

Reworking exam problems to incentivize improved performance in upper-division electrodynamics

Andrew J. Mason

Department of Physics and Astronomy, University of Central Arkansas, Conway, AR 72035

John S. Colton

Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602

Abstract: A previous study showed that incentivizing students to correct mistakes on unit exam problems within an upper-division quantum mechanics course improved students' problem-solving efforts on those same problems in a final exam environment, relative to a comparison group of students who were not incentivized. We attempt to replicate the quantitative portion of this study within a first-semester upper-division electromagnetism course, specifically examining students' invoking correct concepts and applying those concepts correctly. A statistical comparison of students who accepted the offer to rework unit exam problems for partial credit, versus students who declined the offer, demonstrates a better improvement for students who chose to rework relative to students who declined. As the results suggested that unit exam performance might provide a covariate within the comparison of choice to rework between groups, the results were analyzed using ANCOVA; to understand the effect size, a pre-post normalized gain comparison was also made; statistical results were consistent across both measurements. Results additionally appear to show that incentivization works more specifically for invoking correct concepts on a primarily conceptual problem, and more specifically for applying concepts correctly on a primarily algorithmic problem. Future plans include a more complete analytical framework using think-aloud protocol interviews for students from the sample, as well as more statistical detail to determine the interaction between unit exam score and choice to rework problems.

I. INTRODUCTION

Even at the upper-division level, undergraduate physics majors have novice-like tendencies in courses such as first-semester quantum mechanics [1]. Students often miss the opportunity to learn from problem solving in that they do not reflect upon their problem-solving approaches or their mistakes [2,3]. Cognitive apprenticeship [4,5] recommends reinforcement via scaffolding to assist students in metacognitive reflection upon their own mistakes.

Brown et al. found that explicitly incentivizing students to rework problems in quantum mechanics caused the incentivized students to significantly improve their performance on the final exam, relative to a comparison group that was not incentivized to rework the problems [3]. Their identification of difficulties within QM topics led to future research on specific topics, e.g. conceptual topics, such as wave-particle duality in the double-slit experiment [6], and mathematical methods, e.g. calculations of expectation values [7].

The work done by Brown et al. inspired the authors of this paper to conduct a similar investigation into upper-division electrodynamics. While student difficulties in upper-division topics will have similarities across different courses, the question is raised as to the differences between upper-division electrodynamics and quantum mechanics in students' approaches to problem solving. In addition, studies on upper-division electromagnetism are relatively sparse in the literature, most prominently conceptual understanding via tutorials [8]. At explicit look at problem solving in electromagnetism similar to Brown et al. is warranted.

One aspect of student experience that differs between upper-division electromagnetism and quantum mechanics is arguably that students typically have an introductory physics course to prepare them conceptually for upper-division electromagnetism. While many undergraduate curricula include all or part of a modern physics course to prepare students for quantum mechanics, quantum mechanics courses typically introduce new concepts (e.g. physical states as a linear combination of eigenstates) which are a cause for student struggle in and of themselves [6,7,9]. However, topics at the introductory level lay groundwork for further depth at the upper-division level of electromagnetism; accordingly, difficulties with introductory concepts, e.g. Gauss' law [10], may cause issues with understanding upper-division electromagnetism. At the same time, potential for difficulties at the upper-division level may also entail difficulties with algorithms based on higher

mathematical rigor, even if the concepts are well understood from an introductory level.

The authors' research questions are as follows. First, there is an overall question about whether incentivized reworking of unit exam problems will help students perform better on the same problems featured verbatim on the final exam semester in a first-semester upper-division electromagnetism course, in keeping with the model set by Brown et al. [3] for a first-semester quantum mechanics course.

The second question is to determine whether different aspects of problem solving, specifically, invoking the correct concepts versus applying them correctly, will have different effects on primarily conceptual versus primarily algorithmic problems. Three selected problems, as explained in Section II, are respectively a mix of conceptual and algorithmic problem, a more purely conceptual problem, and a more purely algorithmic problem. The second research question is whether there is a particular effect on invoking correct concepts, vs. applying those concepts correctly, that may vary from problem to problem based upon conceptual vs. algorithmic focus.

II. PROCEDURE

Three semesters of upper-division Electromagnetism 1 (EM 1) were taught by the same instructor at a large private research university in the Mountain West region of the US: the Fall 2019 (F19) section (27 overall students), the Spring 2021 (S21) section (19 overall), and the Fall 2021 (F21) section (44 overall). The F19 and F21 semesters were regular 15-week semesters, while the S21 semester was a 7-week semester with double the instruction time per week. Students were provided with a consent form to sign granting permission for data collection and analysis in the study. Of the overall course populations, 25 students consented to participate from the F19 semester, 14 students from the S21 semester, and 36 students from the F21 semester, for a total of 75 students overall.

For each course, the instructor used mostly traditional lectures, but with specific active learning components added. Discussions at the beginning of each class were typically in the form of formative (zero-point) conceptual quizzes, to review topics from previous lectures and introduce topics from the current lecture. Other active learning techniques included periodic questioning of students during the lecture (Socratic dialogue), either as individuals or as part of paired activities (e.g. student volunteers to assist with solving worked problems).

Three unit exams and a final exam were assigned as part of the EM 1 course. All exams were administered at a dedicated testing center at the host university's campus, outside of regular class time. Students were

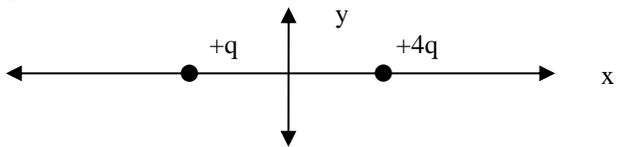
| Name | Unit exam | Text |
|------|------------------------------|---|
| P1 | Exam 1 (F19, S21, F21) | Two point charges of equal but opposite charge are separated by a distance d , the $+q$ charge being on the left and $-q$ on the right. If the charges are each moved a distance $d/2$ away from each other, what is the change in potential energy of the system? Specify whether the potential energy has increased or decreased, and give a conceptual explanation for why this is the case. |
| P2 | Exam 1 (S21, F21) | Two unequal charges are assembled on the x -axis as shown below.  <p>a) Make a sketch of the electric field lines on the figure. b) In the space below make a rough plot of E_x vs. x for points along the x-axis. Don't worry about any numbers, just the general shape of what happens to the x-component of the field – which can be positive or negative depending on whether \mathbf{E} points to the right or the left.</p> |
| P3 | Exam 2 (S21, F21) | A uniform charged rod with linear charge density $+\lambda$ lies on the z -axis, extending from $z = 0$ to $z = -d$ as shown. Suppose you want to calculate the potential at point P which lies on the positive z -axis. a) What is the monopole contribution to the potential at point P, V_{mono} ? b) In order to get a little more accuracy than the monopole potential, consider the dipole potential: $V \approx V_{\text{mono}} + V_{\text{dip}}$. (You may or may not have noticed, but the dipole formula does not actually require there to be both positive and negative charges, although typically we think of dipoles in those terms.) What is the dipole contribution to the potential at point P, V_{dip} ? c) In order to get even a little more accuracy than the monopole and dipole potentials combined, consider the quadrupole potential: $V \approx V_{\text{mono}} + V_{\text{dip}} + V_{\text{quad}}$. What is the quadrupole contribution to the potential at point P, V_{quad} ? |

FIG. 1. The three problems chosen for analysis in this study, as well as which unit exam (of three, not including final) each problem was featured in. Included in parentheses is the semester(s) for which each problem was used.

permitted to bring a single page of handwritten notes to the exam, as well as a calculator if needed, but nothing else. Students were allowed to take the exam any time within a limit of a certain number of days (five for the regular F19 and F21 semesters, three for the abbreviated S21 semester), without a specific time limit to finish the exam once it was started. After each unit exam was graded and returned, students were provided the opportunity to rework missed points on all problems on each question, with a deadline to resubmit the work typically within five days of initially receiving back the graded exams.

Certain unit exam problems were chosen to be featured verbatim on the final exam for each course, respectively labeled P1, P2, and P3 (see Fig. 1). P1 presents a combination of an algorithmic solution and a conceptual follow-up question; P2 is more fully a conceptual set of questions concerning electric field lines around point charges; and P3 is more fully an algorithmic question concerning spherical harmonics for multiple polar moments of electric potentials. Only

P1 was available for the F21 cohort; however, all three problems were available for the S21 and F21 cohorts. For these specific problems, one of the authors created rubrics similar to those created for the QM study, and consulted with the other author and external third parties to ensure validity. As with Brown et al. [3], each rubric differentiated the problem solution attempt into two main categories: invoking the correct concepts, and applying those concepts correctly. Overall rubric scores were calculated as an unweighted average of the invoked score and the applied score for each individual student. Inter-rater reliability was established with a sample of unit exam student solutions for problems selected from the unit exams, some of which were included in the study and some of which were not, for which two raters independently scored solutions with the rubrics. The raters were able to achieve 81.5% initial agreement on rubric scoring for the chosen sample before discussion, and were able to ultimately establish 90.8% agreement on these after double-checking the evaluations individually. Once inter-rater reliability was

TABLE I. Summary of students who respectively were and were not eligible for reworking problems on each featured problem in the unit exams and final exams. See Fig. 1 for the text associated with each problem ID number.

| Semester (total #) | Problem | Eligible? (< 90% on unit exam attempt) | | | Total eligible attempts (reworked + no rework) |
|-----------------------|---------|--|---------------------|----|---|
| | | Yes, reworked | Yes, did not rework | No | |
| F19 (25) | P1 | 8 | 7 | 10 | 15 |
| S21 (14) | P1 | 4 | 4 | 6 | 8 |
| | P2 | 7 | 2 | 5 | 9 |
| | P3 | 4 | 1 | 9 | 5 |
| F21 (36) | P1 | 6 | 17 | 13 | 23 |
| | P2 | 11 | 6 | 19 | 17 |
| | P3 | 9 | 13 | 14 | 22 |

established, the researchers were thus able to rate all students across the selected problems on both the unit exam and the final exam.

For each selected problem, quantitative analysis was used to compare students who had the opportunity to improve their unit exam scores, split into students who chose to rework unit exam problems and students who chose not to. Students who scored at least 90% on the unit exam version of each problem were not included in quantitative study for that problem, on the grounds that a 90% score implied sufficient mastery to have little need and/or incentive for reworking. Table I shows how students in each semester section fared on the selected problems. Students were free to choose whether to rework any of the three featured problems, e.g. one student may choose to rework one problem but not the other two, while another student may choose to rework two problems but not the other problem. Accordingly, in Table I, each problem attempt per each student is treated as an individual data point; the data points are separated into two groups, namely attempts in which the problem was reworked and attempts in which the opportunity was declined.

Analysis was performed between groups for pre-post gains in two separate ways. First, 1-way ANOVA was used for pre, post, and normalized gain scores between

groups, with consideration for equality or inequality of variances; this approach afforded a preliminary view of effect sizes on these measurements via Cohen's d. Second, while there was little statistical difference between groups on either the unit exam attempt or the final exam attempt of any problem, a few comparisons showed a difference in average unit exam scores that suggested statistical significance. We therefore also include results for ANOVA-based ANCOVA, considering the unit exam score as a potential covariate, and determining whether pre-post gains were truly dependent on choice of rework.

III. RESULTS

Table II shows the results for unit exam attempt, final exam attempt, and normalized gains for all three student sections, analyzing between groups of eligible students who chose to rework a given unit exam problem and those who did not for each problem across P1, P2, and P3. Table III shows the p-values and effect size (by Cohen's d) for normalized gains via 1-way ANOVA, as well as the p-values for ANCOVA-based interpretation of pre-post gains; both analyses give similar p-value results, with more significance for P2 with the ANCOVA method.

TABLE II. Comparisons between groups of students who reworked each problem and students who declined to rework, among the eligible students across all three sampled semesters. Analysis interprets the unit exam attempt as a pretest and the final exam attempt as a posttest. Unit/final exam attempt values are average percentages correct.

| | Problem | P1 (F19, S21, F21) | | P2 (S21, F21) | | P3 (S21, F21) | |
|--------------------|------------|--------------------|----------|---------------|----------|---------------|----------|
| | | Reworked | Declined | Reworked | Declined | Reworked | Declined |
| | N | 18 | 28 | 18 | 8 | 13 | 14 |
| Unit exam attempt | Invoke (%) | 60.3 | 75.9 | 64.1 | 72.9 | 67.7 | 67.2 |
| | Apply (%) | 67.1 | 73.8 | 59.3 | 64.6 | 54.0 | 69.9 |
| | Total (%) | 63.7 | 74.8 | 61.7 | 68.8 | 60.9 | 68.6 |
| Final exam attempt | Invoke (%) | 83.8 | 77.7 | 87.2 | 74.5 | 85.3 | 84.3 |
| | Apply (%) | 89.5 | 76.7 | 90.7 | 74.0 | 90.8 | 76.2 |
| | Total (%) | 86.6 | 77.2 | 89.0 | 74.2 | 88.1 | 80.3 |
| Gain <g> | Invoke | +0.513 | +0.211 | +0.638 | +0.337 | +0.509 | +0.519 |
| | Apply | +0.621 | +0.226 | +0.757 | +0.412 | +0.788 | +0.310 |
| | Total | +0.683 | +0.276 | +0.709 | +0.379 | +0.629 | +0.482 |

TABLE III. Analysis for statistical significance (in p-values) and effect size (in Cohen’s d) between groups described in Table II. Statistically significant results (defined as $p < 0.05$) are in bold.

| | | P1 (F19, S21, F21) | | P2 (S21, F21) | | P3 (S21, F21) | |
|--|--------|--------------------|--------------|---------------|-------------|---------------|--------------|
| | | p-values | Effect size | p-values | Effect size | p-values | Effect size |
| Gain $\langle g \rangle$, 1-way ANOVA | Invoke | 0.071 | 0.559 | 0.14 | 0.643 | 0.96 | 0.021 |
| | Apply | 0.011 | 0.799 | 0.12 | 0.691 | 0.014 | 1.017 |
| | Total | 0.0066 | 0.861 | 0.11 | 0.713 | 0.39 | 0.339 |
| ANOVA- based ANCOVA | Invoke | 0.144 | | 0.029 | | 0.92 | |
| | Apply | 0.016 | | 0.052 | | 0.022 | |
| | Total | 0.012 | | 0.040 | | 0.13 | |

Although normalized gains for P2 are not statistically significant in 1-way ANOVA test, when ANCOVA is done with unit exam score accounted for as a covariate, the analysis suggests significance or near-significance for choice of reworking for P2. The significance also seems to be more robust for invoking than for applying; this may be due to P2 being a more conceptually focused question, in which students had to draw electric field lines rather than solve for a quantity with algebra and/or calculus-based algorithms. On the other hand, P3, which was more explicitly an algorithmic problem, and P1, which had both conceptual and algorithmic portions, show more significance for applying concepts correctly than for invoking the correct concepts.

The normalized gain comparisons for which there are large ($d > 0.8$) effect sizes suggest that reworking problems P1 and P3 appears to particularly benefit applying correct concepts, more so than invoking correct concepts on the two problems with an algorithmic focus. P2 appears to have a moderate-to-large effect size ($0.5 < d < 0.8$) for students who chose to rework that problem.

IV. DISCUSSION

A. Preliminary Answers to Research Questions

To answer the first research question, students who chose to rework unit exam problems in order to correct their mistakes performed better on the final exam attempts of those same problems than did students who chose not to rework the problems. This appears to be shown by way of consistent effect sizes across almost all normalized gains, and moreover accounts for unit exam score as a potential covariate. To answer the second research question, reworking the problem that was conceptually focused (P2) appeared to help students more with invoking the correct principles, whereas reworking the problems that were fully or partially algorithm-focused (P1 and P3) appeared to help

students more with correctly applying principles that were valid for those problems.

Due to the relatively small sample sizes, we acknowledge that there are limitations with regard to the statistical rigor in this preliminary study. We note that comparisons between groups across all three problems yielded mostly null results for both the unit exam and the final exam, hence this tempers the interpretation of statistical significance for pre-post measurements.

B. Discussion and Future Research

The presented research represents a portion of a planned analytical framework to further understand students within upper-division electromagnetism [11]. First, qualitative data from interviews using think-aloud protocol of students who have finished the EM 1 course sections will identify expert-like vs. novice-like approaches to learning problem solving within the presented course structure, as has been previously done with quantum mechanics [1,9], and seeking evidence of issues noted in recent literature for this course [12]. Second, the statistical significance of the respective comparisons the pretest data in Table II suggests that there may be some interaction between unit exam score and choice to rework the problem, e.g. students who tend to perform worse on the unit exam are more likely to rework the problem. More detailed statistical analysis, e.g. regression analysis of partial vs. full vs. complete models, will shed further light upon this interaction.

Finally, the potential for coordination across different upper-division topics [13] prompts the authors to look more closely at similarities and differences between quantum mechanics and electromagnetism, ideally within the same cohort of students.

V. ACKNOWLEDGEMENTS

We thank C. Singh, R. P. Devaty, and J. McCardell for helpful conversations to initiate this study, and P. White for statistical analysis interpretation assistance.

-
- [1] A. Mason and C. Singh, “Do advanced physics students learn from their mistakes without explicit intervention?” *Am. J. Phys.* 78(7), 760-767 (2010).
- [2] C. Henderson and K. Harper, Quiz corrections: “Improving learning by encouraging students to reflect on their mistakes,” *Phys. Teach.* 47, 581 (2009).
- [3] B. Brown, A. Mason, and C. Singh, “Improving performance in quantum mechanics with explicit incentives to correct mistakes,” *Phys. Rev. PER* 12, 010121 (2016).
- [4] A. Collins, J. S. Brown, and S. E. Newman, “Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics,” in *Knowing, learning, and instruction: Essays in honor of Robert Glaser*, edited by L. B. Resnick, pp. 453-494 (1989).
- [5] V. P. Dennen and K. J. Burner, “The cognitive apprenticeship model in educational practice.” In *Handbook of research on educational communications and technology, 3rd edition*, Ed.s J. M. Spector, M.D. Merrill, J. van Merriënboer, and M. P. Driscoll, L Erlbaum Associates, pp. 425-439 (2008).
- [6] R. Sayer, A. Maries, and C. Singh, “Quantum interactive learning tutorial on the double-slit experiment to improve student understanding of quantum mechanics.” *Phys. Rev. Phys. Educ. Res.* 13(1), 010123 (2017).
- [7] E. Marshman and C. Singh, “Investigating and improving student understanding of the expectation values of observables in quantum mechanics.” *Eur. J. Phys.* 38(4), 045701 (2017).
- [8] e.g. J. P. Zwolak and C. A. Manogue, “Assessing student reasoning in upper-division electricity and magnetism at Oregon State University,” *Phys. Rev. ST – PER* 11, 020125 (2015); C. Baily, Q. X. Ryan, C. Astofoli, and S. J. Pollock, “Conceptual assessment tool for advanced undergraduate electrodynamics,” *Phys. Rev. PER* 13, 020113 (2017); B. Xue, “Identifying and Addressing Student Difficulties with the Source-Field Relationship using Tutorials in Upper-Division Electromagnetism,” dissertation, University of Washington (2021).
- [9] A. Mason & C. Singh, “Surveying graduate students’ attitudes and approaches to problem solving.” *Phys. Rev. ST – PER* 6(2), 020124 (2010).
- [10] e.g. A. Maries, S.-Y. Lin, and C. Singh, “Challenges in designing appropriate scaffolding to improve students' representational consistency: The case of a Gauss's law problem.” *Phys. Rev. Phys. Educ. Res.* 13(2), 020103 (2017).
- [11] e.g. Q. Ryan, B. Wilcox, and S. Pollock, “Student difficulties with boundary conditions in the context of electromagnetic waves.” *Phys. Rev. PER* 14(2), 020126 (2018).
- [12] e.g. D. N. Chari, H. D. Nguyen, D. A. Zollman, and E. C. Sayre, “Student and instructor framing in upper-division physics,” *Am. J. Phys.* 87, 875 (2019).
- [13] e.g. B. Wilcox, M. D. Caballero, C. Baily, H. Sadaghiani, S. Chasteen, Q. Ryan, and S. Pollock, “Development and uses of upper-division conceptual assessments.” *Phys. Rev. ST – PER* 11(2), 020115 (2015).