct final answer because the purpose of grading was to see if the students knew how to solve the problem correctly and arrive at the final answer. Other TAs stated that grading using a rubric is too time consuming for the students (because they have to spend more time writing down their process and explanations) and the graders (because they have to spend more time grading on the process and explanations as opposed to only checking that the answer is correct). Therefore, they preferred to use their intuition to grade rather than using the rubric.

Some TAs indicated in the interviews, in-class discussions, and in end-of class discussions with the course instructor that they remained unsure about the purpose of grading students on the process rather than the final answer and continued to hold the belief that the primary purposes of grading are to assess student understanding and assign a grade, i.e., grading serves mainly a summative purpose. This apparent disbelief in the potential for grading to serve as a formative assessment for students despite extensive class discussions on this issue may have led some TAs to grade the short solution SSE using less stringent criteria and to assign a larger portion of credit to the correct answer. Despite the course instructor trying to convince them otherwise via class discussions and reflections, TAs often noted that if the students performed poorly because they did not know how to arrive at the correct final answer, they should realize that they need to start working harder. They did not think that it was the TAs' job to expect detailed solutions from students that explicate the problem-solving process in order to help students learn. These TAs often drew parallels between what they would do when they performed poorly and what their students should do if they performed poorly. In the view of many of the TAs, it was not the systematic approach to problem solving that was important when grading but rather whether the students had arrived at the correct final answer.

Polling of 20 faculty members at the same university suggests that except for exams, very few physics instructors require that their TAs use rubrics to evaluate their students on a regular basis in homework and quizzes. Most faculty members were not as concerned about the TA grading on quizzes and homework because they also did not view grading as serving a formative purpose. They noted that student grades were mostly determined by their exam performance so using a rubric for exam grading was useful for fairness and consistency in assigning scores.

It is also possible that some TAs were impacted by the grading rubric intervention but this impact was not reflected in their grading practices at the end of the semester because they were in a state of cognitive conflict and it was challenging for them to assimilate what they had learned in the TA professional development class with their views about grading that they had held for a long time as students. Similar findings have been reported in the context of learning rules, e.g., for balancing, in which students have difficulty taking into account the impact of both lever arm and the weights hanging from the two sides [53]. It was found that the students were in a "mixed" state even after several rounds of intervention and sustained intervention was needed to help them internalize the rules [53]. TAs, in general, seemed unfamiliar with the concept of using a rubric that focuses on process to grade students in order to enhance their learning (except to give partial credit to students for fairness) and found it difficult to accept and apply what the professional development course emphasized (which is that solutions that did not explicate the problem-solving process should be penalized). After being exposed to the rubric and discussing the pros/cons of the rubric, TAs could either dismiss the rubric completely, accept and internalize the rubric, or remain in a “mixed” state [53] in which they may recognize the formative benefits of grading using the rubric but do not necessarily resolve to use the rubric when grading (especially, when explication of the problem-solving process was missing but the final answer was correct). Our written and oral data from class discussions and individual interviews suggest that while some TAs may have dismissed the rubric provided (or the idea of a rubric altogether, preferring to use their intuition alone to grade), others may have needed more time to internalize the rubric since it was an unfamiliar grading tool and penalizing students for the process when the final answer was correct was too discomforting for them.

Even though changes in TAs’ grading practices were not apparent, some TAs indicated in interviews that they were still contemplating the value of grading students for the process of problem solving as a result of the rubric activity several weeks after the TA professional development course was over. These TAs may need more time and more exposure to reflect on the benefits of the rubric. It is possible that with more time and exposure to reflect on the formative benefits of grading using a rubric that explicates the problem-solving process, they would realize that rubrics which focus on the process of problem-solving can help students develop effective problem-solving approaches and learn physics.

5.5.2 Does Convincing TAs to Shift Burden of Proof onto Students Involve a "Paradigm Shift"?

Interviews and class discussions suggest that most TAs, who had not shown the process of arriving at a final answer in their own solutions in the past, had generally managed to get full scores if their answers were correct. Interviews and class discussions also suggest that for many of these TAs, throughout their education, their solutions have been graded based upon correctness only, and there was often an unspoken sense of shared expectation that if the final answer is correct, the student must know how to solve the problem correctly. In class discussions and interviews, the TAs mentioned that students who write short solutions are generally likely to be "brilliant" students who can do the problem in their heads. Even asking the TAs to contemplate situations in which a student copied the final step from the student next to him or her was not sufficient to convince the TAs that they should penalize students for not explicating the problem-solving process. Moreover, trying to convince the TAs that if all students are given the grading rubric that penalizes students for not explicating the problem-solving process, there will not be any excuse for students not to show their work (regardless of whether they could do the problem in their heads) did not seem to convince most TAs in the professional development course. It appears that, partly because of their own past experiences, it may be particularly difficult for a majority of TAs to change their beliefs about grading and their grading practices and discern the benefits of using a rubric that focuses on the problem-solving process to grade their students' solutions.

In summary, this case study suggests that the shift from the focus on the correctness of the solution to the process in students' solutions while grading can be a difficult leap for many TAs to make. TAs have seldom been penalized themselves for it, and they also want to avoid student complaints about their grading particularly because the grading rubric promoted a grading approach which was not the "norm." Furthermore, the time required on the parts of the students to write a detailed solution to each problem and the time required for the TAs to grade them was also a concern despite the fact that the discussions in the TA professional development class focused on the formative benefits of placing the BOP of explicating the problem-solving process on students.

Based upon individual interviews and class discussions with the TAs, we propose that the views about grading may be so ingrained in many TAs' minds that the challenges in helping TAs focus on the benefits of the process as opposed to the correctness of the final answer are somewhat similar to the challenges in helping introductory physics students learn Newtonian physics and overcome their prior naïve conceptions related to force and motion. The "paradigm shift" [54] from naïve notions to Newtonian physics related to force and motion makes the transition to Newtonian thinking very challenging for many introductory physics students, especially because these naïve notions have become highly ingrained over a long period of time trying to (often implicitly) make sense of everyday experiences. Similarly, the "paradigm shift" from grading mainly on the correctness of the final answer to grading on the process (involving initial qualitative analysis of the problem, planning and decision making, reflection, etc.) may be challenging for most TAs, especially because most TAs have strongly ingrained views about grading that have been developed over a long period of time. In fact, we hypothesize that it may possibly be even more challenging to observe discernible changes in TAs' grading practices than changes in introductory students' naïve notions about force and motion. In particular, the established laws that govern the physical universe are encapsulated in compact mathematical forms and an instructor can help students learn to unpack them to make sense of physical phenomena related to force and motion without ambiguity even though it becomes more challenging due to students' naïve prior notions. On the other hand, there are no mathematical laws that govern how one should grade effectively (or, for that matter, how to teach effectively) in order to enhance student learning. Since grading is a more subjective activity than learning about how to make sense of force and motion based upon the established laws of physics, it may turn out to be more challenging to establish guiding principles for grading effectively to maximize student learning and convince TAs to change their grading practices based upon those guidelines (e.g., that place more emphasis on the process of problem solving and less on the final answer). The challenges in changing the grading practices of the TAs despite extensive activities and discussions in this investigation attest to this difficulty.

5.5.3 Implications

Leaders of the professional development courses/programs for physics graduate TAs and physics education researchers can take advantage of the findings of this study. Future studies can build on this research and investigate strategies to get buy-in from the TAs so that they consistently use a rubric and value a rubric that appropriately weights the process of problem solving. Helping TAs value and grade students' solutions on the process of problem solving requires extended time, discussion, support, feedback and practice. The professional development courses/programs can allow more time and support for the TAs to internalize how grading can be used for formative assessment and the fact that grading using a well-designed rubric can support students in developing better problem solving practices and learning physics better. It may be helpful to have the TAs use the rubric they were provided to grade students' solutions in the recitations that they are teaching in a particular semester and track the changes in the problem solving practices and learning over the semester for those students. Over time, this practice may allow the TAs to observe how a good rubric can make grading more objective and encourage students to adopt effective problem-solving strategies. In addition, physics instructors who supervise graduate TAs can collaborate with their TAs in creating grading rubrics, since taking part in developing a rubric can get them to think more deeply about the value of a rubric. It is possible that the TAs will then begin to gradually make a transition and view grading using a rubric as a means to support student learning in addition to all of its other benefits.

5.6 ACKNOWLEDGEMENTS

We thank the members of the physics education research group at the University of Pittsburgh as well as the TAs involved in this study.

5.7 CHAPTER REFERENCES

1. E. Yerushalmi, C. Henderson, K. Heller, P. Heller, and V. Kuo, Physics faculty beliefs and values about the teaching and learning of problem solving. I. Mapping the common core, Phys. Rev. ST PER **3,** 020109 (2007).

2. K. Heller and P. Heller, *The Competent Problem Solver for Introductory Physics* (McGraw-Hill, New York, NY, 2000); 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).

3. A. Van Heuvelen, Learning to think like a physicist: A review of research based instructional strategies, Am. J. Phys. **59**, 891 (1991).

4. F. Reif, Systematic problem solving, in *Applying Cognitive Science to Education: Thinking and Learning in Scientific and Other Complex Domains* (MIT Press, Cambridge, MA, 2008), pp. 201-227; J. Heller and F. Reif, Prescribing effective human problem solving processes: Problem description in physics, Cognition and Instruction **1**, 177 (1984); F. Reif, Millikan Lecture 1994: Understanding and teaching important scientific thought processes, Am. J. Phys. **63**, 17 (1995).

5. J. Mestre, J. Docktor, N. Strand, and B. Ross, Conceptual problem solving in physics, in *Psychology of Learning and Motivation*, edited by J. Mestre and B. Ross (Academic Press, Vol. 55, 2011), pp. 269-298.

6. C. Singh, When physical intuition fails, Am. J. Phys. **70**(11), 1103 (2002).

7. D. Maloney, Research on problem solving: Physics, in *Handbook of Research on Science Teaching and Learning*, edited by D. Gable (MacMillan, New York, NY, 1994), pp. 327-256.

8. W. Leonard, R. Dufresne, and J. Mestre, Using qualitative problem solving strategies to highlight the role of conceptual knowledge in solving problems, Am. J. Phys. **64**, 1495 (1996).

9. T. Nokes-Malach, K. Van Lehn, D. Belenky, M. Lichtenstein, and G. Cox, Coordinating principles and examples through analogy and self-explanation, Eur. J. Psych. Educ. **28**, 1237 (2013).

10. R. Atkinson, A. Renkl, and M. Merrill, Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps, J. Educ. Psychol. **95**, 774 (2003).

11. J. Pellegrino, N. Chudwosky, and R. Glaser (Eds.), *Knowing What Students Know: The Science and Design of Educational Assessment. Committee on the Foundations of Assessment* (National Academy Press, Washington, DC, 2001).

12. P. Black and D. William, Assessment and classroom learning, Assessment in Education **5**(1), 7 (1998).

13. A. Schoenfeld, When good teaching leads to bad results: The disasters of “well-taught” mathematics courses, Educational Psychologist **23**(2), 145 (1988).

14. A. Elby, Another reason that physics students learn by rote, Am. J. Phys. **67**(7), S52 (1999); T. Crooks, The impact of classroom evaluation practices on students, Review of Educational Research **58**(4), 438 (1988).

15. T. Angelo and K. Cross, *Classroom Assessment Techniques: A Handbook for Faculty* (National Center for Research to Improve Postsecondary Teaching and Learning, Ann Arbor, MI, 1998).

16. A. Schoenfeld, Toward a theory of teaching-in-context, Issues in Education **4**(1), 1 (1998).

17. A. Brown, Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings, Journal of the Learning Sciences **2**, 141 (1992).

18. A. Collins, Toward a design science of education, in *New Directions in Educational Technology*, edited by E. Scanon and T. O'Shey (Springer, New York, NY, 1992), pp. 15-22.

19. P. Cobb, J. Confrey, A. diSessa, R. Lehrer, and L. Schauble, Design experiments in education research, Educational Researcher **32**(1), 9 (2003).

20. E. Yerushalmi, C. Henderson. W. Mamudi, C. Singh, and S. Lin, The group administered interactive questionnaire: An alternative to individual interviews, *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE*, edited by S. Rebello, C. Singh, and P. Engelhardt (AIP Conf. Proc., Melville, NY, 2012), p. 97.

21. S. Lin, C. Singh, W. Mamudi, C. Henderson, and E. Yerushalmi, TA-designed vs. research-oriented problem solutions, *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE*, edited by S. Rebello, C. Singh, and P. Engelhardt (AIP Conf. Proc., Melville, NY, 2012), p. 255.

22. E. Yerushalmi, E. Marshman, A. Maries, C. Henderson, and C. Singh, Grading practices and considerations of graduate students at the beginning of their teaching assignment, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN*, edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2015), p. 287.

23. C. Henderson, E. Marshman, A. Maries, E. Yerushalmi, and C. Singh, Instructional goals and grading practices of graduate students after one semester of teaching experience, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN*, edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2015), p. 111.

24. D. Schwartz, J. Bransford, and D. Sears, Efficiency and innovation in transfer, in *Transfer of Learning From a Modern Multidisciplinary Perspective*, edited by J. Mestre (2005), p 1.

25. G. Hatano and K. Inagaki, Two courses of expertise, in *Child Development and Education in Japan*, edited by H. Stevenson, J. Azuma, and K. Hakuta (W. H. Freeman & Co., New York, 1986), pp 262–272.

26. P. Black and D. William, Inside the black box: Raising standards through classroom assessment: Formative assessment is an essential component of classroom work and can raise student achievement, Phi Delta Kappan **92**, 81 (2010).

27. P. Black and D. William, Developing the theory of formative assessment, Educational Assessment, Evaluation and Accountability **21**(1). 5 (2009).

28. A. Schoenfeld, Toward a theory of teaching-in-context, Issues in Education **4**(1), 1 (1998).

29. D. William, Formative assessment: Getting the focus right, Educational Assessment **11**, 283 (2006).

30. P. Black, C. Harrison, C. Lee, B. Marshall, and D. William, *Assessment for Learning: Putting It Into Practice* (Open University Press, Buckingham, 2003).

31. E. Yerushalmi, E. Cohen, A. Mason, and C. Singh, 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).

32. B. White and J. Frederiksen, Inquiry, modeling, and metacognition: Making science accessible to all students, Cognition and Instruction **16**(1), 3 (1998).

33. A. Mason and C. Singh, Helping students learn effective problem solving strategies by working with peers, Am. J. Phys. **78**, 748 (2010).

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

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

36. C. Henderson, E. Yerushalmi, V. Kuo, P. Heller, and K. Heller, Grading student problem solutions: The challenge of sending a consistent message, Am. J. Phys. **72**, 164 (2004).

37. H. Andrade, Understanding rubrics, Educational Leadership **54**(4), 44 (1997); H. Andrade, Using rubrics to promote thinking and learning, Educational Leadership **57**(5), 13 (2000); H. Andrade, Teaching with rubrics: The good, the bad, and the ugly, College Teaching **53**, 27 (2005).

38. S. Brookhart, *How to Create and Use Rubrics for Formative Assessment and Grading* (ASCD, Alexandria, VA, 2013).

39. E. Panadero and J. Anders, The use of scoring rubrics for formative assessment purposes revisited: A review, Educational Research Review **9**, 129 (2013).

40. M. Heritage, J. Kim, T. Vendlinski, and J. Herman, From evidence to action: A seamless process in formative assessment?, Educational Measurement: Issues and Practice **28**(3), 24 (2009).

41. G. Taggart, S. Phifer, J. Nixon, and M. Wood, *Rubrics: A Handbook for Construction and Use* (R&L Education, Lanham, MA, 2005).

42. J. Docktor and K. Heller, Robust assessment instrument for student problem solving, in *Proceedings of 82nd NARST Annual International Conference*, 2009 (see

http://groups.physics.umn.edu/physed/People/Docktor/talks_papers/Docktor_NARST09_paper.pdf).

43. For example, see http://www.aps.org/programs/education/graduate/conf2008/index.cfm.

44. L. McDermott, Millikan lecture 1990: What we teach and what is learned—closing the gap, Am. J. Phys. **59**, 301 (1991).

45. F. Lawrenz, P. Heller, and R. Keith, Training the teaching assistant, J. Col. Sci. Teach. **22**(2), 106 (1992).

46. E. Jossem, Resource Letter EPGA-1: The education of physics graduate assistants, Am. J. Phys. **66**, 502 (2000); C. Sandifer and E. Brewe (Eds.), *Recruiting and Educating Future Physics Teachers: Case Studies and Effective Practices* (American Physical Society, PhysTEC, 2015); A. Maries and C. Singh, Performance of graduate students at identifying introductory physics students' difficulties related to kinematics graphs, *Proceedings of the 2014 Phys. Ed. Res. Conference*, *Minneapolis*, *MN*, edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2015), p. 171 ; Improving one aspect of pedagogical content knowledge of teaching assistants using the TUG-K, Phys. Rev. ST PER **9**, 020120 (2013); Teaching assistants' performance at identifying common introductory student difficulties in mechanics revealed by the Force Concept Inventory, Phys. Rev. PER (2016).

47. C. Singh, Categorization of problems to assess and improve proficiency as teacher and learner, Am. J. Phys. **77**, 73 (2009).

48. C. Singh, Rethinking tools for training teaching assistants, *Proceedings of the 2009 Phys. Ed. Res. Conference, Ann Arbor, MI*, edited by M. Sabella, C. Henderson, and C. Singh (AIP Conf. Proc., Melville, NY, 2009), p. 59.

49. S. Lin, C. Henderson, W. Mamudi, E. Yerushalmi, and C. Singh, Teaching assistants' beliefs regarding example solutions in introductory physics, Phys. Rev. ST PER **9**, 010120 (2013).

50. See https://www.aip.org/statistics/data-graphics/demographic-profile-physics-phds-classes-2010-2011-2012-combined.

51. E. Yerushalmi and B. Eylon, Supporting teachers who introduce curricular innovations into their classrooms: A problem-solving perspective, Phys. Rev. ST PER **9**, 010121 (2013).

52. M. Hull, E. Kuo, A. Gupta, and A. Elby, Problem-solving rubrics revisited: Attending to the blending of informal conceptual and formal mathematical reasoning, Phys. Rev. ST PER **9**, 010105 (2013).

53. R. Seigler and Z. Chen, Developmental differences in rule learning: A microgenetic analysis, Cog. Psych **36**, 273 (1998).

54. T. Kuhn, *The Structure of Scientific Revolution* (University of Chicago Press, Chicago, IL, 1996).

5.8 APPENDIX D

Figure 5-10 shows the distribution of TAs' assigned scores for SSD vs. SSE in the quiz context both before (left) and after (right) the rubric intervention. TAs who are below the diagonal line in the graphs in Fig. 5-10 score SSD higher than SSE. On average, students graded solution SSD slightly higher than solution SSE before the rubric intervention and SSE slightly higher after the intervention. Some individual TAs graded solution SSE much lower than solution SSD in both contexts (for example, one TA gave SSE a homework score of zero and another TA gave SSE a score of one). If these outlier scores are removed, the remaining distributions of scores are more consistent with the scores observed in prior semesters of the course [22,23]. Eight TAs (out of 15) gave SSE a score greater than or equal to SSD in both the homework and quiz contexts.

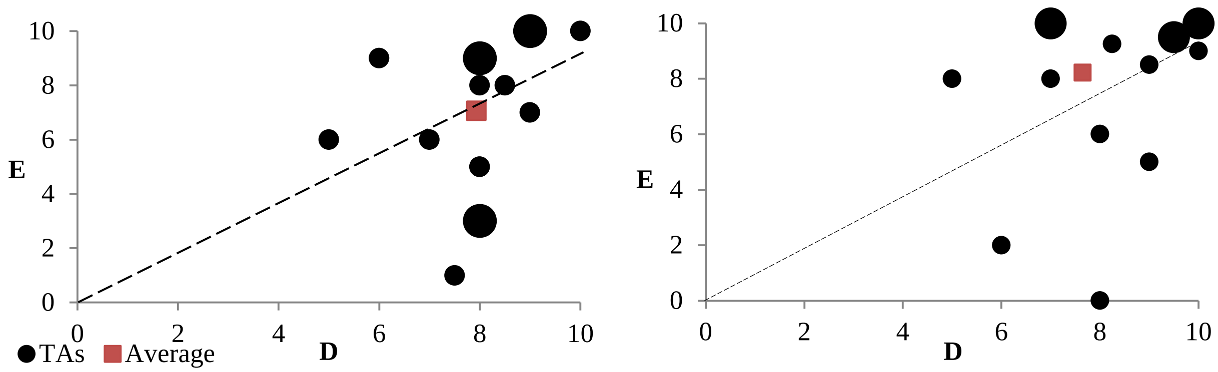


Figure 5-10 Distribution of 15 TAs' scores assigned to SSE vs. SSD in the quiz context (Left) before the rubric intervention and (Right) after the rubric intervention. The larger bubbles represent two TAs at that particular point.

6.0 CONTRASTING CRITERIA USED TO GRADE INTRODUCTORY PHYSICS PROBLEMS AND QUANTUM MECHANICS PROBLEMS: A CASE STUDY OF PHYSICS GRADUATE TEACHING ASSISTANTS

6.1 INTRODUCTION

At large research institutions in the U.S., graduate students in physics often play an important role in the education of undergraduate students in physics courses at all levels. It is quite common for physics graduate Teaching Assistants (TAs) to teach introductory physics recitations or labs and grade student work in introductory and advanced courses. Common goals for physics courses at all levels [1] are to help students learn physics [2-7], develop students' problem-solving and reasoning skills and help them make better use of problem solving as an opportunity for learning [8-10]. It is important that TAs' teaching practices promote these learning goals.

At the Graduate Education in Physics Conference jointly sponsored by the American Physical Society and the American Association of Physics Teachers, discussions with faculty about teaching assistantships suggest that the majority of physics departments at research institutions in the U.S. employ physics graduate students as TAs for introductory physics courses and for grading in courses at all levels [11]. The TAs are generally expected to do most of the grading, including grading exams in introductory courses and homework and quizzes in both introductory and upper-level courses. Many of the physics departments provide very brief training to the TAs (half day or less) to help them learn how to carry out their teaching responsibilities [11]. However, a handful of departments have provided a semester-long TA professional development program similar to the one associated with the present study. Other than the training provided by the department, most conference participants noted that the TAs usually carry out the tasks without significant guidance from their supervising instructor except for a general discussion about how to carry out recitations or how to grade [11].

TAs are often responsible for grading students' work in undergraduate physics courses at all levels. TAs' grading approaches can help shape student learning and communicate instructors' goals and expectations to the students [16-20]. Physics education research suggests that placing the burden of proof for explicating the problem solving process on students in both introductory and advanced courses can help students develop problem solving skills and learn physics. Most TAs receive very little training or guidance about grading, and they may not have had the opportunity to reflect on their goals for grading or develop grading practices that promote learning [12,13]. TAs' grading beliefs and practices are often based upon their own experiences as students, the expectations of their supervising instructor, and their workload [14]. Moreover, TAs may perceive the difficulty of a problem they are grading from their own perspective instead of the perspective of their students [15]. These factors can impact TAs' beliefs about grading and shape their grading practices in different courses, and their grading beliefs and practices may change depending on the course level. For example, TAs have significantly more expertise in solving introductory physics problems, and they may not think about the difficulty of an introductory physics problem from their students' perspective. They may assume that the answers to introductory physics problems are obvious and students do not need to show their work while solving them [15]. As a result, when grading introductory physics solutions, TAs may not require that students explicate the problem solving process. On the other hand, since TAs may not yet be experts in an advanced course such as quantum mechanics (QM), they may perceive a QM problem to be difficult. As a result, when grading QM student solutions, it is possible that the TAs expect students to explicate the problem solving process.

Since grading plays a crucial role in student learning, TAs can benefit from an opportunity to reflect upon their grading goals and practices. Contemplating and reflecting on the reasons for the differences in their grading practices in courses at different levels can help clarify their beliefs and improve their grading practices. This research study investigates whether physics graduate TAs are aware of solution features that are conducive to learning when preparing introductory physics and quantum mechanics problem solutions for their students. We also investigated whether physics graduate TAs grade student solutions in introductory physics and quantum mechanics using different criteria, and if so, what are the reasons for the differences. By asking TAs to grade student solutions in both introductory physics and quantum mechanics and compare their grading in the two contexts, TAs were given an opportunity to reflect on their grading goals and practices and resolve possible conflicts between their goals and practices. The findings of the study can inform professional development leaders interested in helping TAs improve their grading practices.

This case study involved 15 graduate TAs participating in a semester-long professional development program at a research university in the U.S. The data collection tool was designed to probe implicit and potentially conflicting perceptions regarding the goals of grading and grading practices. TAs were given an introductory physics problem and a QM problem and were asked to create solutions to the problems that would help their students learn. Then, TAs were given a set of introductory student solutions that were used in prior studies to investigate faculty grading practices [1] as well as a set of QM student solutions that have solution features similar to the introductory student solutions (e.g., some solutions explicate the problem solving process while other solutions briefly provide the correct answer but do not explicate the problem solving process). All the steps in the shorter solutions to each problem were included in the longer solution that explicated the problem solving process (but the longer solution had additional steps). The contrasting solution features in the short and long solutions to the same problem were designed to encourage graders to reflect on various problem solving approaches that educational literature suggests promote desired problem-solving practices [2-8, 21-23]. The TAs were asked to grade the student solutions for the introductory physics and QM problems and explain whether they used different criteria when grading student solutions in the two different contexts. They were also asked to explain why they used different criteria in the two contexts (if they used different criteria).

In particular, the study was designed to investigate the following research questions:

1. What features do TAs include in their own problem solutions when creating solutions for students in introductory physics and QM?
2. Do TAs grade students' solutions to an upper-level QM problem differently than students' solutions to an introductory physics problem?
3. What solution features do TAs grade on in upper-level QM versus introductory physics?
4. What are the TAs' stated reasons for whether (or not) their grading is different for an introductory problem versus a QM problem?

We begin with a literature review before discussing the methodology. Then, we present the findings and follow up with a discussion and summary.

6.2 BACKGROUND

One of the common goals for physics courses at all levels is to help students develop expertise, i.e., gain a robust understanding of physics and develop effective problem-solving skills. Many prior studies [1,2,12,13,16,17,24,25] have documented differences between experts and novices in a particular domain when approaching problems. Both use heuristics to guide their search process in identifying the gap between the problem goal and the state of the solution and taking action to bridge this gap. However, novices differ from experts in the types of heuristics they use to solve problems. Novices approach problems in a haphazard manner, typically searching for appropriate equations first and plugging in numbers until they get a numerical answer [24]. Furthermore, novices often draw on their naive knowledge base rather than formal physics knowledge [18]. Novices also engage in pattern matching, i.e., attempting to solve a problem using another previously solved problem with similar surface features, even if the underlying concepts and principles are different [18]. On the other hand, experts devote time and effort to qualitatively describe the problem situation, identify principles and concepts that may be useful in the analysis of the problem, and retrieve effective representations based on their better organized domain knowledge [1,2,12,13,17,24-27]. In addition, experts devote time to plan a strategy for constructing a problem solution by devising a useful set of intermediate goals and means to achieve them, frequently by working in a backward manner [1-2,12,16]. Experts also spend more time than novices in using diverse representations to analyze and explore problems (especially when they are not sure how to proceed) [16]. Experts also engage more than novices in self-monitoring by evaluating previous steps and revising their choices as needed [12,16,17,19]. They utilize problem solving as a learning opportunity more effectively by engaging in self-repair—identifying and attempting to resolve conflicts between their own mental model and the scientific model conveyed by peers' solutions or worked-out examples [10].

One way to help students develop expertise in physics is to encourage them to use effective approaches to problem solving (e.g., starting with a conceptual analysis of the problem, planning and making decisions before implementing the plan, and then checking the reasonability of the solution obtained and reflecting upon the problem-solving process to learn) instead of a plug and chug approach (e.g., starting by looking for a formula that matches the quantities in the problem statement). Prior research suggests that students in a traditionally taught physics course who were required to use effective problem solving approaches performed better than students who were allowed to use any problem solving approach they preferred as the complexity of the problems increased [4]. The students in these two groups were matched in terms of their prior performance in the physics class (for example, students who had a C grade were matched with other students with a similar grade). This study was conducted in a one-on-one situation outside of the class and many students were surprised at how they were able to solve complex problems when they were required to use a systematic approach [4]. The researchers of this study noted that the students often had a tendency to start looking at the formula sheet before doing a qualitative analysis of the problem and planning the problem solution [4].

Since students often value what they are graded on, grading their solutions on the explication of the problem-solving process can put the burden of proof of understanding students' thought processes while solving a problem on the students and can encourage students to use a systematic approach to problem solving. In the spirit of formative assessment [28,29], grading can provide feedback that can improve student learning and communicate to learners what practices are useful in learning the discipline and for developing problem solving skills [14]. Effective grading practices can also communicate to students what to focus on in future learning activities [29-33]. Such practices can encourage students to explain their reasoning (i.e., requiring that the students explain the reasoning underlying their solutions) and provide them with an artifact to reflect on and learn from after problem solving (i.e., from their own graded clearly articulated solution in which the problem-solving process is explicated) [11]. Thus, grading in physics courses at all levels should reward the use of effective problem-solving strategies such as drawing a diagram, listing known and unknown quantities, clarifying considerations in setting up sub-problems, and evaluating the reasonability of the problem solution.

However, TAs may not have had the opportunity to think about the goals for the physics course in which they are TAs or develop teaching practices that support those learning goals. Prior research has investigated common beliefs and practices among physics TAs that have implications for effective teaching [12,24,25,34-39]. For example, research suggests that sometimes graduate TAs struggle to understand the value of thinking about the difficulty of a problem from an introductory students' perspective [36,37]. Also, while graduate TAs state that they have the goal of helping students develop effective problem solving approaches, they do not notice features in example solutions that are supportive of helping students develop effective problem solving approaches [24,25,38]. Furthermore, the TAs do not always engage in grading practices which are conducive to helping introductory physics students learn desired problem-solving approaches and develop a coherent understanding of physics [12,13]. On the other hand, in advanced courses, it is possible that TAs are more easily able to recognize the difficulty of a problem and identify effective problem solving approaches.

Here, we discuss an investigation focusing on possible differences in TAs' beliefs about grading and grading practices in introductory physics and QM to uncover possible discrepancies in the two contexts. We find that there are differences in TAs' beliefs about grading and grading practices. The findings of the study can be useful for professional development of TAs and can be used to help TAs reflect on and resolve conflicts in their grading goals and practices in introductory physics and QM so that their grading practices in both cases are aligned with improving student learning.

6.3 METHODOLOGY AND DATA COLLECTION

6.3.1 Description of TA Professional Development Course

In this investigation, we collected grading data from a mandatory, semester-long TA professional development course led by one of the authors. The course met for two hours each week for the entire semester and was meant to prepare the TAs for their teaching responsibilities. The TAs in general were asked to do one hour of homework each week pertaining to the professional development course, e.g., related to grading, that was graded for completeness. During class meetings, TAs generally discussed their homework assignment from the previous week in small groups. At the end of the class, they shared what they had discussed in groups while the instructor gave input. The TAs had also attended a one-day new teaching assistant workshop facilitated by the university, but this workshop was general and did not focus on discipline-specific issues in teaching and learning physics. There were 15 first-year graduate students enrolled in the course. The majority of the first-year graduate students were TAs. Most of the

TAs were teaching recitations for introductory physics courses for the first time. A few other TAs were also assigned to facilitate a laboratory section or grade students' work in various physics courses for the first time. In the same semester, a majority of the TAs were also tutors in a physics resource room where introductory students can receive help on assignments such as homework and laboratory reports. The participants consisted of a mix of domestic and international students from nations such as China, India, Turkey, etc. There were 4 female TAs and 11 male TAs. The demographics of the TAs in this course are somewhat similar to national norms [40].

6.3.2 Data Collection

6.3.2.1 Development and validation of the data collection tool

The data on TAs' beliefs about grading and grading practices were collected using a group administered interactive questionnaire (GAIQ) previously developed and validated by three of the authors in collaboration with two graduate student researchers in physics education for use with TAs/instructors [24]. This tool consists of a series of activities involving worksheets which are designed to clarify a TA/instructor's ideas about helping students learn physics content and effective problem solving approaches. The GAIQ worksheets and artifacts encourage reflection on various facets of teaching physics problem solving: designing problems on a particular physics topic with features effective for use in different situations (e.g., questions for clicker and class discussion, homework, quiz, exams, collaborative learning, etc.), designing solutions to problems suitable for distributing to their students that will help students learn, and grading student solutions. Questionnaires on each facet of teaching problem solving (e.g., problem types, instructors' example solutions, or grading) involve three stages: 1. TAs/Instructors are individually asked to solve an introductory physics problem (Fig 6-1) and complete a worksheet eliciting TAs/Instructors' initial ideas about teaching problem solving; 2. TAs/instructors work in groups of three to discuss their ideas from the pre-class activity and then a whole class discussion takes place in which groups share their ideas; 3. TAs/instructors individually complete another worksheet in which they can modify their previous answers and connect their ideas to a list of pre-defined features about teaching problem solving developed by the researchers.

> You are whirling a stone tied to the end of a string around in a vertical circle having a radius of 0.65 m. You wish to whirl the stone fast enough so that when it is released at the point where the stone is moving directly upward it will rise to a maximum height of 23 meters above the lowest point in the circle. In order to do this, what force will you have to exert on the string when the stone passes through its lowest point one-quarter turn before release? Assume that by the time you have gotten the stone going and it makes its final turn around the circle, you are holding the end of the string at a fixed position. Assume also that air resistance can be neglected. The stone weighs 18 N.
>
> The correct answer is 1292 N.

Figure 6-1 Core Problem 1.

The initial versions of the worksheets used in the GAIQ were developed using the findings of semi-structured interviews with faculty members based on an "artifact comparison" approach [24]. In these interviews, faculty members were asked to make judgments about instructional artifacts which were similar to those they often use in their classes. The specific types of artifacts that were presented to instructors during interviews were designed to reflect those that would be familiar to physics instructors. In particular, the three types of instructional artifacts were instructors' example solutions, student solutions, and problem types (e.g., problems in multiple-choice format, context-rich form, divided into sub-problems, with and without diagrams, etc.). The artifacts presented to the instructors during interviews were designed to create a context which would activate beliefs that could influence decisions when they select instructional material or pedagogical techniques while teaching [24]. All of the original GAIQ activities about instructors' solutions, student solutions, and problem types refer to an introductory physics problem shown in Fig. 6-1 (which was used in this study) [24]. The introductory physics problem was designed, validated and approved by four physics instructors who taught introductory physics courses at the University of Minnesota and was used on final exams. The problem was also sent to several other instructors of physics courses and all of them reported that the problem was difficult enough to require an average student to use an exploratory decision making process as opposed to an algorithmic procedure [24]. The problem involves synthesis of several important physics concepts and principles. The problem included several features of a context-rich problem [24] (i.e., it was set in a realistic context, was not broken into parts, and did not include a diagram, etc.) and is rich enough to allow for interesting variations in students' solutions. Students could potentially solve the problem in different ways. Thus, the problem allows for a spectrum of more or less desired problem solving practices. The student final exam solutions were available, providing a source of authentic student solutions which were used both in Ref. [24] and in the present study. The specific artifacts involving grading activities included five student solutions (see an example of two student solutions in Fig. 6-2), which were based upon actual students' common responses to the final exam. The artifacts were chosen to reflect differences between expert and novice problem solving from the research literature such as including a diagram describing the problem, explication of sub-problems, justification of solution steps, evaluation of the final answer, explication of the scientific principles used, evidence of reflective practices, etc. [24].

Instructors' responses to interview questions about the instructional artifacts revolving around the introductory physics problem and five student solutions were used to create the initial GAIQ worksheets, including the worksheets on grading [1]. The GAIQ is meant to take the place of individual TA/instructor interviews about the teaching and learning of problem solving. While the development and validation of the GAIQ was a very time consuming process [24], the GAIQ requires significantly less time than interviews for data collection and analysis. Equally important, it avoids researcher intervention in the process of clarifying the interviewees' responses, and the inter-rater agreement on the coding of the data obtained and interpretation of the data is excellent. Thus, the GAIQ worksheets can be used by researchers and professional developers at different institutions to collect and analyze data and data across different institutions can readily be compared with relative objectivity.

The initial version of the GAIQ was iterated between the researchers and physics instructors and modified to a version which was administered in the context of professional development for Israeli pre-service and in-service teachers many times [24]. After each initial implementation and feedback from the teachers, the GAIQ was refined further until a version satisfactory to the researchers was developed. The GAIQ tool was then adapted for a professional development program for physics teaching assistants in the U.S. The TA professional development program in this study anchored the professional development activities in collaborative reflection with peers (other TAs) on classroom experiences [12,13,25]. Reflection on practice serves to enrich instructors' interpretations of classroom experiences, widen the inventory of possible actions instructors might use, clarify instructional goals, examine practice in view of these goals, and provide motivation for the adoption of new instructional strategies. Following these suggestions, the activities in the TA professional development program elicited TAs' initial ideas on different facets of teaching and learning. Then, the instructor facilitated peer discussions about their ideas on those facets of teaching and learning, enabled entire class discussions in which the instructor provided ideas for "best practices", and also provided opportunities for TAs to reflect on their ideas (for example, opportunities to think about discrepancies in their ideas about teaching and learning and reflect on changes in their initial ideas).

The GAIQ including the grading activities were implemented in three different semesters of a TA professional development program in the U.S., and after each implementation, the researchers iterated the version several times between them. A graduate student researcher in PER observed the three semesters of the TA professional development program when TAs worked on the GAIQ. The graduate student researcher and two of the authors revised and iterated the GAIQ based upon the TAs' comments and responses. This validation process in the context of the TA professional development program ensured that TAs interpreted all components of the GAIQ appropriately as the researchers had intended.

The artifacts about grading introductory students' solutions have also been used as the basis of a previous investigations on faculty members' grading practices [1]. In that previous study [1], faculty members were asked to solve the core problem (see Fig. 6-1) and compare and make judgments about two student solutions (see Fig. 6-2) to the core problem. These two solutions were chosen because they trigger conflicting instructional considerations in assigning a grade [1]. In the study presented here, since one of the problems for which graduate TAs were asked to grade introductory student solutions was this problem, we suggest that the readers examine the student solutions (see Fig. 6-2) and think about how to grade them. Clearly incorrect aspects of the solutions are indicated by boxed notes. Both solutions end up with the correct answer. The solution SSD includes a diagram, articulation of the principles used to find intermediate variables, and clear justification for the final solution. The elaborated reasoning in SSD reveals two canceling mistakes, involving misreading of the problem situation as well as misuse of energy conservation to imply circular motion with constant speed. On the other hand, the solution SSE is brief with no explication of reasoning, and it does not give away any evidence for mistaken ideas. However, lines of work very similar to the three lines of work in SSE are also present in SSD, suggesting that Student E could have been guided by a similar thought process as Student D. In this investigation, we will focus on comparing the TA grading of introductory solutions shown in Fig. 6-1 with the same TA's grading of student solutions to a QM problem.

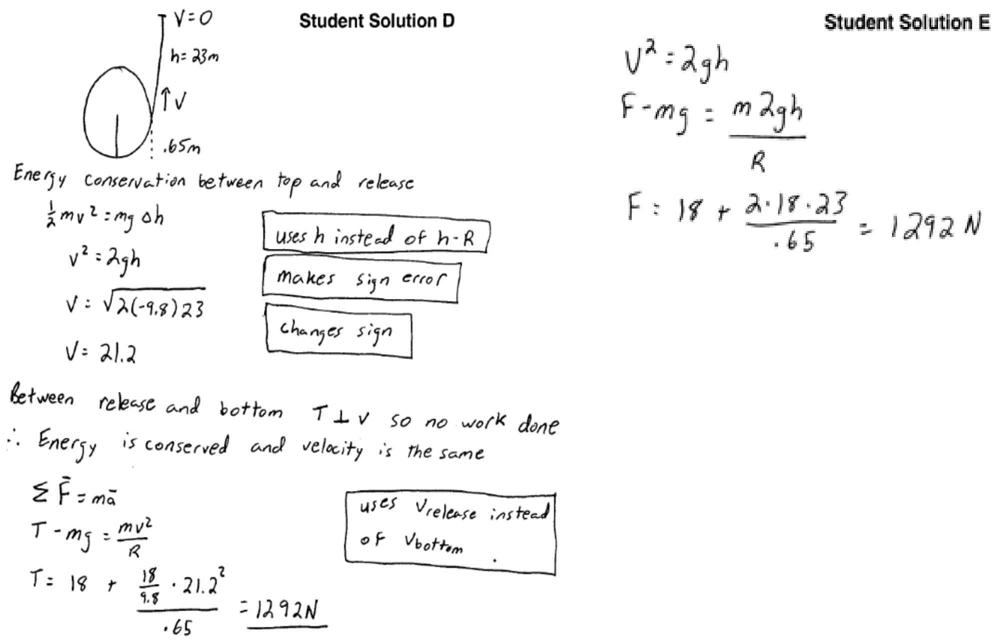


Figure 6-2 For the introductory physics problem, Student Solution D (SSD) and Student Solution E (SSE).

To investigate TAs' grading practices when grading student solutions to QM problems and compare them with their introductory physics grading, we incorporated a QM problem (see Fig. 6-3), two student solutions to this problem (see Fig. 6-4), and a grading worksheet to the GAIQ grading activities for QM in the present study. The QM problem was developed and iterated over a period of more than ten years and had been used on midterms and exams in several advanced QM courses at a large research university. The QM problem is difficult enough to require an average student in a quantum mechanics course to use an exploratory decision-making process as opposed to an algorithmic procedure. An initial qualitative analysis of the problem and planning can greatly facilitate the problem solving process. Two student solutions (Student Solution 1 (SS1) and Student Solution 2 (SS2)) to this problem were developed and iterated several times between three researchers based on actual student responses to the QM problem from previous years with common conceptual difficulties (see Fig. 6-4).

> For an electron in a one-dimensional infinite square well with well boundaries at $x = 0$ and $x = a$, measurement of position yields the value $x = a/2$. Write down the wave function immediately after the position measurement and without normalizing it show that if energy is measured immediately after the position measurement, it is equally probable to find the electron in any odd-numbered energy stationary state.

Figure 6-3 The upper-level quantum mechanics problem.

Student Solution 1

$$P = \left| \int \psi_n^*(x)\, \Psi(x)\, dx \right|^2$$

$$= \left| \sin \frac{n\pi}{a} \frac{a}{2} \right|^2$$

$$= \left| \sin \frac{n\pi}{2} \right|^2$$

which is 0 for n even and equal for n odd.

Student Solution 2

$\psi_n(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right)$ energy eigenstates for infinite square well

$\Psi(x) = \frac{a}{2}$ wave function after measuremen

Probability for measuring $E_n = |\langle \psi_n | \Psi \rangle|^2$

$$\text{Prob} = \left| \int \psi_n^*(x)\, \Psi(x)\, dx \right|^2$$

$$= \frac{2}{a} \left| \int \sin \frac{n\pi x}{a}\, \Psi(x)\, dx \right|^2$$

integral goes away because of delta function, so

$$\text{Prob} = \frac{2}{a} \left| \sin\left(\frac{n\pi}{\cancel{a}} \cdot \frac{\cancel{a}}{2} \right) \right|^2$$

$$= \frac{2}{a} \left| \sin \frac{n\pi}{2} \right|^2$$

for even n, Prob = 0

for all odd n, Prob $= \frac{2}{a}$

Figure 6-4 Student Solution 1 (SS1) and Student Solution 2 (SS2) to the quantum mechanics problem.

To make comparisons in the grading approaches of the TAs for the introductory physics solutions and QM solutions, the QM solutions developed were made analogous to the two introductory physics solutions, i.e., SS1 for QM is similar to SSE for introductory physics and SS2 for QM is similar to SSD for introductory physics. Both SS1 and SS2 include the correct answer. SS2, like SSD, includes articulation of the principles used to find intermediate variables, and clear justification for the final result. Similar to SSD, the elaborated reasoning in SS2 reveals a mistake involving writing the wave function immediately after measurement as $\psi(x) = \frac{a}{2}$ rather than $\psi(x) = A\delta\left(x - \frac{a}{2}\right)$, though the delta function is mentioned later in the solution. Like the brief introductory solution SSE, the quantum solution SS1 is brief with no explication of reasoning, and it does not give any evidence for mistaken ideas on the part of the student. However, the three lines of work in SS1 are also present in SS2, suggesting that Student 1 might be guided by a similar thought process as Student 2.

6.3.2.2 Implementation of the data collection tool

The outset of the course revolved around the group administered interactive questionnaire (GAIQ) encouraging reflection on grading. Table 6-I shows the sequence of grading activities. The activities served as a data collection tool in order to study TAs' grading decisions and considerations in a simulated environment as well as a learning experience within the professional development program [24].

The GAIQ included several stages (see Table 6-I). At the beginning of the semester, TAs were asked to create a solution to the introductory physics problem (see Fig. 6-1) and the QM problem (see Fig. 6-3) that they would give to their students to help them learn. The TAs were also asked to individually grade introductory physics solutions SSE and SSD for both homework (HW) and quiz contexts out of a total score of ten points, list characteristic solution features, and explain their choice of weights for the different features to obtain a final score (see Figure 6-5). The TAs were told to assume that 1) they were the instructors of the class and could structure their grading approaches to improve learning; 2) they had the authority to make grading decisions; and 3) they had told their students how they would be graded. An example response (transcribed) is shown in Fig. 6-5.

During the in-class stage of the GAIQ for introductory physics (see Table 6-I), the TAs worked in groups of 3-4 in which they were asked to discuss and try to reach an agreement regarding grading the student solutions SSD and SSE. After they had graded the solutions, a representative from each group shared their grading approaches with the entire class. Two of the authors were present in the class. One researcher coordinated the class work and the discussion at the end of the class which highlighted "best practices" of grading, i.e., grading approaches that promote desired problem solving. The instructor of the professional development program also discussed with TAs the disadvantages of grading which focused exclusively on correctness. One researcher observed and documented the TAs' comments during the class discussions.

Table 6-I Sequence of TA grading activities.

Time		Activity
Beginning of Semester	Homework	• Individually, TAs were asked to create a solution to the introductory physics problem (see Fig. 6-1) and the QM problem (see Fig. 6-3) that they would give to their students to help them learn. • Individually, TAs completed a worksheet which asked them to grade student solutions (see Fig. 6-2) to the introductory problem (see Fig. 6-1) in homework (HW) and quiz contexts, list features of each solution, and explain their choice of weights for the features to arrive at a final score.
	In Class	• In groups of 3-4, TAs graded the student solutions SSD and SSE using a group worksheet and then participated in a whole-class discussion in which the groups shared their grading approaches.
Immediately After the Introductory Student Grading Activities	Homework	• TAs were given a solution to the QM problem shown in Fig. 6-3. • Individually, TAs completed a worksheet which asked them to grade SS1 and SS2 (see Fig. 6-4) corresponding to the quantum mechanics problem (see Fig. 6-3) in HW and quiz contexts, list features of each solution, explain their choice of weights for the features to arrive at a final score, and identify differences in their grading practices compared to when grading the introductory problem solutions.

Right after the introductory grading activities were over, the TAs were given the solution to the QM problem shown in Fig. 6-3 and were asked to grade two student solutions to the QM problem: Student Solution 1 (SS1) and Student Solution 2 (SS2) (see Fig. 6-4), for both the homework (HW) and quiz contexts out of a total score of ten points. See Fig. 6-6 for an example response. TAs were also asked to list characteristic solution features of SS1 and SS2, and explain their choice of weights for the different features to obtain a final score. TAs were also asked the following questions regarding their grading practices:

1. Was your grading approach different when grading introductory physics student solutions vs. upper-level quantum mechanics student solutions? If so, why? If not, why not?
2. How did your grading considerations change when grading introductory physics student solutions vs. upper-level quantum mechanics student solutions? What are the reasons for these differences?

Features: Solution E	Score		Reasons: explain your reasoning for weighing the different features to obtain your assigned score.
	HW	Q	
The answer is correct, the approach is correct. The steps for getting $v^2 = 2gh$ are not written	8	10	I gave this student a lower grade on HW because I think that students have enough time to write down all steps, and they should. This answer looks like it has been written just to get a grade, not that the student was learning something while doing the HW. I think that since the approach and the answer are right, this answer gets a full grade on a quiz.

Figure 6-5 One component of a sample TA's worksheet (transcribed) related to introductory student solution SSE which was part of the pre-grading activity.

Features: Solution 2	Score		Reasons: explain your reasoning for weighing the different features to obtain your assigned score.
	HW	Q	
-Organizing/setting up -One mistake $\Psi(x) \neq \frac{a}{2}$ -Missing $\lvert A \rvert$ -Explaining himself adequately -Knowns and unknowns	9	9.5	This student understands the problem and using the correct approach. However, his statement $\Psi(x) = \frac{a}{2}$ is not correct and he omitted $\lvert A \rvert$. For homework, he will lose one point, but for quizzes ½ point is enough.

Figure 6-6 One component of a sample TA's worksheet (transcribed) related to advanced QM student solution SS2 which was part of the pre-grading activity.

6.3.2.3 Post-course interviews

After an initial analysis of the collected data, in the following semester, seven of the TAs in the study volunteered to be interviewed to provide further clarification of their grading beliefs and practices and to investigate whether the grading activities carried out in the TA training class impacted their beliefs about their grading in some manner not captured in their written responses. The interviewer had some pre-determined questions to ask the TAs (e.g.: "Can you elaborate on the differences in grading solutions to introductory physics problems compared to grading solutions to QM problems? Did your approach to grading students' solutions to introductory physics problems change after the grading activity involving QM solutions? Have your experiences as a TA in introductory physics caused you to reflect on your grading approach? Have your beliefs about grading changed due to the interventions in the TA professional development course? What caused the change in beliefs?"). However, the interviewer also asked additional follow-up questions on-the-spot to examine TAs' reasoning and also to give them an opportunity to clarify their written responses on the GAIQ worksheets if there were any ambiguities in their responses.

6.4 RESULTS

6.4.1 What Features Do TAs Include in Their Own Solutions to the Introductory Physics and QM Problems?

To investigate Research Question 1 related to the features the TAs included in their own solutions they would give out to their students, we examined TAs' own solutions to the introductory physics problem and QM problem. For the QM problem, many TAs stated in class that the assignment to create a solution to it to help their students learn was difficult for them and they struggled to solve it. However, we find that although many of the TAs struggled to solve the QM problem, all but one of the TAs' written solutions to the QM problem demonstrated effective problem solving strategies and explication of problem solving. The majority of the TAs included an explanation and justification of their reasoning while solving the problem. In addition, most of the TAs broke the problem down into intermediate steps in order to solve it. We also examined the TAs' solutions to the introductory physics problem and found that all of the TAs' solutions demonstrated effective problem-solving strategies. All but one of the TAs' solutions to the introductory physics problem contained explication and justification of the steps. The majority of TAs included a diagram, broke the problem into sub-problems, and listed knowns and unknowns.

Thus, the problem solutions created by the TAs for their students show that, in both introductory physics and QM contexts, they recognized the value of explicating the problem solving process in their solutions for their students. The features included in their solutions suggest that the TAs knew how to solve problems using an effective problem solving approach. However, in the following section, we discuss findings that suggest that although TAs created solutions that included effective problem-solving approaches in both the QM and introductory physics contexts, they often did not penalize solutions in which these features were missing in the introductory physics context. In contrast, in the QM context, TAs were more likely to grade on explication of problem solving and explicit demonstration of conceptual understanding.

6.4.2 Scores Assigned by TAs on Introductory and QM Student Solutions

To investigate Research Question 2 (Do TAs grade students' solutions to an advanced QM problem differently than students' solutions to an introductory physics problem?), TAs' assigned scores on the QM solution and the introductory physics solution were analyzed. Table 6-II shows the average scores and standard deviations when TAs graded the introductory solutions and the QM solutions in both the homework and quiz contexts. TAs tended to grade elaborated solutions higher and brief solutions lower in both the introductory and QM contexts, but the difference was more pronounced for the QM solutions than for the introductory physics solutions. The highest disagreement among TAs was about what scores to assign the brief solution to the introductory problem SSE (Std. dev. $= 3.16$ for the HW context and 2.71 for the quiz context). We performed t-tests for comparison, and found that the differences in averages were statistically significant between the QM solutions SS1 and SS2 in both the HW context ($p < 0.001$) and quiz context ($p = 0.008$) but not statistically significant for the introductory solutions (see Table 6-II).

Table 6-II Average scores assigned to the brief and elaborated solutions to the introductory and QM physics problems in the homework (HW) and quiz (Q) contexts, with corresponding standard deviations (Std. Dev.) for each score and *p*-values for comparison between brief and elaborated solution scores as well as between introductory physics and QM solution scores.

		Introductory Physics Solutions			**QM Solutions**		
		Brief (SSE)	Elaborated (SSD)	p	Brief (SS1)	Elaborated (SS2)	p
HW	Average	6.00	7.40	0.130	4.93	7.67	< 0.001
	Std. Dev.	3.16	1.30		1.87	1.63	
Q	Average	7.07	7.93	0.274	6.57	8.47	0.008
	Std. Dev.	2.71	1.24		2.06	1.55	

Figure 6-7 shows the distribution of TAs' assigned scores for the elaborated solutions to the QM problem versus the introductory problem in the quiz (left) and homework (right) contexts. The smallest bubbles represent one TA, and a larger bubble shows that many TAs are clustered at that point (the number of TAs at a particular point is proportional to the relative size of the bubble). TAs who are above the diagonal line in the graphs score the QM solution higher than introductory physics solution. While the scores were mostly grouped near the upper right corner, the scores were somewhat higher for the QM problem than for the introductory problem, though a t-test shows that the difference between the means was not statistically significant for either the quiz ($p = 0.307$) or the homework ($p = 0.625$) contexts.

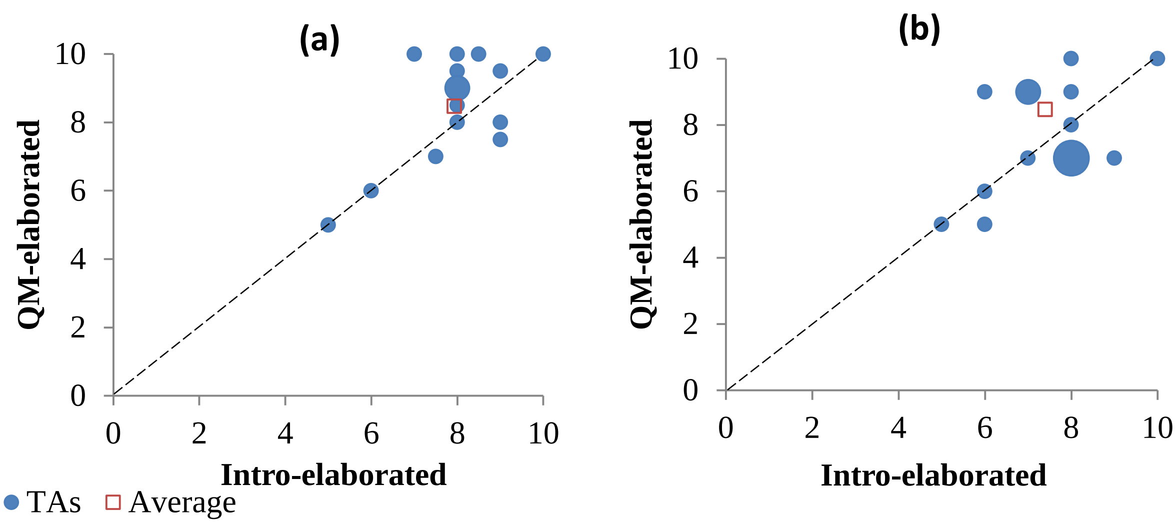


Figure 6-7 (a) Distribution of individual scores assigned to the elaborated solutions SS2 (QM) versus SSD (Intro) in the quiz context ($p = 0.307$). (b) Individual scores assigned to the elaborated solutions SS2 (QM) versus SSD (Intro) in the homework context ($p = 0.625$). The relative size of the bubble represents the number of TAs at a particular point ($N = 15$).

Figure 6-8 shows the distribution of TAs' assigned scores for the brief solutions to the QM problem versus the introductory physics problem in the quiz (left) and homework (right) contexts. TAs tended to grade the brief QM solution somewhat lower than the brief introductory solution (fewer TAs are above the diagonal line), though the difference in the means was not statistically significant for either the quiz ($p = 0.574$) or the HW ($p = 0.273$) contexts.

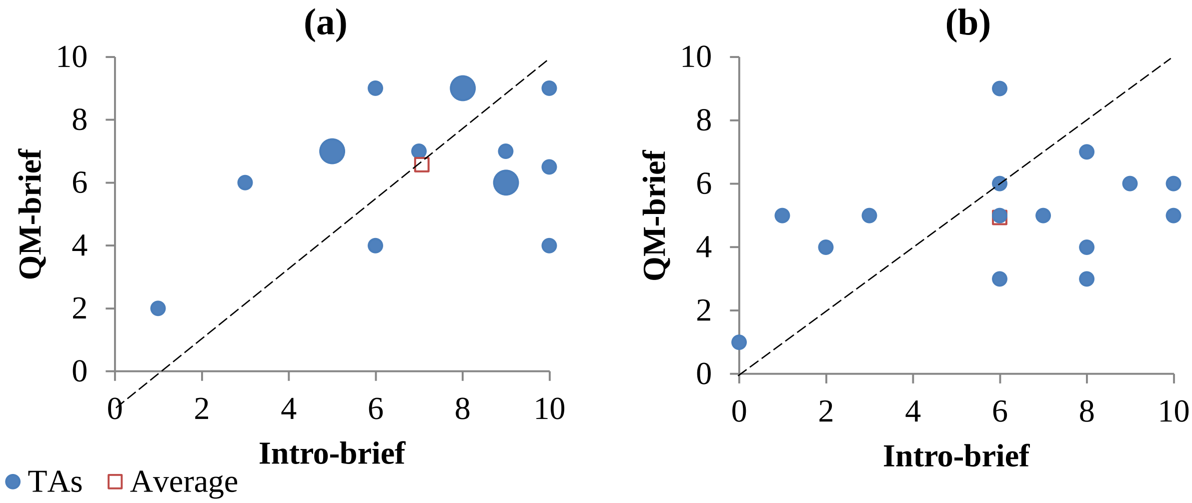


Figure 6-8 (a) Distribution of individual scores assigned to the brief solutions SS1 (QM) versus SSE (Intro) in the quiz context ($p = 0.574$). (b) Individual scores assigned to SS1 (QM) versus SSE (Intro) in the homework context ($p = 0.273$). The size of the bubble represents the number of TAs ($N = 15$).

6.4.3 Grading Criteria

In order to investigate Research Question 3 (What solution features do TAs grade on in advanced QM vs. introductory physics?), the solution features TAs graded on in the introductory physics solution and the QM solution were analyzed. In the GAIQ worksheets, TAs were asked to grade introductory physics solutions and the QM solutions in a HW and a quiz context, list features of each solution, and explain their choice of weights for the features to arrive at a final score. Data analysis involved coding the features listed by TAs in the worksheets into a combination of theory-driven and emergent categories. Twenty-one solution features were identified. The coding was done by two of the researchers. In cases where disagreement occurred, this was usually due to vagueness in the wording of TAs' written statements. After comparing codes, the researchers discussed any disagreements during multiple meetings until agreement better than 90% was reached.

To facilitate interpretation of the data, the features were analyzed by grouping them into 5 clusters, as shown in Table 6-III. Each solution feature listed by a TA was entered into only one cluster. Cluster 1 (C1) includes features related to desired problem solving practices [2-8] (i.e., initial problem analysis as well as evaluation of the final result). Cluster C2 also involves features related to desired problem solving practices such as explication of reasoning (i.e., articulation and justification of principles). Cluster 3 (C3) includes domain-specific features, such as invoking relevant physics principles and applying them properly. Cluster 4 (C4) includes features related to elaboration which emerged during the coding process, e.g., "written statements," "good presentation," "solution in steps," and "conciseness." These features were not assigned to the "explication" category C2 because they were imprecise. Cluster C2 is focused on the explication and justification of the physics principles, whereas C4 is more about general communication of the solution. For example, we could not differentiate whether a TA who wrote "written statements" meant that the student solution includes an explicit statement of a principle in writing, explicit justification of a principle in writing, or simply a written statement. Thus, we coded "written statements" as belonging in the general category C4. Similarly, if a TA noted that a solution is "organized" he/she could mean that the solution is neatly written or that it is systematic. Cluster C4 also involves solution features related to lack of elaboration, e.g., conciseness (this feature was mentioned most often by TAs when they graded the brief student solution SSE). Finally, Cluster 5 (C5) focuses on correctness of algebra and the final answer.

Table 6-III Sample features sorted into clusters and sample citations.

<table>
<tr><td>C1
Problem description
& evaluation</td><td colspan="2">Visual representation (e.g., “diagram,” “figure,” “graph”); articulating the target variables and known quantities (e.g., “knowns/unknowns,” “list of variables,” “nothing labeled”); evaluation of the reasonability of the final answer (e.g., “check,” “double check what they did”)</td></tr>
<tr><td>C2
Explication of
problem-
solving approach</td><td colspan="2">Articulation of principles (e.g., “labels energy conservation use,” “text showing knowledge of concepts”); justifying principles (e.g., “explained the reason he used the formulas,” “explanation for constant velocity,” “no demonstration for why the first equation holds”)</td></tr>
<tr><td>C3
Domain knowledge</td><td colspan="2">Essential principle invoked (e.g., “sums forces, energy conservation,” “has not written [the stationary state for an infinite square well] explicitly,” “knows how to calculate the probability of an event” “does not write wave function after measurement,”) ; essential principle is applied adequately (e.g., “mistake $\psi(x) \neq a/2$”, “wrong $\psi(x)$, correct probability”)</td></tr>
<tr><td rowspan="3">C4
Elaboration</td><td>4.1</td><td>Explanation; written statements (e.g., “verbal explanations,” “narration”, “no text,” “doesn’t explain anything,” “no words,” “no statements”)</td></tr>
<tr><td>4.2</td><td>Organization (e.g., “good presentation”); showing algebraic steps (e.g., “solution in steps”)</td></tr>
<tr><td>4.3</td><td>Conciseness (e.g., “short and concise”)</td></tr>
<tr><td>C5
Correctness</td><td colspan="2">Algebraic errors (e.g., “makes sign error”); correct final answer (e.g., “final result right”)</td></tr>
</table>

Figure 6-9 shows the percentages of TAs who graded on solution features in the five clusters in the elaborated QM solution SS2 and the elaborated introductory physics solution SSD when treating the student solutions in a homework and quiz context. When grading the elaborated solutions, many TAs focused on domain knowledge in both introductory physics and quantum mechanics. However, TAs were more likely to grade on cluster C2 (explication) in QM as opposed to introductory physics. In addition, TAs were less likely to grade on cluster C5 (correctness) in QM as opposed to introductory physics.

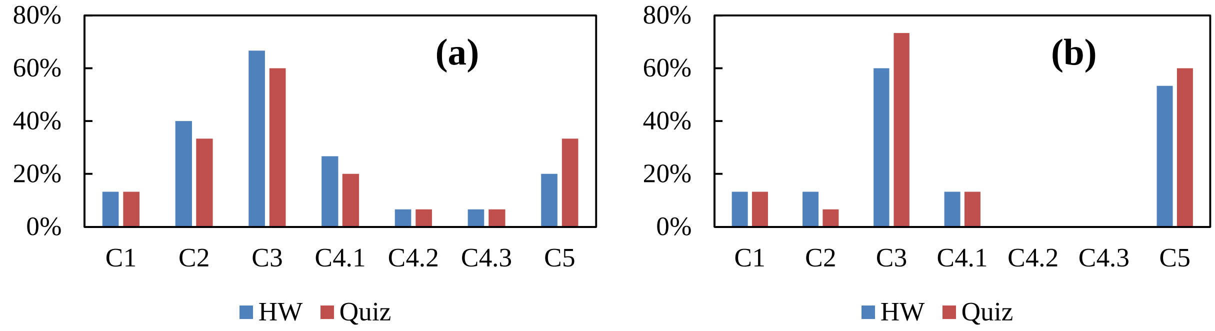


Figure 6-9 (a) Percentage of TAs who graded on solution features in the five clusters on the elaborated QM solution SS2 in the homework (blue) and quiz (red) context. (b) Percentage of TAs who graded on solution features in the five clusters on the elaborated introductory physics solution SSD in the homework context ($N = 15$ TAs).

Figure 6-10 shows the percent of TAs who graded on solution features in the five clusters on the brief QM solution SS1 and the brief introductory physics solution SSE in the homework and quiz contexts. Again, many of the TAs were focused on correct domain knowledge in both introductory physics and QM. However, the TAs were more likely to grade on C1 (problem description and evaluation) and C2 (explication) in QM as opposed to introductory physics.

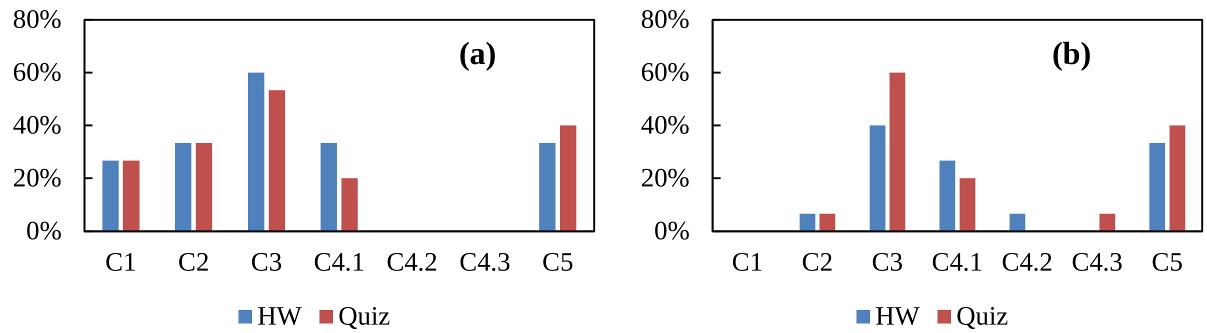


Figure 6-10 (a) Percentage of TAs who graded on solution features in the five clusters on the brief QM solution SS1 in the homework (blue) and quiz (red) context. (b) Percentage of TAs who graded on solution features in the five clusters on the brief introductory physics solution SSE in the homework (blue) and quiz (red) context ($N = 15$ TAs).

These findings suggest that TAs expect students in QM to show evidence of understanding via problem description, evaluation, and explication of their problem-solving approach. However, in grading introductory students, the TAs were mainly focused on domain knowledge and correctness. To investigate the reasons why TAs graded on different criteria in the two contexts, we discuss below TAs' stated reasons for why they graded differently in the two contexts found in written responses and interviews.

6.4.4 TAs' Reasons for Grading Differently in the QM Context and the Introductory Physics Context in Written Responses and Interviews

To investigate Research Question 4 (What are the TAs' stated reasons for whether (or not) their grading is different for introductory problems vs. QM problems?), TAs were asked to write responses to the following two questions, which were part of the QM grading activity: 1. "Was your grading approach different when grading introductory physics student solutions vs. upper-level quantum mechanics student solutions? If so, why? If not, why not?" 2. "How did your grading considerations change when grading introductory physics student solutions vs. upper-level quantum mechanics student solutions? What are the reasons for these differences?" In addition, a subset of the TAs were interviewed approximately one month after the professional development course to further clarify their views about grading solutions to QM and introductory physics problems. In their written responses, 10 of the 15 TAs (67%) noted that they would grade the QM and introductory physics problems differently, while 5 TAs (33%) noted that they would not grade differently in the two contexts. TAs' written responses about the reasons why they would grade differently or not in the two contexts were analyzed using open-coding to generate initial categories grounded in the actual data [41]. Once initial categories emerged from the data, the coding was completed by two of the researchers separately. After comparing codes, any disagreements were discussed and the categories were refined until better than 90% agreement was reached. Table 6-IV shows the categories of TAs' written responses for why (or why not) they would grade differently in the introductory physics and QM contexts, example citations, and the percentages of TAs who mentioned each category. We note that TAs could have written more than one reason for why they graded differently in the QM and introductory solution contexts.

Table 6-IV Explanation of categories used for coding TAs' stated differences/similarities when grading student solutions for introductory versus QM physics problems and percentages of TAs mentioning each category. TAs could mention more than one category so the percentages do not add up to 100%.

Category	Definition	Examples	% of TAs
More important to demonstrate understanding in QM than in introductory physics	Demonstrating understanding is more important in QM, either because it is expected of advanced students or because the subject is more complex.	--"Expect more explanations (in QM) because it's a more difficult course." --"For the upper level courses, the concepts are more complex, need more explanation."	40%
Focus more on concepts for QM and equations in introductory physics	Grading should focus more on conceptual understanding in QM and more on procedures (use of equations, calculations, solving steps, correct math) in introductory physics	--"If a student is majoring in physics, they should be able to understand all the concepts perfectly to be able to solve complicated problems." --"I will consider (grading) more on the interpretation of problems when grading upper level quantum mechanics students. As for the introductory level students, I will consider more on their calculation, solving steps".	53%
Diagrams/lists are more important in introductory physics than QM	Problem features such as diagrams and lists of unknown quantities are more important for introductory physics problems than for QM problems.	--"Upper level student should not waste time on drawing graphs that they are familiar with, they can decide if they need a diagram to help themselves." --"Focus more on concept understanding than diagram/list (in QM)."	20%
Both should have the same standards	The grading standards should be the same for introductory and QM physics problems.	--"I think whether a student majors in the field or not, they should be held up to the same standard in grading, because the difference already exists in how hard the questions are, and to reach the objective of the course, students should be expected to do things right even in introductory courses." --"I would put equal weight on different criteria and look for whether they are present/absent and correct/incorrect. That means an equal framework for both seniors (QM) and freshmen (introductory physics)."	33%

Over half of the TAs expected that students should explicitly demonstrate their understanding when solving QM problems as opposed to introductory physics problems. These

TAs put the burden of proof on the students in QM to explicate the problem-solving process. Some TAs mentioned that since QM is a more complex subject than introductory physics, advanced students in QM should demonstrate the process of problem solving and explain their reasoning in order to get credit. For example, one interviewed TA explained that she focused significantly more on proof of understanding when grading solutions to QM problems, stating: "*In QM, I don't expect people [advanced students] to be able to do things in their mind, so if they're not writing it down I kind of feel they don't know it*." Another interviewed TA also stated that since QM concepts are more abstract, advanced students should explain their reasoning when solving QM problems to get credit and added that the difference between QM concepts and introductory physics concepts is that "*in introductory physics, we can make an example to understand the questions more clearly, but for QM we [do not] have many concrete examples. We only have very abstract concepts and principles.*" This TA emphasized that the abstractness of quantum mechanics necessitates that students show their work to get credit. However, the concrete contexts in introductory physics make it easier for the TA to understand what the students' thought processes are even if students do not show their work explicitly. The TAs with these types of responses typically did not think about the abstractness of introductory physics from the perspective of an introductory student (even if introductory physics problems are often posed in concrete contexts). In particular, an introductory physics problem is challenging from the perspective of a student. However, following a systematic approach (i.e., performing a conceptual analysis of the physics problem, considering what is given and what the goals are, dividing the problem into sub-problems and making decisions regarding which principle should be applicable for different sub-problems before implementing the plan, and performing a reasonability check) can help an introductory student solve the problem correctly, and develop problem solving skills.

Other TAs claimed that advanced students should demonstrate their understanding when solving QM problems because they are already expected to have learned physics concepts as well as problem solving skills. For example, one interviewed TA stated that "*high-level students have gone through many years of training, what they need is interpret the problem [to get credit]*." This TA felt that after many years of training, students should be able to articulate their thought processes explicitly in their solution in order to receive credit. Another interviewed TA clarified her considerations when grading solutions to QM and introductory physics problems stating, "*if a student is majoring in physics, they should be able to understand all the concepts perfectly to be able to solve complicated problems. In upper-level courses, I think the student should understand everything they are doing, they are not allowed to just use an equation because they have seen people use [that equation] before.*" This TA emphasized that, in her view, a formula-fitting approach was acceptable in introductory physics courses but not in advanced physics courses for physics majors and advanced students should not receive most of the credit unless they showed their work. Another TA who valued justification of answers in the context of QM stated, "*QM students should know by now that they should justify their answers. So they should still lose points for not showing their work properly.*" Other TAs also felt that there was a distinction between physics majors and non-majors in terms of how strictly they should be graded and whether they should be penalized for not showing their work. One TA stated in an interview that he would be stricter when grading physics majors: "*I would like to be a bit stricter when grading a physics major because he's a physics major. He should grasp the idea better than those [non-major] students [and show his work to get full credit]*." These TAs did not think it was necessary to put the burden of proof for explicating the problem solving process on introductory students and they felt that a "plug-and-chug" approach was perfectly fine for a basic physics course. On the other hand, in the context of QM, these same TAs put the burden of proof on advanced students for explicating the problem solving process (in order to get a higher grade). It appears from the interviews that none of these TAs had reflected on how grading can help introductory students develop expertise, i.e., by helping them learn physics and develop effective problem solving skills. Most of these TAs felt that grading is a summative assessment only and did not realize that grading can also serve as a formative assessment activity, helping students at any level learn physics as well as develop problem solving skills.

Some interviewed TAs claimed that advanced students should focus more on demonstrating conceptual understanding while solving quantum mechanics problems but introductory students should focus mostly on formulas to solve problems. For example, one interviewed TA stated, "*in the upper level quantum mechanics, there are abstract principles and ideas that are more difficult to understand [without explanations]. So I will give more points to their correct understanding of the problems and basic ideas….*" Another TA claimed that her grading focused more on concepts in QM and that "*in introductory physics (assuming the students are not majoring in physics) it's okay if they only learn how to use equations and how to solve problems because they might have not seen physics problems before in their life, so I think they should learn step by step*." Another interviewed TA discussed differences in grading solutions to QM problem and introductory physics problem as follows, "*I will consider more the interpretation of problems when grading upper-level quantum mechanics students. As for the intro-level students, I will consider more their calculation solving steps.*" These TAs in general were more demanding of advanced students than introductory students in terms of whether they needed to clearly explain why they were using some concepts to solve a problem. However, some of these TAs who cared about students demonstrating their conceptual understanding in QM were not as critical of mathematical mistakes in QM. For example, an interviewed TA noted the differences in mathematical complexity between introductory physics and QM stating, "*the introductory physics involves more fundamental mathematics while the upper-level quantum mechanics always requires integrals or other upper-level mathematics. So I will be more tolerant to the mathematical mistakes in the upper-level physics course [but he would not tolerate if QM solutions did not clearly explain why some concepts were applied]*." The TA stated that "*in [advanced physics] exams and quizzes, skipping some steps are tolerable and minor issues compared with intro students.*"

Also, while a majority of TAs expected QM solutions to demonstrate conceptual understanding and explication of the problem solving process, about 20% of the TAs stated that drawing a diagram and listing what is known and what one is looking for are not important when solving QM problems (although they are important when solving introductory physics problems). These TAs may not have realized that drawing diagrams when appropriate and creating lists of known and unknown quantities can be useful heuristics in successfully solving a QM problem as well. In fact, for the particular QM problem the TAs were asked to grade in this study, the answer could have been checked by drawing a diagram of the wave function after the measurement of position (with a delta function in the middle of the well) and also drawing the stationary state wave functions for an infinite square well. By drawing these diagrams, one can rationalize that since the even-numbered stationary state wave functions are zero at the center of the well, the probability of finding the particle in an even stationary state would be zero.

Some TAs held contradictory beliefs regarding the importance of conceptual understanding in QM. For example, in an interview, one TA stated, "*in introductory physics we expect that the student is still learning, but when you are doing something like QM we expect that you understand the basic physics and you can easily implement it in your advanced work, so we expect somewhat more understanding.*" When asked if students learning QM face similar challenges to students learning introductory physics, the same TA continued, "*QM is in itself a difficult thing to understand…so this is a factor…the problem-solving pattern will be the same for both but the concept may be different. Conceptually, I will be lenient [when grading QM]…when it comes to getting the answer perfectly and reaching a good result...*" This TA first claims that he expects more understanding in QM, but then states that he would grade more leniently on conceptual understanding in QM since QM is difficult.

Five out of the 15 TAs noted that both introductory physics and QM should have the same grading standards. Although five TAs stated that their grading practices would be similar in the QM and introductory physics contexts, a comparison of their scores on the QM and introductory physics solutions shows that three out of these five TAs scored the brief introductory physics solution higher than the elaborated introductory physics solution on a quiz. On the other hand, none of these five TAs scored the brief QM solution higher than the elaborated QM solution in a quiz context. Thus, there is contradiction in what some of these TAs claim they would do and what they actually do.

One TA explained why she graded QM and introductory student solutions similarly stating, "*they should be held up to the same standard in grading, because the difference already exists in how hard the questions are, and to reach the objective of the course, students should be expected to do things right even in introductory courses.*" In the interview, this same TA stated, "*the difference should be embodied in the difference of the questions, not the grading. The way you do things should be held up to the same standard for all kinds of students.*" This TA appears to have realized that even though the topic may be different in introductory physics and QM, the grading should focus on similar standards for all students. Another TA, who had graded the brief solutions lower than the elaborated solutions in both the QM and introductory physics contexts, noted that he did not think that grading should change based upon the level of the student: "*it is the same physics, different concepts. If s/he is at this [advanced] level, no need to grade different.*" Furthermore, another TA noted that he would use similar standards for grading in both the QM and introductory physics contexts, stating "*I would put equal weight on different criteria and look for whether they are present/absent/correct/incorrect. That means an equal 'framework' for both seniors [advanced students] and freshmen [introductory physics students]*."

6.5 SUMMARY AND DISCUSSION

We find that most of the TAs were aware of problem solution features that can help students learn in both introductory physics and QM and included those features in their prepared solutions for their students. In particular, all of the solutions that the TAs were asked to create for the introductory physics problem and the QM problem included effective problem-solving approaches. This finding indicates that most TAs realize that giving students worked-out solutions that include good problem-solving strategies and explication of reasoning may be useful for helping them learn.

However, the TAs usually did not want to penalize introductory students for not explicating the problem-solving process. In particular, many TAs graded introductory student solutions differently than student solution in QM. Only one-third of the TAs stated that they would grade introductory student solutions and QM solutions in a similar manner. An analysis of their actual grading shows that even among those TAs, some were stricter in grading QM solutions than the introductory solutions. A majority of the TAs expected elaborated solutions that explicated the problem solving process and explicitly demonstrated conceptual understanding from students in a QM course but not from students in introductory physics courses. In other words, many TAs put the burden of proof of explicating the problem solving process on the students in QM. On the other hand, in introductory physics, they often put the burden of proof of explicating the problem solving process on themselves and inferred correct understanding when there was no evidence of it.

The differences in grading in QM and introductory physics are partly due to the fact that TAs had not thought about how grading can help students learn physics and develop problem solving skills even in an introductory physics course. Many of the TAs had not thought about learning from an introductory student's perspective and claimed that solving introductory problems using a formula centered approach was fine. Prior research has shown that TAs often view solutions to introductory physics problems as obvious [15]. In the study presented here, many TAs explicitly noted that introductory physics was easy and mainly required matching formulas to the knowns and unknowns in the problem, and they did not expect introductory students to explicate the problem solving process and show conceptual understanding. This viewpoint was a factor in why they graded introductory solution mainly on correctness as opposed to on the explication of the problem-solving approach. On the other hand, in the QM context, many TAs were able to think from a student's perspective more easily than in the introductory physics context. They perceived the QM problem to be more difficult than an introductory physics problem, and expected students in QM to explicate their reasoning and show good problem solving approaches in order to obtain a higher grade. These TAs failed to realize that introductory physics is also highly abstract for introductory students. They did not realize that, similar to advanced students learning QM, demanding that introductory students explicate and demonstrate evidence of conceptual understanding in their problem solving can help them learn physics and develop problem solving skills.

The differences in grading QM and introductory physics are also partly due to the fact that TAs had not thought about how grading can serve as a formative assessment tool and support student learning. Many TAs claimed that introductory students are novices and it is fine for them to focus only on the formulas during problem solving. On the other hand, these same TAs claimed that students in QM should demonstrate conceptual understanding and how they arrived at their answer clearly because they have learned more physics and are expected to use effective problem solving approaches. TAs who claimed that advanced students had learned more physics and developed better problem solving skills should be graded on explication of problem solving process but introductory students, who were not good at problem solving and physics, need not be held accountable for showing their work did not think about how effective grading practices can help students learn physics and develop problem solving skills. These findings suggest that TAs were not cognizant of the role of grading in promoting effective problem solving approaches and learning physics. Even though many TAs were aware that students could learn from worked-out solutions that included effective problem solving approaches, they did not realize that grading on explication of the problem solving process can help introductory students develop expertise, i.e., help them learn physics and develop problem solving skills. Most TAs thought of grading solely as a summative assessment of student learning (thought that the sole purpose of grading was to evaluate what students had learned so far) as opposed to a formative assessment that can help students learn better.

Some interviews suggest that at least some of the TAs' beliefs about grading in QM and introductory physics may have been positively impacted by the intervention in the professional development course. For example, one interviewed TA stated, "*In QM I don't expect people to be able to do things in their mind, so if they're not writing it down I kind of feel they don't know it, but in introductory physics since I can do it in my mind, I [used to] think that intro students can do it too. But then I learned that that's not the case, if I can't do quantum in my mind then they can't do [introductory physics] in their mind.*" Further conversation with the TA suggests that she was learning to put herself in her students' shoes and was beginning to recognize that advanced students learning QM are similar to introductory students learning introductory mechanics. At the beginning of the semester this TA gave SSE a score of 8/10 in the HW context and noted in her explanation of her score that "*the final answer is correct.*" However, when asked at the end of the semester to grade SSE once again, she gave SSE in the HW context a score of 6/10 and wrote the following: "*As a homework problem, the student has to show me that they understand what is going on and write down the steps.*" The interview and grading data suggest that she was starting to realize that she needed to put the burden of proof on the student and require evidence of understanding even in introductory physics student solutions. It is possible that this shift in her opinion was the combined effect of the grading activities in the professional development course and her own experiences in teaching and learning.

6.6 IMPLICATIONS

This case study investigated whether physics graduate TAs grade students in introductory physics and quantum mechanics using different criteria and the reasons for the differences. Our findings suggest that many TAs expect students to demonstrate conceptual understanding and desired problem solving practices when grading QM problems but do not necessarily penalize introductory physics solutions in which those features are missing. Moreover, it appears that the TAs in general struggled to put themselves in the shoes of their introductory physics students with regards to the difficulty of the subject matter. They often claimed that solving introductory problems is a straightforward plug and chug activity in which conceptual understanding is not as important as finding formulas and correct implementation of mathematical steps. Several TAs felt that quantum mechanics is more challenging and they expected that students in QM courses should show their work in order to be rewarded with a good grade. These same TAs felt that introductory physics is relatively easy and they did not expect introductory physics students to show their work to get a higher grade. Thus, the TAs did not think about the difficulty of a subject matter from the students' perspective, e.g., the fact that introductory physics is challenging for introductory students even though it is easy for the TAs.

Some TAs noted that it is appropriate to demand explication of the problem solving process from advanced students taking QM because advanced students have learned more physics and developed better problem solving skills. However, they did not expect introductory students to use effective problem solving strategies since they may not have learned these strategies and may not be good at using them. The interviews, class discussions, and written responses suggest that many of the TAs had not thought about their learning goals for introductory students. In particular, they had not thought about how grading can support learning goals and can serve as a formative assessment tool (instead of only thinking about the value of grading for summative assessment). For example, demanding that students explicate the problem solving process appropriately in order to get good grades can help students learn physics and effective problem solving strategies. Interviews and class discussions suggest that for most TAs, even in the context of QM, TAs' insistence that students show their work serves only for the grader to understand what the students knew so that they can be graded fairly. TAs often did not realize that explicating the problem solving process could aid students in becoming better at problem solving and help them learn physics.

TAs may benefit from an early discussion of the difficulties that introductory physics students face when solving introductory-level problems and the importance of grading criteria in helping students learn and develop better problem-solving approaches. In particular, the professional development of TAs may be improved by "framing" for TAs via activities and explicit discussions that the challenges encountered by introductory students when solving introductory physics problems are analogous to those that advanced students face in solving QM problems. It is also possible that as the TAs gain more experience working with introductory physics students, they may understand their difficulties better and develop better appreciation for the ideas brought up by the activities in the TA professional development course (including how grading can serve a formative purpose and students at all levels should be asked to explicate the problem solving process while solving problems).

The findings of this study have implications for the professional development of the TAs. Asking TAs to grade student solutions in both introductory physics and QM using the same general rubric and comparing their performance in the two contexts is an effective probe for understanding their grading beliefs and practices. Alternatively, in a professional development course, TAs can be asked to discuss and reflect on the findings of this study in order to improve their grading beliefs and practices. Leaders of professional development courses/programs for physics graduate TAs and physics education researchers can take advantage of these findings. Helping TAs value and grade students' solutions on the process of problem solving requires extended time, discussion, support, feedback and practice. The professional development courses/programs can allow more time and support for the TAs to internalize how grading can be used for formative assessment and support students in developing better problem solving practices and learning physics better [28-30]. It may be helpful to encourage the TAs to explicitly think about their own problem-solving approaches when they solve the QM problem and why those approaches would also help introductory students when they solve problems in introductory physics. TAs might be asked to list the ways in which they are similar to introductory physics students, possibly helping TAs realize that students learning introductory physics face similar challenges to advanced students learning QM. As a result of these multi-faceted professional development experiences, TAs may begin to view grading as a means to support student learning in both introductory and advanced physics.

6.7 CHAPTER REFERENCES

1. E. Yerushalmi, C. Henderson, K. Heller, P. Heller, and V. Kuo, Physics faculty beliefs and values about the teaching and learning of problem solving. I. Mapping the common core, Phys. Rev. ST PER **3**, 020109 (2007).

2. K. Heller and P. Heller, *The Competent Problem Solver for Introductory Physics* (McGraw-Hill, New York, NY, 2000); 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).

3. A. Van Heuvelen, Learning to think like a physicist: A review of research based instructional strategies, Am. J. Phys. **59**, 891 (1991).

4. F. Reif, Systematic problem solving, in *Applying Cognitive Science to Education: Thinking and Learning in Scientific and other Complex Domains* (MIT Press, Cambridge, MA, 2008), pp. 201-227.

5. J. Mestre, J. Docktor, N. Strand, and B. Ross, Conceptual problem solving in physics, in *Psychology of Learning and Motivation, Vol. 55*, edited by J. Mestre and B. Ross (Academic Press, New York, NY, 2011), pp. 269-298.

6. C. Singh, When physical intuition fails, Am. J. Phys. **70**(11), 1103 (2002).

7. D. Maloney, Research on problem solving: Physics, in *Handbook of Research on Science Teaching and Learning*, edited by D. Gable (MacMillan, New York, NY, 1994), pp. 327-256.

8. W. Leonard, R. Dufresne, and J. Mestre, Using qualitative problem solving strategies to highlight the role of conceptual knowledge in solving problems, Am. J. Phys. **64**, 1495 (1996).

9. T. Nokes-Malach, K. VanLehn, D. Belenky, M. Lichtenstein, and G. Cox, Coordinating principles and examples through analogy and self-explanation, Eur. J. Psych. Educ. **28**, 1237 (2013).

10. R. Atkinson, A. Renkl, and M. Merrill, Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps, J. Educ. Psych. **95**, 774 (2003).

11. C. Henderson, E. Yerushalmi, V. Kuo, P. Heller, and K. Heller, Grading student problem solutions: The challenge of sending a consistent message, Am. J. Phys. **72**, 164 (2004).

12. E. Yerushalmi, E. Marshman, A. Maries, C. Henderson, and C. Singh, Grading practices and considerations of graduate students at the beginning of their teaching assignment, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN*, edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2015), p. 287.

13. C. Henderson, E. Marshman, A. Maries, E. Yerushalmi, and C. Singh, Instructional goals and grading practices of graduate students after one semester of teaching experience, *Proceedings of the 2014 Phys. Ed. Res. Conference, Minneapolis, MN*, edited by P. Engelhardt, A. Churukian, and D. Jones (AIP Conf. Proc., Melville, NY, 2015), p. 111.

14. A. Schoenfeld, Toward a theory of teaching-in-context, Issues in Education **4**(1), 1 (1998).

15. A. Mason and C. Singh, Surveying graduate students' attitudes and approaches to problem solving, Phys. Rev. ST PER **6**, 020124 (2010).

16. J. Pellegrino, N. Chudwosky, and R. Glaser (Eds.), *Knowing What Students Know: The Science and Design of Educational Assessment* (National Academy Press, Washington, DC, 2001).

17. P. Black and D. William, Assessment and classroom learning, Assessment in Education **5**(1), 7 (1998).

18. A. Schoenfeld, When good teaching leads to bad results: The disasters of "well-taught" mathematics courses, Educational Psychologist **23**(2), 145 (1988).

19. A. Elby, Another reason that physics students learn by rote, Am. J. Phys. **67**(7), 552 (1999); T. Crooks, The impact of classroom evaluation practices on students, Review of Educational Research **58**(4), 438 (1988).

20. T. Angelo and K. Cross, *Classroom Assessment Techniques: A Handbook for Faculty* (National Center for Research to Improve Postsecondary Teaching and Learning, Ann Arbor, MI, 1998).

21. M. Chi, M. Lewis, P. Reimann, and R. Glaser, Self-explanations: How students study and use examples in learning to solve problems, Cognitive Science **13**, 145 (1989).

22. M. Chi and K. VanLehn, The content of physics self-explanations, J. Learn. Sci. **1**, 69 (1991).

23. 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 (Erlbaum Associates, Hillsdale, NJ, 2000), pp. 161-238.

24. E. Yerushalmi, C. Henderson. W. Mamudi, C. Singh, and S. Lin, The group administered interactive questionnaire: An alternative to individual interviews, *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE*, edited by S. Rebello, C. Singh, and P. Engelhardt (AIP Conf. Proc., Melville, NY, 2012), p. 97.

25. S. Lin, C. Singh, W. Mamudi, C. Henderson, and E. Yerushalmi, TA-designed vs. research-oriented problem solutions, *Proceedings of the 2011 Phys. Ed. Res. Conference, Omaha, NE*, edited by S. Rebello, C. Singh, and P. Engelhardt (AIP Conf. Proc., Melville, NY, 2012), p. 255.

26. D. Schwartz, J. Bransford, and D. Sears, Efficiency and innovation in transfer, in *Transfer of Learning From a Modern Multidisciplinary Perspective*, edited by J. Mestre (Information Age, Greenwich, CT, 2005), pp 1-51

27. G. Hatano and K. Inagaki, Two courses of expertise, in *Child Development and Education in Japan*, edited by H. Stevenson, J. Azuma, and K. Hakuta (W. H. Freeman and Co., New York, NY, 1986), pp. 262-272.

28. P. Black and D. William, Inside the black box: Raising standards through classroom assessment, Phi Delta Kappan **92**(1), 81 (2010).

29. P. Black and D. William, Developing the theory of formative assessment, Educational Assessment, Evaluation and Accountability **21**(1), 5 (2009).

30. P. Black, C. Harrison, C. Lee, B. Marshall, and D. William, *Assessment for Learning: Putting It Into Practice* (Open University Press, Buckingham, 2003).

31. E. Yerushalmi, E. Cohen, A. Mason, and C. Singh, 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).

32. B. White and J. Frederiksen, Inquiry, modeling, and metacognition: Making science accessible to all students, Cognition and Instruction **16**(1), 3 (1998).

33. A. Mason and C. Singh, Helping students learn effective problem solving strategies by working with peers, Am. J. Phys. **78**, 748 (2010).

34. F. Lawrenz, P. Heller, and R. Keith, Training the teaching assistant, J. Col. Sci. Teach. **22** (1992).

35. C. Sandifer and E. Brewe (Eds.), *Recruiting and Educating Future Physics Teachers: Case Studies and Effective Practices* (American Physical Society, College Park, MA, 2015).

36. C. Singh, Categorization of problems to assess and improve proficiency as teacher and learner, Am. J. Phys. **77**(1), 73 (2009).

37. C. Singh, Rethinking tools for training teaching assistants, *Proceedings of the 2009 Phys. Ed. Res. Conference, Ann Arbor, MI*, edited by M. Sabella, C. Henderson, and C. Singh (AIP Conf. Proc., Melville, NY, 2009), p. 59.

38. S. Lin, C. Henderson, W. Mamudi, E. Yerushalmi, and C. Singh, Teaching assistants' beliefs regarding example solutions in introductory physics, Phys. Rev. ST PER **9**, 010120 (2013).

39. J. Larkin, J. McDermott, D. Simon, and H. Simon, Expert and novice performance in solving physics problems, Science **208**, 1335 (1980); A. Schoenfeld, What's all the fuss about metacognition?, in *Cognitive Science and Mathematics Instruction*, edited by A. Schoenfeld (Lawrence Erlbaum, Hillsdale, New Jersey, 1987), pp. 189-215.

40. See https://www.aip.org/statistics/data-graphics/demographic-profile-physics-phds-classes-2010-2011-2012-combined.

41. A. Strauss, *Qualitative Analysis for Social Scientists* (Cambridge University Press, London, 1987).

FUTURE OUTLOOK

The studies discussed in this dissertation can be extended in several possible ways. The study discussed in Chapter 2 can be extended by including several modifications in the use of JiTT and peer instruction with clicker questions in a future quantum mechanics course. For example, a future course could implement the changes proposed in the discussion in Chapter 2, such as including more pointed questions into the pre-lecture assignments in which students would have to apply what they learned to specific situations. Also, the amount of time devoted to lecture compared to clicker questions could be modified so that students are given more opportunities for peer discussions. A comparison of clicker results from the upper-division quantum mechanics course with the results from a large introductory classical mechanics course could help determine whether the use of peer instruction in a quantum mechanics course yields gains that are comparable to those observed at the introductory level.

The study discussed in Chapter 3 can be expanded upon by investigating the effectiveness of the component of the QuILT that was developed to help students connect their conceptual understanding and reasoning in terms of “which-path” information with a mathematical formalism. This component of the QuILT accounts for photon polarization states and polarizers with various orientations in a double-slit experiment involving single photons sent one at a time through the slits. The students can make predictions about the pattern that will appear on the screen after a large number of photons are sent to the screen, then use the math to check their predictions. A future study could incorporate this component of the QuILT and include additional pre- and post-test questions which involve some of the mathematical aspects covered by the homework component.

The study discussed in Chapter 4 can be modified by reversing the order in which the double-slit experiment QuILT and Mach-Zehnder interferometer QuILT are administered in an upper-division quantum mechanics course. The order of the QuILT administration has been reversed for a set of graduate students but not for any undergraduate students. Further investigation of transfer in undergraduate students may yield interesting results. The study in Chapter 3 showed that graduate students underperformed compared to undergraduate students on the post-test. It is possible that undergraduate students may demonstrate a different degree of transfer compared to graduate students, especially if the difference in grade incentive between the two groups potentially affects their performance.

Chapters 5 and 6 focus on studies involving the grading beliefs and practices of TAs regarding the use of rubrics when grading and the use of different criteria when grading QM vs. introductory physics problems, respectively. The rubric study could be extended by investigating the grading practices of the TAs after they have had more experience in their roles as graders. It is possible that the discussions and grading assignments in the TA training class left some TAs in a state of disequilibrium by the end of the study, and that those TAs hold contradictory beliefs regarding grading (as evidenced by interviews with them). By extending the study to include a more longitudinal investigation of the evolution of the TAs' beliefs and practices, it is possible that the intervention could yield greater long-term differences than indicated by the current study. Additionally, it would be interesting to compare the grading considerations of TAs before and after they complete an assignment as a grader in a QM course. By gaining more experience with grading in QM, TAs may have the opportunity to reflect more deeply on the role of grading in QM and introductory physics courses. This could potentially strengthen their connections with the ideas discussed in their TA training course and help them to discern the similarities between the challenges faced by students in an introductory physics course and the challenges faced by students in a QM course.