

The effect of giving explicit incentives to correct mistakes on subsequent problem solving in quantum mechanics

Benjamin R. Brown,¹ Chandralekha Singh,¹ and Andrew Mason²

¹*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260*

²*University of Central Arkansas, Conway, AR 72035*

One attribute of experts is that they are likely to learn from their own mistakes. Experts are unlikely to make the same mistakes when asked to solve a problem a second time, especially if they have had access to a correct solution. Here, we discuss a study spanning several years in which advanced undergraduate physics students in a quantum mechanics course were given identical problems in both the midterm exam and final exam. Approximately half of the students were given incentives to correct their mistakes in the midterm exam and they could get back up to 50% of the points lost on each midterm exam problem. The solutions to the midterm exam problems were provided to all students in both groups but those who corrected their mistakes were provided the solution after they submitted their corrections to the instructor. The performance on the final exam on the same problems suggests that students who were given incentives to correct their mistakes significantly outperformed those who were not given an incentive. The incentive to correct the mistakes had the greatest impact on the final exam performance of students who performed poorly on the midterm exam.

I. INTRODUCTION

One characteristic of experts is that they are likely to use problem solving as an opportunity for learning [1–3]. In particular, experts are likely to reflect upon their mistakes in their problem solution automatically in order to repair, extend and organize their knowledge structure. Unfortunately, for many students in physics courses, problem solving is a missed learning opportunity [4–7]. Without guidance, many students do not reflect upon the problem solving process after solving problems in order to learn from them, nor do they make an effort to learn from their mistakes after the graded problems are returned to them [4–7].

However, closing the “performance gap” between high and low achieving students and ensuring that all students excel in physics courses are important goals of physics education research [8, 9]. Prior research suggests that only providing students worked examples is insufficient [10] and effective approaches to learning involve engaging students in meta-cognition or self-monitoring while they solve problems [11, 12]. For example, 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 up until a consolidation lecture by the instructor [13]. However, these students from the productive failure condition significantly outperformed their counterparts from the lecture and practice condition on both well- and ill-structured problems on the posttests [13]. After the posttest, they also demonstrated significantly better performance in using structured-response scaffolds to solve problems on a new topic not even covered during instruction. Similarly, Schwartz et al. have proposed invention tasks to prepare students for future learning [14].

It is often implicitly assumed that, unlike students in introductory physics courses, most students in advanced courses have become independent learners. They will invest time to learn from their own mistakes, even if the instructors do not reward them for fixing their mistakes, e.g., by explicitly asking them to correct their mistakes by giving them grade in-

centives [4, 5]. Contrary to these beliefs, our earlier investigation found that advanced students in a quantum mechanics course did not automatically improve their performance from midterm to final exam on identical questions even when they were provided the correct solutions and their own graded exams [6]. There was a lack of reflection by supposedly mature students; many students did not make use of mistakes identified on a midterm exam as an opportunity to repair and better organize their knowledge. In individual interviews with some of the students, we also found evidence that even in these advanced courses, many students do not automatically learn from their mistakes and they often resort to rote learning strategies for getting through the course.

These issues are particularly important considering that the diversity in the prior preparation of students at all levels has increased. Many students need explicit guidance not only in the introductory courses, but also in the advanced courses. One instructional strategy that may help is to explicitly prompt students to learn from their mistakes by rewarding them for correcting the mistakes. This activity can also help them learn to make use of problem solving as a learning opportunity.

II. GOAL AND METHODOLOGY

Here we report on an investigation focused on how well students in a junior-senior level undergraduate quantum mechanics course learn from their mistakes on midterm exam problems when provided with time and grade incentives for re-submitting written corrected solutions after the graded midterm exams were returned to them. Using a comparison group that did not self-diagnose mistakes on the midterm exams, we investigated the effects of this self-diagnosis of mistakes on subsequent problem solving on some of the same problems repeated on the final exam.

The study took place over four years with the same instructor and textbook and the course was primarily taught in a traditional lecture format. Students were assigned weekly homework throughout the fifteen-week semester. For each course,

there were two midterm exams and a final exam (all exams and homework were identical across the four years). Both midterm exams covered only limited topics while the final exam was comprehensive. Students were instructed in all of the relevant concepts before the exams, and homework was assigned each week from the material covered in that week. The first midterm exam took place approximately eight weeks after the semester started; the second midterm exam took place four weeks after the first midterm exam. For this study, two problems were selected from each of the midterm exams and were repeated verbatim on the final exam along with other problems not asked earlier.

In the first and third year of the study (total 33 students), students were not provided any explicit incentives to learn from their mistakes on the midterm exam (comparison group). However, in the second and fourth years (total 30 students), students were given a grade incentive to correct their mistakes (incentivized group): they could earn up to 50% of the points they had lost for the corrections. Students in the incentivized group were directed to work on their own while correcting mistakes, but were free to use any resources including homework, notes, and books for help. Solutions to all of the midterm exam questions were available to students on a course website except that the incentivized group was provided the solutions after the students had submitted their corrected midterm exam solutions.

Our goal was to explore the extent to which students in each group use their mistakes as a learning opportunity and whether their performance on the same problems administered a second time on the final exam (which we call the posttest) is significantly better than the corresponding performance on the midterm exams (which we call the pretest). In particular, we looked for correlation between the midterm exam score and final exam score on the four common problems for each group (incentivized and comparison groups).

All problems were graded according to a rubric developed by the researchers [6] with better than 95% inter-rater reliability for 25% of the students graded separately (see the summary of the rubric used for problem 1 in Table I). As shown in Table I, on the problem, each student received an overall score, which is the average of the Invoking and Applying scores. Each of these general criteria (invoking and applying) scores is in turn the average of the specific criteria scores (see Table I), which can take on the values 0, 1, 1/3, 1/2, 2/3, or N/A. "N/A" arises in situations when certain criteria are not relevant in determining a score.

III. RESULTS

The comparison group and incentivized group had nearly identical average performance on the pretest as shown in Table II (analysis of variance or ANOVA shows $p=0.972$). However, the distribution of posttest performance of the incentivized group is significantly better than the comparison group with a p -value of 0.001. Table II also shows that generally the students with lower performance on the pretest benefit more from the explicit incentive to correct their mistakes in the pretest.

TABLE I. Summary of the rubric used for problem 1. In this problem, students are asked to find the expectation value of an observable Q in terms of eigenstates and eigenvalues of the corresponding operator.

General Criteria	Specific Criteria
Invoking appropriate concepts	Spectral decomposition expressing identity operator in terms of a complete set of eigenstates $ \psi_n\rangle$: $\hat{I} = \sum \psi_n\rangle \langle \psi_n $
	Or expressing general state in terms of the eigenstates of \hat{Q} : $ \psi\rangle = \sum c_n \psi_n\rangle$, where $c_n = \langle \psi_n \psi\rangle$
	Make use of $\hat{Q} \psi_n\rangle = \lambda_n \psi_n\rangle$
	$\langle \psi_n \psi\rangle^* = \langle \psi \psi_n\rangle$
	Using legitimate principles or concepts that are not appropriate in this problem.
Applying concepts	Using invalid principles or concepts (for instance, confusing a general state $ \psi\rangle$ with an eigenstate $ \psi_n\rangle$).
	Inserting spectral decomposition into the expression for expectation value
	Eigenvalue evaluated and treated as number.
	Probability expressed in terms of $\langle \psi_n \psi\rangle$ and $\langle \psi \psi_n\rangle$

TABLE II. Pretest, posttest, and gain for students in the comparison and incentivized groups broken down into low, medium and high performance categories based on students pretest score and for all students. While the pretest scores are comparable for the comparison and incentivized groups, the posttest scores are significantly higher for the incentivized group. The gap between low and high categories shrinks from 42.2% in the comparison group to 19.9% in the incentivized group.

	Comparison Group ($N = 33$)			Incentivized Group ($N = 30$)		
	Pre	Post	Gain	Pre	Post	Gain
<i>Low</i>	34.6	50.8	+16.2	33.2	78.2	+45.0
<i>Medium</i>	64.7	66.3	+1.6	67.3	88.5	+21.3
<i>High</i>	96.0	93.0	-2.9	96.9	98.1	+1.2
All	67.9	71.5	+3.6	67.6	88.4	+20.8

Figure 1 shows the average gain (defined as the arithmetic difference between posttest and pretest scores; gain can therefore range from -100% to +100%) vs pretest score (which can range from zero to 100%) for each student on the four questions repeated from the pretest to posttest in the comparison group and incentivized group.

The regions of possible gain are shaded according to posttest score performance categories: green for High posttest performance, yellow for Medium posttest performance and orange for Low posttest performance. The performance categories are defined as follows: High, for scores from 85% to 100%; Medium for scores from 50% to 85%; and Low for scores from zero to 50%. The researchers agreed upon a 50% cutoff somewhat arbitrarily for separating Low and Medium categories and 85% cutoff was chosen so that roughly one third of the students scored in the High performance category on the pretest.

Figure 1 and Table II show that students with poor performance on the midterm exam were likely to benefit from self-diagnosis activities in which they submitted the corrected midterm solutions for 50% of the points lost on each problem. Therefore, the gap between the High and Low performers on the midterm exam was reduced for this incentivized group on the repeated problems on the final exam. On the other hand, for the comparison group, the gap remained, i.e., scores did not substantially improve, and remained diverse.

The data were analyzed by breaking the students into three groups based on their pretest performance as shown in both Figure 1 and Table II. The initially high-performing students from both the comparison and incentivized groups (scoring 85% and higher on the pretest) generally performed very well on the posttest regardless of the intervention (see Figure 1). Most of these students who start in the High pretest category stay in that category. Students who initially performed at a Medium level on the pretest (scoring between 50% and 85%) in the incentivized group perform better on the posttest than the corresponding students who were in the comparison group. In the comparison group, students in the Medium performance category on the pretest were as likely to improve on the posttest (above the horizontal axis in Figure 1A) as they were to deteriorate (below the horizontal axis). In contrast, in the incentivized group, almost all of the students in the Medium category on the pretest improved on the posttest (see Figure 1B).

Furthermore, about half of these students in the incentivized group improved as much as possible on the posttest, saturating the boundary for maximal improvement (see Figure 1B). Among the initially Low performing students (pretest scores less than 50), many students in both comparison and incentivized groups improved on the posttest. However, the degree to which these struggling students performed on the posttest is highly dependent on whether or not they received a grade incentive to improve. The students in the Low category on the pretest in the incentivized and comparison groups had an average gain of 45.0% and 16.2%, respectively (see Table II). In summary, the gains are much larger for the incentivized group, bringing the average of the Low category to the level of the Medium category, and the Medium category to the High category (see Figure 1 and Table II).

IV. DISCUSSION

An easy-to-implement intervention substantially reduced the performance gap between low- and high- achieving students in quantum mechanics. Explicitly providing grade incentives to correct mistakes positively affected the performance of students with a diverse spectrum of performance in the midterm exam on subsequent solving of the same problems in the final exam. Our research suggests that even students in advanced physics courses such as quantum mechanics are more motivated to engage with instructional material in a meaningful way if they are provided an explicit incentive [15–17]. Considering the relative ease with which instructors in physics courses at all levels can implement the intervention in

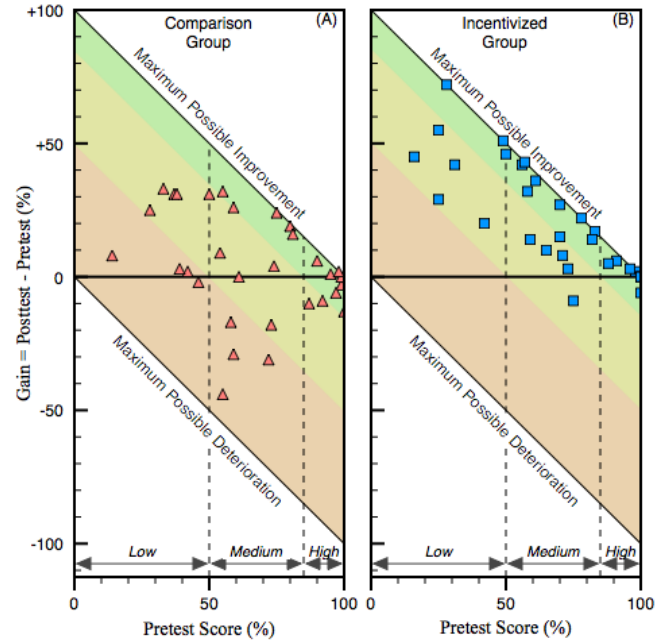


FIG. 1. Average gain (defined as the difference between posttest and pretest score) vs pretest score for each student on the four questions repeated from the pretest to posttest in the comparison group (A) and incentivized group (B). A few students' data points overlap. Red-filled triangles are for each student from the comparison group and the blue-filled squares are for each student in the incentivized group in which students received an explicit grade incentive to correct their own mistakes on the pretest before answering the same questions on the posttest. Students whose scores improved are above the horizontal axis; students whose performance deteriorated are below the horizontal axis.

which students are given grade incentives to correct and learn from their mistakes, instructors at all levels should consider giving students this opportunity to learn. Asking students to correct their mistakes in several courses may also help students understand the importance of learning from mistakes and the role of appropriate struggle in learning.

Providing students with incentives to learn from their mistakes may positively impact learning without a significant additional effort on the part of the instructors and without requiring a substantial initial or recurring investment of resources. We find that in quantum mechanics, the corrections themselves take very little time to grade because students seize the opportunity and produce expert-like solutions even when they are given only a few days to correct their own mistakes. We note that while students submitted both their initial and corrected solutions, explanations of the mistakes they had made and how they corrected them was limited since the instructor (who graded them) did not explicitly tell them that they would be graded on such explanations.

While these results are encouraging, caution is urged in interpreting improvement. Our findings support the claim that students improve on problems administered a second time when expert-like behavior (self-diagnosing and correcting

their mistakes) is explicitly incentivized. It does not necessarily follow that students have become adept at self-monitoring skills from just two such interventions in the two midterm exams in quantum mechanics. In particular, we compared the performance of students in the incentivized and comparison groups on four other problems (on other topics) on the same final exam for which incentivized group students did not diagnose their mistakes. We find that while the incentivized group scored somewhat higher than the comparison group, the results are not statistically significant on student performance on those four other problems on the final exam.

V. SUMMARY AND CONCLUSION

We find that the performance of students in the group in which no incentives were provided shows that while some advanced students performed equally well or improved compared to their performance on the midterm exam on the questions administered a second time, a comparable number of students obtained lower scores on the final exam than on the midterm exam. The wide distribution of students' performance on problems administered a second time suggests that many advanced students do not automatically exploit their mistakes as an opportunity for learning, and for repairing, extending, and organizing their knowledge structure. On the other hand, the performance on the final exam on the same problems suggests that students who were given incentives to correct their mistakes significantly outperformed those who were not given an incentive. The incentive to correct the mistakes on the midterm exam had the greatest impact on the final exam performance on those problems for students who performed poorly on the midterm exam.

An explicit incentive to correct their mistakes can be an effective formative assessment tool [18]. Offering grade incentives to diagnose and correct mistakes can go a long way to close the performance gap between struggling and high-performing students. If this type of easy-to-implement intervention is implemented routinely in all physics courses, students are likely to use their mistakes as a learning opportunity and may even develop better self-monitoring skills over time.

Considering the relative ease with which instructors in physics courses at all levels can implement the intervention in which students are given grade incentives to correct and learn from their mistakes, instructors should consider giving students at all levels such an opportunity to learn.

ACKNOWLEDGMENTS

We thank NSF for award PHY-1202909.

Appendix A: An Example Solution

Below, we provide a student response on problem 1 from the incentivized group to show how a typical student improved from the pretest to posttest.

1) The eigenvalue equation for an operator \hat{Q} is given by $\hat{Q}|\psi_i\rangle = \lambda_i|\psi_i\rangle$, with $i = 1\dots N$. Find an expression for $\langle\psi|\hat{Q}|\psi\rangle$, where $|\psi\rangle$ is a general state, in terms of $\langle\psi_i|\psi\rangle$.

$$\begin{aligned} \langle\psi|\hat{Q}|\psi\rangle &= && \text{Pretest} \\ &= \sum_{i=1}^N \langle\psi|\lambda_i|\psi_i\rangle &= \sum_{i=1}^N \langle\psi|(\langle\psi_i|\psi\rangle)|\psi_i\rangle \\ &= \sum_{i=1}^N |\langle\psi_i|\psi\rangle|^2 \end{aligned}$$

$$\begin{aligned} \langle\psi|\hat{Q}|\psi\rangle &= \sum_{i=1}^N \langle\psi|\hat{Q}|\psi_i\rangle \langle\psi_i|\psi\rangle && \text{Posttest} \\ &= \sum_{i=1}^N \langle\psi|\lambda_i|\psi_i\rangle \langle\psi_i|\psi\rangle \\ &= \sum_{i=1}^N \lambda_i \langle\psi|\psi_i\rangle \langle\psi_i|\psi\rangle &= \sum_{i=1}^N \lambda_i |\langle\psi_i|\psi\rangle|^2 \end{aligned}$$

-
- [1] F. Reif, J. Larkin and G. Brackett, *Am. J. Phys.* **44**, 212 (1976).
[2] M. Chi, R. Glaser and M. Farr, *The Nature of Expertise*, (Lawrence Erlbaum, Hillsdale, NJ, 1988).
[3] K. Ericsson and J. Smith, *Toward a General Theory of Expertise: Prospects and Limits*, (Cambridge, UK, 1991).
[4] E. Yerushalmi et al., *Phys. Rev. STPER* **8**, 6 (2012).
[5] E. Yerushalmi et al., *Phys. Rev. STPER* **8**, 5 (2012).
[6] A. Mason and C. Singh. *Am. J. Phys.* **78**, 760 (2010).
[7] C. Henderson and K. Harper, *The Physics Teacher*, **47**, 581 (2009).
[8] J. Watkins and E. Mazur, *J. Coll. Sci. Teach.* **42**, 36 (2013).
[9] M. Lorenzo, C. Crouch and E. Mazur, *Am. J. Phys.* **74**, 118 (2006).
[10] R. Atkinson, S. Derry, A. Renkl and D. Wortham, *Review of Educational Research* **70**, 181 (2000).
[11] A. Schoenfeld, in *Cognitive science and mathematics education* edited by Schoenfeld (Lawrence Erlbaum, Hillsdale, NJ, 1987).
[12] G. Schraw, *Instructional Science*. **26**, 113 (1998).
[13] M. Kapur, *Cognition and Instruction*, **26**, 379 (2008); *Learning: Research and practice*, **1**, 51 (2015).
[14] D. Schwartz et al., *Cognition and Instruction*, **22**, 129 (2004).
[15] D. Stipek, *Motivation to learn: From theory to practice*, (Allyn and Bacon, Boston, MA, 1993).
[16] C. Ames, *J. Ed. Psych.* **84**, 261 (1992).
[17] C. Dweck and E. Leggette, *Psych. Rev.* **95**, 256 (1988).
[18] P. Black and D. Wiliam, *Assessment in Education: Principles, Policy and Practice*, **5**, 7 (1998).