Comparing retrieval-based practice and peer instruction in physics learning

Tianlong Zu,1,* Jeremy Munsell,1 and N. Sanjay Rebello1,2
1Department of Physics and Astronomy, Purdue University, 525 Northwestern Ave., West Lafayette, Indiana 47907, USA
2Department of Curriculum and Instruction, Purdue University, 100 N. University St., West Lafayette, Indiana 47907, USA

(Received 20 September 2018; published 10 January 2019)

Even though it has been shown that retrieval practice could foster deeper learning and better long-term retention in other domains such as psychology, it is rarely studied in the context of physics learning where students need to solve more complicated problems. To an even lesser degree is comparing retrieval-based learning with other active learning style methods adopted in the physics classroom. In this study, we compared the effects of retrieval-based learning and peer instruction based restudying on physics problem solving and transfer. In both conditions (retrieval and peer instruction), participants were first presented video lectures explaining the definition of speed and energy conservation. In the training session that immediately followed the video lectures, the retrieval condition was asked to recognize, recall, and apply the relevant physics concepts to solving problems; while the peer instruction condition was asked to discuss the two video lectures with each other in a group of three or four members. After this training, an immediate and a delayed (one-week) final test were administered. Both tests contained an initial task, isomorphic to the training materials, near transfer tasks, and far transfer tasks for each topic. Subjective judgment of learning (JOL) was collected immediately after watching the video lectures, after the training, and right before the delayed final test. We found a retention advantage of the retrieval practice for the initial tasks, but not for the near and far transfer tasks. We also found an advantage of the retrieval practice on far transfer on the immediate final test. Peer instruction training inflated the participants’ JOL compared to the retrieval practice. Both JOL scores were significantly lower than after the video lecture but did not differ from before the delayed final test.

DOI: 10.1103/PhysRevPhysEducRes.15.010105

I. INTRODUCTION

Educators expect students to learn the relevant concepts and apply them to solve tasks in similar as well as novel situations compared to those used in instruction. These objectives have been the focus of education studies for a long time [1]. Recognition and retrieval of the relevant concepts are necessary steps toward successful problem solving and transfer [2]. In educational psychology studies, a burgeoning line of research using retrieval-based testing purports the idea that repeated testing after initial instruction promotes more learning and retention than just studying once or restudying the same material [3–7]. Given the aforementioned advantages of retrieval-based learning, studies have also found evidence that retrieval-based learning can promote transfer [8,9].

Even though it has been shown that retrieval practice could foster deeper learning and better long-term retention in other domains such as psychology, it is rarely studied in the context of physics learning where students need to solve more complicated problems. To an even lesser degree is comparing retrieval-based learning with other active learning style methods adopted in the physics classroom such as peer instruction. According to the social constructivism, interacting with peers after initial learning or problem solving could further increase individual learning as evidenced by previous studies [10,11]. Peer instruction works because it allows more interstudent communication [12], enables students to reach a deeper conceptual understanding, to provide better justifications of their arguments [13], and affords learners the opportunity to co-construct their knowledge [14]. However, few studies looked at how peer instruction might affect transfer and retention.

In this study, we compared retrieval-based testing and peer instruction regarding promoting problem solving, transfer, and retention of knowledge. We were also interested in how each intervention influences students’ metacognitive performance since previous studies have found no significant correlation between judgment of memory
...and actual performance [15]. While research on retrieval-based learning found that students in the repeated studying condition made less accurate judgments of learning than the retrieval-based testing condition [16]. In this study, we asked students to subjectively assess how well they comprehended the learning material. We were also interested in how individual differences such as working memory capacity (WMC) might be correlated with the JOL and performance of the students in our two conditions since it has been found that individual differences in WMC is correlated with reading comprehension [17]. Individual differences could play a factor in determining the best instructional strategies for different students [18]. For example, according to the expertise reversal effect [19], experts learn better from problem solving, but novices learn better from worked examples. Individual differences in working memory capacity could limit the amount of information learners can process during encoding according to cognitive load theory [20]. Another motivation for this study is that it is important to replicate significant new results, since researchers have been repeatedly demanding independent replication of novel findings [21–25]. Thus, in this work, we investigated if the testing effect could be applicable to learners of different working memory capacity.

In summary, our research questions are (i) How does students’ performance compare after retrieval-based practice or peer instructed restudying on immediate and delayed final tests? (ii) How is students’ judgment of learning affected by retrieval-based practice or peer instructed restudying? (iii) How are test performance and judgment of learning affected by individual differences in working memory capacity?

II. METHODS

A. Participants and materials

We recruited 68 students enrolled in a physics class for future elementary teachers to participate in this study. Participants received extra credit equivalent to about 4% of their course grade. Since most of these students did not have previous physics coursework, we prepared two multimedia video lectures explaining the two topics covered for this study: the definition of speed and the energy conservation principle. The two videos were designed closely following the principles of the cognitive theory of multimedia learning [26]. Each video began with the introduction of the relevant concepts, followed by one example task for demonstrating how to apply the concepts. The average duration of the videos was about 6 min. Video lectures similar in style to those used here have been shown to be effective exam preparation for low performing students [27], and for promoting transfer [28]. The two lecture videos used in this study can be accessed following the link in Ref. [29].

There is no clear definition in the literature on the distinction between near transfer and far transfer. In our study, we operationalized near and far transfer tasks as follows. Each task has three dimensions: underlying principle (e.g., energy conservation), context (e.g., car on a road, ball rolling, etc.), and representation (graphical or pictorial). The near transfer task differs from the learning task only in terms of the context. The far transfer task differed from the learning task in both context and representation (example near and far transfer problems can be seen in Fig. 2–3 in Appendix A).

B. Procedure

The study was conducted as part of the regular lab session for the course. Students came to the lab and seated themselves in front of a laptop computer in a group of three or four, with each group seated around a table. After everybody was ready, they were instructed to open a pre-installed Qualtrics survey and then follow this procedure (see Fig. 1):

Step 1: Watch two multimedia videos which were presented to the students in a randomized order.

Step 2: Provide their judgment of learning on a Likert scale in answer to the question: “How well have you comprehended the material? Please rate on a scale from 0 (I have comprehended 0% of the material) to 100 (I have comprehended 100% of the material).”

Step 3: Each group was randomly assigned to one of two conditions: retrieval-based practice and peer instructed restudying.

- Retrieval-based practice condition: For each topic, the procedure followed the three steps proposed by Barnett and Ceci [2]. They suggest that transfer entails a robust understanding of the relevant knowledge domain, recognition of the surface and deep structure of the tasks at hand, successful retrieval of all the relevant conceptual and procedural knowledge, and finally effective execution of the procedural knowledge. Consistent with this notion (i) an example problem was presented which was similar to the example presented in the video lecture. Participants were asked to recognize the best principle that could be used to solve this problem from two options. (ii) Regardless of their selected option, they were asked to recall the definition of the relevant concept and (iii) how to apply it. Then they were asked to solve the example task presented. The same procedure was repeated for the second topic (See Fig. 4–5 in Appendix B).

- Peer instruction condition: A PowerPoint file containing slides from two video lectures was presented to students. Students sat with their usual lab group. This took advantage of the effective group dynamic established since the start of the semester. The group members were asked to take turns explaining each slide to the rest of the group. They were encouraged to raise questions and discuss the slides with each other.
This condition was an enhanced version of restudying since restudying in previous studies was individual. Here participants restudied the material with their peers. We acknowledge this peer instruction is operationalized differently than the one initially proposed by Mazur [11], but it was similar to other adapted versions of peer instruction in which students work in groups on a task [14,30].

Step 4: Students in both conditions were asked to rate their JOL again by answering the question: “Please rate again, how well have you comprehended the material? Please rate on a scale from 0 (I have comprehended 0% of the material) to 100 (I have comprehended 100% of the material).”

Step 5: An immediate final test containing two blocks (one for each topic) was administered. Each block contained: an initial task (similar to the task used in the instruction), a near transfer task, and two far transfer tasks. The order of two blocks was randomized. Initial tasks used in this study were from the Force Concept Inventory (FCI) validated by Hestenes and collaborators [31]. Near and far transfer tasks used in this study had been validated in previous studies [32–35].

Step 6: The same final test was administered again in one week as a retention test. Before the one-week delayed test, students were asked again to rate their JOL, using the same question as before.

Step 7: We measured each participant’s WMC using the operation span task (O-span task) developed by Unsworth and colleagues [36]. The task used in this study was conducted on a computer. Participants solved simple arithmetic operations such as \( \frac{8}{4} + 2 = ? \) and memorized a letter which was shown immediately after each mathematical operation. In each trial, the total number of operations varied from three to seven, hence participants needed to memorize between three to seven letters in each trial. After each trial, participants needed to report the memorized letters back in the presented order by clicking them from a letter matrix. In total, there were 75 operations and thus 75 letters. Participants’ data were used only if their performance on the mathematic operations satisfied an 85% correctness rate. This indicates participants were completing both sections in the O-span task, and not focusing only on memorizing the letters. These data allow us to investigate how individual differences in working memory capacity might affect participants’ performance and JOL.

III. RESULTS

A. Immediate and delayed final test performance

In the retrieval condition, the successful rate of selecting the proper underlying principle from the multiple-choice options was 88.6%, indicating most of the students could recognize the correct concept. The successful rate of recall of the definitions was 70%. Table I shows the average scores and respective S.D.s for the two conditions on the initial, immediate final test, and delayed final tests.
Results can be found in Table II. Overall, we found that far transfer, and their respective delayed tests between. Instructed restudying condition on initial, near transfer, the difference between the retrieval condition and the peer instructed restudying on the immediate as well as delayed test. The effect sizes were small to medium. To be more specific, there is no significant difference between immediate initial task score was significantly higher for the retrieval condition than the peer instruction condition. The scores of the immediate and delayed near transfer tasks did not differ between the two conditions. We found the score on the immediate far transfer tasks were marginally higher for the retrieval-based practice condition compared to the peer instruction condition. However, there was no difference between the far transfer scores on the delayed test from the two conditions.

**B. Correlation between working memory capacity and final test performance**

To look at the influence of individual difference in working memory capacity on test performance, we obtained the total score for the immediate final test as well as the delayed final test by simply summing the scores for each of the three tasks (IN, NT, FT) of the respective tests. As expected, the total scores for the two final tests are positively correlated with each other. Nevertheless, we found no correlation between working memory capacity with the two total final test scores.

**C. Judgment of learning**

We wanted to investigate how the JOLs changed compared to the 1st JOL made immediately after initial encoding with the video lectures. The means and S.D.s can be found in Table III. For the 1st and 2nd JOLs, a repeated measure ANOVA revealed a main effect of condition, $F(1,66) = 4.391, p = 0.040, \eta^2_p = 0.062$, such that the average of 1st and 2nd JOL was significantly higher for the peer instruction condition. It revealed no main effect of time, $F(1,66) = 1.484, p = 0.227, \eta^2_p = 0.022$, such that the average JOL from the delayed test did not differ from the average on the immediate test. It revealed a marginally significant interaction between time and condition, $F(1,66) = 3.715, p = 0.058, \eta^2_p = 0.053$. A one-way ANOVA revealed no difference between the two conditions on the 1st JOL made, $F(1,66) = 1.095, p = 0.299, \eta^2_p = 0.016$, but it revealed a significant main difference between the two conditions on the delayed far transfer task, $F(1,66) = 8.068, p = 0.006, \eta^2_p = 0.109$, such that the JOL made by the peer instruction condition was significantly higher than that made by the retrieval condition.

For the 1st and 3rd JOLs, a repeated measure ANOVA revealed no main effect of condition, $F(1,66) = 1.391,$
In this study, we directly compared the impact of two retrieval-based practice and peer instruction. To an even lesser extent, it has been shown that retrieval practice could foster deeper learning and better long-term retention in other domains such as psychology, it is rarely studied in the context of physics learning where students need to solve more complicated problems. To an even lesser degree is comparing retrieval-based learning with other active learning style methods adopted in physics classroom. In this study, we directly compared the impact of two research-based interventions—retrieval-based practice and peer instruction—on physics problem solving and transfer performance. The tasks used in this study were conceptual tasks that students could encounter on real exams. We also examined the impact of both treatments on participants’ metacognitive judgment of learning (JOL). We also investigated if individual differences in working memory capacity might be correlated with performance as well as JOL.

This study replicates the testing effect that states retrieving the learned material after initial learning leads to better retention performance than simply restudying the material [3–7]. This is evidenced by a significantly better performance of students on the delayed initial task in the testing condition than the peer instructed restudying condition. Our result here indicates a stronger testing effect since the restudying was in the form of peer interaction, which is collaborative learning and demonstrated to promote more conceptual understanding than individual studying during traditional classroom [37]. The resulting pattern in this work is different from typical findings from studies on the testing effect where restudy or repeated restudy outperforms retrieval practice on immediate final test, but is outperformed by retrieval practice on the delayed final test [3]. The difference might be explained by the activities involved during retrieval. For example, when studying educationally relevant material as in Ref. [3], students were asked to simply recall the content of the prose paragraph. Whereas during the retrieval practice in our study, we asked students to recognize the underlying principle first (without feedback), write down the relevant definitions for a given task, solve the given task. The sequence follows closely the proposition of Barnett and Ceci [2] to promote transfer. The initial better performance on an immediate test in favor of restudying could be explained by the readiness of information from primary memory, whereas the initial better performance on far transfer in favor of retrieval could be explained by the process of the retrieval practice in our study. It is possible training with only one task is not sufficient for retention of far transfer given the higher level of complexity of physics problems. It is interesting that there was no difference for the near transfer tasks on both immediate and delayed final tests. This could certainly be caused by a training effect since the order of the tasks was fixed and the near transfer task differs from the initial task only in context. The benefit of retrieval practice for far transfer only exists on the immediate final test, but not on the delayed final test. A future study should include individual studying as a control condition to see if peer instruction and retrieval would do better than control on both near and far transfer tests.

We found that restudying using peer instruction inflates students’ confidence compared to testing condition. This is consistent with the results from Roediger and Karpicke [3], where students from the repeated study condition predicted higher performance than those from the retrieval condition; it turned out that the opposite was true coming to the actual performance. When students were asked to make a JOL before the delayed final test, both groups made a significantly lower JOL than the first time but did not differ from each other. This could possibly mean students were aware of the forgetting during the past week and became more confident.
conservative in their judgment of learning. This is consistent with the finding delayed judgment of learning is more accurate than immediate JOL [38].

We also looked at how individual difference in working memory capacity might affect students’ problem-solving performance. A correlation analysis between total immediate score, total delayed score, and working memory capacity of each student revealed positive correlation between the two final total scores, but no correlation between either score and working memory capacity. The results here seem to be different than previous studies. For example, working memory capacity was found to be correlated with reading comprehension [17] and verbal SAT score [39]. This lack of correlation here might be explained by the complexity of the physics subject matter and the task which requires deeper understanding than reading comprehension. This begs for future careful studies.

We also investigated how individual differences in working memory capacity might affect students’ metacognitive judgment of learning. A correlation analysis between judgment of learning made right after initial learning, right after training, right before the one-week delayed final test, and working memory capacity revealed positive correlation between the three judgments of learning made, but no correlation between any of the judgments of learning and working memory capacity. This indicates students make their metacognitive prediction of learning in a consistent manner at different times. However, the lack of correlation between WMC and JOL was something new here. Better judgment of learning requires better recall of the information to properly assess and predict one’s learning. Studies have found that higher WMC led to better long-term memory recall [40]. Here it is possible that students did not make their JOLs based on effortful recall of the information learned. Given the current “replication crisis” in psychology [21–25], failure of replicating the significant correlation between JOL and WMC in this study requires a more scrutinized methodology. One possible improvement is to adopt multiple-item JOL measures since it could capture more nuances of the expected underlying construct [41,42] and other types of WMC measurements. This certainly begs for future careful study.

This study found some initial evidence supporting retrieval-based practice could do equally well if not significantly better than restudy with peer interaction even though studies in physics education have shown peer instruction to be a very effective strategy of promoting conceptual understanding. Considering the diverse background of students in the physics classroom, exploring and administering different flavors of instruction could benefit more students. We do not intend to advocate one method over another, rather, we will investigate in future experiments whether the combination of retrieval-based practice and peer instruction can outperform either of those two strategies. We plan to implement retrieval practice inside the physics classroom and use a larger sample size to investigate the impact of the retrieval practice on learning of students.

ACKNOWLEDGMENTS

This work is supported in part by the U.S. National Science Foundation Grant No. 1348857. Opinions expressed are those of the authors and not necessarily those of the Foundation.

APPENDIX A: EXAMPLE NEAR AND FAR TRANSFER PROBLEMS USED IN THE STUDY

Example near and far transfer problems used for the two physics topics (definition of speed & transformation of energy) used in this study can be found in this section.

Ball A begins riding downward in an elevator at the same time Ball B is dropped from the roof of an adjacent building. A snapshot of the balls is taken every second. At what point in time does Ball B have the same speed as Ball A?

A runner runs along a track. The following diagram, viewed from above, shows the position of the runner at each second. At which point in time is the runner moving the fastest?

FIG. 2. Example near (left) and far (right) transfer problems (definition of speed).
APPENDIX B: PROCEDURE FOR THE RETRIEVAL PRACTICE

In this section, the procedure used during the retrieval-practice is presented for both physics topics used in this study. This retrieval practice includes three steps: recognition, recall and execution. In the recognition phase, students need to select the physics principle they will be using for obtaining a solution; in the recall phase, students need to retrieve the content knowledge of the related concepts; in the execution phase, students need to apply the activated concepts and solve the problem at hand.

<table>
<thead>
<tr>
<th>Step 1: recognition</th>
<th>Step 2: recall</th>
<th>Step 3: execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which principle or concept is most appropriate for solving this problem? Select one from the two options.</td>
<td>What is the Definition of Speed?</td>
<td>In this motion diagram, a snapshot of a car is taken at each second. How can one determine the speed of the car between 1 second and 2 second?</td>
</tr>
<tr>
<td>• Definition of Speed</td>
<td></td>
<td>0 second</td>
</tr>
<tr>
<td>• Transformation of Energy</td>
<td></td>
<td>1 m</td>
</tr>
</tbody>
</table>

FIG. 3. Example near (left) and far (right) transfer problems (transformation of energy).

FIG. 4. Retrieval practice for the definition of speed implemented in the framework of Barnett and Ceci [2].
FIG. 5. Retrieval practice for transformation of energy implemented in the framework of Barnett and Ceci [2].


