Effective and Adoptable Metacognitive Tools

Dr. John Chen, California Polytechnic State University - San Luis Obispo

John Chen is a professor of mechanical engineering. His interests in engineering education include conceptual learning, conceptual change, student autonomy and motivation, and lifelong learning skills and behaviors.
Effective and Adoptable Metacognitive Tools

Abstract

This paper, an evidence-based practice paper, describes two metacognitive teaching tools that were tested in classroom environments for their efficacy and ease of adoption. Ease of adoption refers to the subjective measures of (a) ease of implementation, (b) minimal displacement of class time, and (c) no requirement for a change in pedagogy. The two tools promote metacognition, which has an extensive evidence base for promoting learning in a wide range of subjects and across grade levels. The first tool tested is the exam wrapper. Examinations or tests provide a measure of student performance and offer feedback to students of their learning and the need to perhaps adjust their learning strategies. Many students, however, focus on the grade rather than the comments or corrections. Even when students make the effort to look at the mistakes, they often miss the opportunity to reflect on the deeper root causes and instead focus on the superficial error. Without deep reflection students may not gain the awareness that they need to confront misconceptions or make strategic changes in their learning. The second tool tested is the assignment correction, a variant of exam wrappers but used for more frequently occurring activities such as homeworks or quizzes. The idea is that, perhaps, improving metacognition requires frequent practice. If the exam wrapper could be adapted for use with graded assignments, it would provide such practice. To remain a tool that is easy to use, however, assignment corrections must be briefer than an exam wrapper, easy to assign, collect and score, and continue to consume little to no class time for completion. The two tools were tested in various engineering courses and mixed results were found: While both tools were adoptable, only the exam wrapper appeared to be efficacious in this study.

Introduction

Metacognition, which has as its simplest definition thinking about one’s thinking, is the modern term used to capture the processes that learners use to reflect upon and take actions to improve their learning. The psychologist John Flavell\(^1\) introduced the term in the 1970’s while advancing research on the topic, but ideas about the usefulness of reflection in improving learning began much earlier, starting with John Dewey\(^2\). Both Piaget and Vygotsky – both recognized widely for their theories in education – wrote of the role of metacognition in the cognitive development of children\(^3\). Finally, in the much-heralded 2000 revision of the taxonomy of educational objectives, originally put forth by Benjamin Bloom and colleagues in 1956\(^4\), metacognition was included as a fourth dimension of knowledge (alongside factual, procedural and conceptual knowledge)\(^5\).

Despite this rich history and a wealth of research supporting its effectiveness in helping students learn, metacognition is not a teaching tool in wide use by engineering educators nor is it intentionally taught as a learning strategy to students. This paper describes the implementation and testing of two metacognitive tools for their efficacy in improving student learning and performance.
Background

Much effort and cost have been devoted to improving STEM learning over the past few decades. The majority of such projects has been focused on improving the learning of subject content and student success through, for example, improving classroom teaching [e.g., refs. 6, 7], curricular reform [e.g., ref. 8], or providing more and better design experiences [e.g., refs. 9, 10]. Little attention, however, has been given to the incorporation of reflection to promote learning and the experience of learning. The psychologist Ellen J. Langer exposes the pitfall of a lack of reflection. Learning without being mindful sometimes leads to rote exercising that could build bad habits and prevent learners from seeing how to apply knowledge learned in one context to other, very different ones. As Lang points out, “[s]tudents who have frequent opportunities to pause and reflect on what they’re practicing will develop deeper understanding and more transferable skills.”

Mindfulness in learning, when combined with self-generated control of that learning, is synonymous with the modern psychological term metacognition. Pintrich defines metacognition as the awareness, knowledge and control of one’s thinking while learning. Experts across widely varied fields, including academic disciplines, chess and various sports, actively practice metacognition, demonstrating perhaps that metacognition is a trait important not only to learning and expertise, but also to the application of that knowledge. Moreover, adept problem-solvers possess strong metacognitive skills. They are aware of flaws or gaps in their knowledge, can readily describe their thought processes, and subsequently alter their solution strategies toward achieving a desired outcome. Novices and students often lack these skills or do not know to apply them when needed.

As described by Pintrich, Flavell suggested a three-part framework for metacognitive knowledge: strategy, task and person. Strategic knowledge consists of knowledge of strategies for thinking, learning and problem solving. They are generally applicable to all or most academic disciplines, as opposed to being specific to one discipline. These may include, for example, rehearsal (e.g., repeated practice) or elaboration (e.g., providing detailed explanation or paraphrasing of a solution) or problem solving (e.g., working backward from a goal). Task knowledge, or knowledge about cognitive tasks, involves recognizing that various tasks can differ in difficulty and may therefore require different strategies. For example, a recognition task (recognizing a word in an unfamiliar language) is easier than a recall task (reproducing that word from memory). Learners must be aware of why and when to apply different skills in such situations. Finally, person- or self-knowledge describes awareness of one’s strengths and weaknesses. If a learner prefers multiple-choice rather than short-answer tests, but is faced with the latter, she would recognize this challenge and alter her study strategies in preparation (e.g., start studying earlier and outline the chapters for review). Self-knowledge also includes aspects of motivation for learning. For example, is the learner pursuing the learning through an intrinsic (“this is interesting”) or extrinsic (“I want a good grade”) orientation, and what about the learner’s self-efficacy?

Research over the past 40 years has conclusively demonstrated the effectiveness of learning accompanied by metacognition [see, for example: refs. 17, 18, 19, 20]. Although few of these studies have been based in engineering or science, the evidence seems clearly extendable to these learning environments. As Pintrich states, “Because metacognitive knowledge in general is
positively linked to student learning, explicitly teaching metacognitive knowledge to facilitate its development is needed.” Furthermore, Bransford et al.\textsuperscript{21}, in their synthesis of research on learning over the past few decades, declares the effectiveness of a metacognitive approach to instruction as one of its three key findings.

An Internet search of the term “metacognitive teaching methods” returns several hundred thousand hits. A perusal of the first two pages of this reveals scholarly web sites or articles that include suggestions for various ways that educators could incorporate metacognition into instruction to promote learning. These include short activities such as predicting outcomes, self-questioning, journaling and critiquing. Longer and more complex activities include using a rubric for self-evaluation, creating mind maps, and reflection writing. Finally, whole-class or extensive metacognitive teaching methods include student-developed tests or grading rubrics, self-assessment of assignment, and creating concept maps. Further study of the search results reveals that many metacognitive teaching methods have been the subject of scholarly study, including, for example, rubrics\textsuperscript{22}, self-assessment\textsuperscript{23}, student-written exam\textsuperscript{24} and concept map\textsuperscript{25}. These studies all demonstrate positive outcomes for student learning, attitude, or both.

Given the overwhelming evidence of effectiveness, the question is why are metacognitive teaching methods not widely adopted in science and engineering disciplines? Given this nation’s need for more and better trained engineers and scientists\textsuperscript{26}, 27, 28, 29, 30, and the greater attention given to improving undergraduate STEM teaching over the past few decades, it seems that more faculty should be incorporating metacognitive approaches in instruction. Susan Ambrose, the educator and co-author of \textit{How Learning Works}\textsuperscript{31}, directly addressed this point to the engineering education community, saying “So, yes, students learn by doing, but only when they have time to reflect on what they are doing – the two go hand in hand. Why, then, don’t engineering curricula provide constant structured opportunities and time to ensure that continual reflection takes place?”\textsuperscript{32}

The reason for the low use of metacognitive instruction is likely to be low awareness and the challenge of adoption. Metacognition is a relatively new construct and thus has low awareness among engineering and science faculty, whose graduate education included little or no educational theory or training. Even if faculty members were made aware of the importance and value of metacognitive instruction, we argue that its adoption would remain low. Many well-known instructional approaches with overwhelming research supporting their effectiveness have yet to be adopted in engineering and science teaching\textsuperscript{33, 34}. The most frequently cited reasons for non-adoption of these innovations include the displacement of course content, the fear of student resistance, and the alteration of the instructor’s preferred pedagogy.\textsuperscript{33} It is clear that overwhelming evidence is insufficient to convince faculty to adopt a new teaching method. It must also overcome the cited barriers.

\textbf{Project Rationale}

It seems that for metacognitive teaching tools (or any teaching tool, for that matter) to be widely adopted, they need to be demonstrably effective at improving some aspect of student performance and to overcome the ease-of-adoption barrier. Ease of adoption refers to the subjective measures of (a) ease of implementation with little training or preparation, (b) minimal displacement of class time, and (c) no requirement for a change in the instructor’s preferred
pedagogy. Two metacognitive tools – exam wrappers and assignment corrections – that seemingly overcome the ease-of-adoption barrier are tested in this study for their efficacy.

Examinations or tests are comprehensive assessments of several topics within a course that serve two major objectives: They provide a straightforward measure of student performance and, more importantly, offer feedback to students of their state of learning and, perhaps, the need to adjust their learning strategies prior to subsequent tests. The latter objective is accomplished through comments and corrections that the instructor provides on each exam, which, in some instances, is intended to promote the students’ use of metacognition. As Schinske and Tanner35 note, however, “the grade trumps the comment.” That is, students tend to focus on the grade rather than the comments or corrections. Even when students make the effort to look at the corrections, they often miss the opportunity to reflect on the deeper root causes and instead focus on the superficial error (e.g., “used the wrong equation” or “bad assumption”). Without deep reflection students may not gain the awareness that they need to confront misconceptions or make strategic changes in their learning.

“Exam wrapper”36 was created to address these issues. The tool is a writing-reflection exercise that is given to students immediately after a graded exam is returned. The exercise, which is typically completed outside of class for little or no credit, prompts students to reflect on and write about (a) what the student did to prepare for the exam, (b) what happened during the exam and (c) what he/she will do to prepare for the next exam to improve performance. Exam wrapper is in use in several disciplines, including college science courses37. There is scant research support for its effectiveness, especially in engineering, but in our opinion the tool has high face validity for promoting metacognition in students. This may account for its use in college teaching.

Assignment corrections are a variant of exam wrappers, but used for more frequently occurring activities such as homeworks or quizzes. The idea is that, perhaps, improving or instilling metacognition requires frequent practice, as both Lang12 and Ambrose32 suggest, which is not achieved using exam wrappers once or twice in an academic term. If the exam wrapper could be adapted for use with frequently graded assignments, it would provide such practice. To remain a tool that is easy to use, however, assignment corrections must be briefer than an exam wrapper (since it will be used frequently), easy to assign, collect and score, and continue to consume little to no class time to implement.

Research Methodology

The study population is a convenience sample of engineering students at a large, state-supported university in the western US. The testing took place at three time points that coincided with three different academic quarters and with the same instructor using generally the same pedagogy. Testing occurred in three different courses: an engineering mechanics course (Statics), introductory thermodynamics and a second thermodynamics course for mechanical engineering majors only. More details of the courses at each time point are provided in Table I.

As shown in Table I, the exam wrapper was tested at all time points while the quiz correction was only tested during the fall 2014 quarter. During each quarter, when two or more sections of the same course were tested, the sections were combined and the results are reported.
for this entire group for each course. Figures 1 and 2 below show the exam wrapper and the quiz correction used.

Table I: Description of courses and study populations in tests

<table>
<thead>
<tr>
<th>Quarter</th>
<th>Tool tested</th>
<th>Course</th>
<th>No. of sections</th>
<th>Total students</th>
<th>Majors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2014</td>
<td>EW, QC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Statics</td>
<td>3</td>
<td>~110</td>
<td>Various</td>
</tr>
<tr>
<td>Winter 2015</td>
<td>EW</td>
<td>Thermodynamics I</td>
<td>2</td>
<td>~70</td>
<td>Various</td>
</tr>
<tr>
<td></td>
<td>EW&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Thermodynamics II</td>
<td>2</td>
<td>~70</td>
<td>Mech. Eng.</td>
</tr>
<tr>
<td>Spring 2015</td>
<td>EW&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Thermodynamics I</td>
<td>3</td>
<td>~100</td>
<td>Various</td>
</tr>
</tbody>
</table>

<sup>a</sup> EW = exam wrapper; QC = quiz correction  
<sup>b</sup> Two in-term examinations were administered in this course  
<sup>c</sup> All students were given the option to complete the exam wrapper

Exam wrapper implementation

Exam wrappers were tested in all classes in the study and, with the exceptions of Thermodynamics II in the winter 2015 quarter and Thermodynamics I in spring 2015, the procedure was identical and as follows: During each 10-week quarter, a mid-term examination was administered to the class around week 6. After the exam was scored and returned, students scoring below a cut-off score were given the option to complete the exam wrapper for an extra credit of 2% added to their exam score. The cut-off score varied slightly from quarter to quarter but was generally between 70-80%. The cut-off was an attempt to separate students who made minor or non-conceptual mistakes (those presumably above the cut-off) from those who made significant conceptual or problem-solving mistakes. In addition to the instructions on the exam wrapper form (Fig. 1), students were instructed to (a) complete the exam wrapper after they have reviewed the exam solutions provided by the instructor and compared it to their own, (b) keep an electronic copy of the completed exam wrapper and (c) submit it within three days and electronically through the on-line class management system. For all courses, a final examination was administered at the end of the quarter. Five days prior to this, the instructor sent the following note via email to those students who completed the exam wrapper:

“You completed an Exam Wrapper after the midterm. Thank you for making the effort to do that. Not everyone did, but the fact that you did means you put effort and thought into it. You gave careful consideration to how you handled the midterm – before (preparing for it), during (taking it) and after (what you would like to do to prepare for the final). Please take a few minutes and read through your ideas. I really believe the thoughts and observations you made can help you on the final.”
Instructions: Answer each question after careful consideration of your specific situation. Avoid cliché answers (which show little thought...) such as “careless oversight” or “lack of preparation”. Think really hard to find the authentic answers to the questions. And, of course, please be truthful in your answers – there is no penalty for this activity, just extra credit for thoughtful responses.

I. Preparing for the midterm

1. Did you do any of the following to prepare for the midterm? 

   Not enough   just right   too much

   a. Review class notes
   b. Review old quizzes & solutions
   c. Review homeworks & solutions
   d. Work extra problems alone
   e. Work extra problems with someone
   f. Go for help from instructor
   g. Read textbook to clarify
   h. Sleep/rest
   i. Exercise or recreation

2. What else, if anything, did you do to prepare for the midterm?

3. What else could you have done before the test to better prepare yourself?

II. During the test

1. What on the midterm gave you problems? That is, what statics or math concepts caused you problems? (Please give this careful thought before answering)

2. What personal (e.g., illness, alertness, etc.) or environmental (e.g., noise, temperature, etc.) issues caused you problems during the midterm? (Please only share information you’re comfortable sharing)

III. Preparing for the final exam (please give deep thought to these answers)

1. Based on your answers and thoughts above, what would you do to better prepare yourself for the final exam?

Fig. 1: Exam wrapper. Version shown is complete but condensed to save space.

In Thermodynamics II in winter 2015, two in-term examinations were administered at approximately weeks 5 and 8. The exam wrapper was administered after each test using the procedure described above. In the three sections of Thermodynamics I in spring 2015, all students were given the opportunity to complete the exam wrapper; i.e., no cut-off score was imposed. In this case, students whose scores were below 70% received two points extra credit for completing the exam wrapper while those at or above 70% received one point for the effort.
Quiz Correction Form

INSTRUCTIONS: Do not erase or alter your original quiz; make separate correction(s) on the quiz (back side is okay) and submit the corrected quiz attached to this form.
Please fill out this form after correcting your quiz and return it attached to the quiz. Remember that this needs to be returned at the start of the next class meeting. Please also complete this form truthfully; there is no penalty or effect on the scoring of your corrected quiz. Thanks.

1. What was the reason for your mistake(s)? (For example, careless oversight, time rush, misunderstanding of concept, etc.)

2. How could you avoid similar mistakes as this/these in the near future?

3. Did you figure out your mistake(s) by yourself or did you look at a classmate’s quiz? (Please circle one)

- Figured it out myself
- Looked at someone’s quiz

Fig. 2: Assignment (quiz) correction. Version shown is complete but condensed to save space.

Quiz correction implementation

Quiz corrections were only tested once, during the fall 2014 quarter with three sections of Statics. The course structure was identical for all three sections in terms of the instructor’s pedagogy, the homework assignments, the quizzes, and the examinations. The only significant difference between the sections was the time of class, which was at 12 pm, 1 pm and 2 pm. During the quarter, 14 quizzes were administered (coinciding with the submission of the 14 homework assignments), and quiz corrections were implemented for 11 of these. After each quiz was completed, all quizzes were electronically scanned into a digital file. The instructor graded each quiz electronically using a digital tablet device with pen input. If a student scored higher than 80%, the paper copy of the quiz was marked up as usual and returned to the student. If a student scored 80% or lower, the paper copy was only marked with the final score and then returned to the student with the option to complete a quiz correction (Fig. 2). Nearly all qualifying students chose the option.

Students choosing to complete the quiz correction must (a) figure out their mistake(s) on the quiz (any method to do this was acceptable, as seen in Fig. 2, question #3) and correct it on another page, (b) fill out the quiz correction, and (c) submit both the quiz correction form and the corrected quiz at the start of the next class meeting. The instructor then grades the corrected quiz on paper and assigns a final score. The maximum score that a student can receive after doing a quiz correction is the original score plus one-half of the missed portion of the original score. For example, if a student scores 60% on the original quiz, then the maximum possible score achievable, if he returns the form and a flawless corrected quiz, is 80% \((60\% + 0.5*(100-60)\%)\). The student could of course score lower if the correction has errors.
Results and Discussion

Exam wrapper results

In fall 2014 the exam wrapper was administered to three sections of Statics using the standard procedure. Of the 112 students in the combined sections, 69 students were below the cut-off score and had the option to complete the exam wrapper; 55 chose to do so and 14 did not. The performance of these latter two groups (which we call “wrappers” and “non-wrappers” respectively) is contrasted to test for possible effects of the intervention.

Figure 3a shows, for the 14 non-wrappers, the paired performances on the midterm (grey bar) and final (black bar) exams. The figure shows the difference between an individual’s score and the mean in that individual’s section, divided by the standard deviation in that section’s test score. Showing the difference as measured in units of standard deviation will allow us to better compare performance between different sections and courses. As can be seen, the results are mixed. Some students improved from the midterm to final (scoring toward a more positive value), while others performed worse. Figure 3b shows the performance of the 55 wrappers. Again, some improved while others did not. A direct comparison of the percentage of students who improved from the midterm to the final demonstrates that the wrappers (61.8%) outperformed the non-wrappers (42.9%).

While Fig. 3 seems to demonstrate that a higher portion of students improved their test performance after completing the exam wrapper, what about quantifying the size of that improvement? Figure 4a shows the paired midterm (grey data points) and final (black) exam scores for each individual non-wraper, scaled by the mean of all scores (including those who scored above the cut-off and did not have the option to complete the exam wrapper) from the three sections. The hoped-for goal is that students will move from their midterm values to one that is higher. Again, the mixed results are obvious. Figure 4b shows the same paired-values for the wrappers. At a glance it appears that for this group more students achieved the hoped-for goal and scored higher, relative to the mean, and this is supported quantitatively. As a group the non-wrappers scored an average of 0.92 (score/mean) on the midterm and 0.93 on the final exam, which are nearly identical. The wrappers scored an average of 0.90 on the midterm (slightly below the non-wrappers) and 0.98 on the final exam – nearly achieving the class average. It is further worth pointing out that 15 out of the 16 lowest performing wrappers on the midterm achieved substantially higher scores on the final exam (Fig. 4b), moving from an average score of 0.73 to 0.99.
Fig. 3: Performance on the midterm exam (grey bar) and final exam (black bar) for (a) the 14 no-wrappers and (b) the 55 wrappers. The plot uses increasing midterm score as its abscissa. The y-axis represents an individual score as the difference of that score from the mean, divided by the standard deviation.

Fig. 4: Paired scores for the (a) non-wrappers and (b) wrappers showing each individual’s midterm (grey data point) and final (black) exam score scaled by the mean in that individual’s section. The plot uses increasing midterm score as its abscissa.
While the results from fall 2014 showed promise for the effectiveness of exam wrappers, they warrant cautious interpretation. First, the study was conducted for only one course and at one time point. Second, the sample size was small, especially for the non-wrappers (N=14). Finally, this particular test was confounded by the use of a second intervention, the quiz correction, which may have influenced the results.

To alleviate the first and third concerns the exam wrapper intervention was repeated in the next quarter in two sections of Thermodynamics I using the same standard procedure. The student population was similar to the first study in terms of the mix of majors and the time point in their academic studies. In this case 36 out of 70 students had the option to complete the exam wrapper after the midterm exam and all but two did so. The low number of non-wrappers eliminated the possibility of a meaningful comparison with the wrappers so the results here focus only on the wrappers group relative to all students comprising the two sections.

The midterm scores of the wrappers, scaled on the mean for all scores for both sections, ranged from 0.47 to 0.98, with an average of 0.84. The final exam score of the wrappers was dramatically improved, ranging from 0.60 to 1.11, with an average of 0.98. Like the first study, the wrappers as a group nearly achieved the class average on the final exam. A more profound result is the effect on the lowest achieving students. A comparison can be made between the lower-performing half of the wrappers based on the midterm (an arbitrary choice) and the entire population in this case. For this group, the scaled midterm average was 0.73 (compared to 0.84 for all wrappers) while the scaled final-exam average was 0.97 (compared to 0.98 for all wrappers), demonstrating the dramatic gain made by this group. These results seem to confirm the findings from the first study for (1) the improvement in performance on a subsequent exam for all wrappers and (2) an even stronger effect for the lowest-performing students who completed the exam wrapper.

In the spring 2015 quarter the study procedure was changed to give all students the option to complete the exam wrapper and to measure the effect of this change. Testing took place with three sections of Thermodynamics I. In this case, students whose scores were below 70% received two points extra credit for completing the exam wrapper while those at or above 70% received one point for the effort. Of the total 103 students, 60 completed the exam wrapper after the midterm. It is interesting to note that only 58% of the students completed the exam wrapper in this case, whereas in the fall 2014 quarter, in a different course but with a similar population and the same extra-credit value, 80% (55 out of 69) of eligible students completed the extra credit. The low completion percentage is not due solely to the high-scoring students choosing to skip the exam wrapper, as one might guess. The proportion of students who completed the exam wrapper for the lowest-, middle- and top-third of the class (based on the midterm) were 65%, 59% and 51%, respectively. Thus, while the completion percentage does decrease with higher midterm scores, it was low for all groups, and was lower than all previous cases in which only those students below the cut-off score were given the option to complete the exam wrapper. This finding seems to suggest that one negative consequence of allowing all students to complete the exam wrapper for extra credit is that they devalue its worth and more students – from across the performance spectrum – choose not to complete it.

Comparing the wrappers to non-wrappers in this case showed other differences from previous cases. The performance difference (again using a scaled score – the midterm or final
exam score scaled by the mean from the student’s section – as the measurand) from midterm to final exam was essentially nil, changing from 1.04 to 1.03 for the non-wrappers and 0.976 to 0.995 for the wrappers. In terms of the percentage of students who improved their performance from midterm to final exam, there again was no difference: 49% for the non-wrappers and 48% for the wrappers. These contrasting findings are likely due to the higher-performing wrapper students confounding the results. If the wrappers are further stratified into a lower- and an upper-half (based on their midterm score), the impact of the exam wrapper emerges: The lower-half wrappers improved from 0.83 to 0.96 (midterm to final) and the upper-half regressed from 1.13 to 1.03. In addition, the percentage of each group that improved from midterm to final was dramatically different: 77% for the lower-half and 20% for the upper-half. These results seemingly confirm again the positive and strong effect of the exam wrapper for the low-performing students.

The final study looked at whether the impact of the exam wrapper is sustained if it is administered twice in a term. In two sections of Thermodynamics II in the winter 2015 term, two in-term tests (at weeks 5 and 8) and one final exam were administered. The exam wrapper was implemented after both in-term tests. The standard procedure was followed otherwise. In this case the cut-off score on either test was substantially reduced in comparison to the other cases reported; this change was an arbitrary one and not by design. This resulted in much fewer students being eligible to complete the exam wrapper. Out of 140 total pre-final tests, only 34 cases were eligible for the exam wrapper; 25 of these 34 opportunities (74%) were taken by 20 unique students. The relatively few students who completed the wrapper make the sample size for this case quite small, which warrant caution in considering the results presented here. On the other hand, this population is composed of those who performed poorest on the tests, and prior results showed that this group benefitted the most from the exam wrapper.

The results are stratified and examined by whether a student completed (a) only the first exam wrapper, (b) only the second one, or (c) both. Figure 5(a), for each student who completed only the first exam wrapper, shows the scaled scores for test 1, test 2 and final exam. Six of the eight students improved substantially from test 1 (mean of scaled score = 0.88) to test 2 (mean = 1.04), but there was a slight drop-off on the final (mean = 0.95). This suggests, if the statistics are trustworthy, that (1) an initial exam wrapper improves performance on a subsequent test, and (2) the exam wrapper’s effect on the final exam (as described for previous cases) can survive a second, similar test in the time between the first test and the final exam, though it may suffer a slight degradation. Figure 5(b), for each student who completed only the second exam wrapper, shows the scaled scores for test 2 and final exam. Again, most students improved substantially (mean test 2 = 0.74; mean final exam = 0.93). This result suggests that the exam wrapper is effective in the very short term, which is not surprising. Finally, Fig. 5(c) shows the scaled scores for students who completed exam wrappers after both tests. The results show that the exam wrappers did not help in this instance – in either time point – as the mean scores were 0.86, 0.72 and 0.89 for test 1, test 2 and final exam, respectively. This suggests, perhaps, that low metacognition is not the sole or dominant weakness for this group of students.

Quiz correction results

Quiz corrections were implemented in three sections of Statics in fall 2014. A caveat to the findings reported here is the simultaneous use of the exam wrapper with this study population,
which obviously could have confounded the results. During this single study, all 111 students were grouped to form a single population. Fifty-five students submitted either zero or one quiz correction, and we will ignore this group since they did not use this tool with significantly frequent practice, which is theorized to be important for developing metacognition. The frequency of use of the quiz correction by the remaining 56 students is shown in Fig. 5.

(a)  (b)  (c)

Fig. 4: Scaled scores for students who completed (a) only the first exam wrapper, (b) only the second one, or (c) both. The plotted values are the student’s score on a particular test scaled by the mean in that student’s section of class.

Fig. 5: Frequency of use of the quiz correction among study population (N=56).
Figure 5 shows there are two distinct groups of students who used quiz corrections with some regularity throughout the quarter. One group clustered between 2-4 uses and the other between 5-7 uses over the 11 quizzes administered during the quarter. We can disaggregate the sample into these two clusters (referred to as the “low use” and “high use” groups) and examine their performance on the quizzes. One would hope that, with frequent practice of metacognition, more practice would lead to improved performance. This was not supported by the data. The quiz average of the low-use group was 87.4% while that of the high-use group was 81.6%, which are significantly different (p=0.005). This suggests that perhaps the high-use group may have other weakness(es) that cannot be overcome with more metacognition practice, or that metacognition is not a significant weakness for this group, or that the quiz correction is not an effective metacognitive tool. Despite the lack of an effect of the quiz correction on student performance, we note that this study was highly impeded by the simultaneous use of the exam wrapper. In retrospect, we would have liked to evaluate the effect of quiz corrections on the students’ performance on the midterm and final exams, which was of course not possible in this case.

Summary and Conclusion

Two metacognitive tools – the exam wrapper and the quiz correction – were tested for their effectiveness in improving student performance and ease of adoption. On the latter point, the instructor for this study found both tools to meet the criteria for easy implementation, no displacement of class time and no required change in pedagogy (so long as quizzes or tests are already a part of an instructor’s course structure). For the quiz correction, the method implemented by the investigator to correct the quizzes may be seen by some instructors as an impediment to use, but alternative methods can be developed to suit each instructor’s preferred evaluation method. The exam wrapper was truly an easy-to-adopt teaching tool.

In addition to its ease of adoption, the exam wrapper was found to be highly effective. Eligible students who elected to use the exam wrapper consistently outperformed those who were eligible but did not choose to use it: A higher proportion improved their performance on the next (final) exam and their final-exam scores were substantially higher, reaching nearly the class average in each instance. Even more impressive is the exam wrapper’s dramatic impact on students who performed the poorest on the first test, moving them up on average to nearly achieving the class average on the final exam. Although easy to adopt, the quiz correction did not show any effect on improving quiz performance by students who used it frequently. The single study of this tool, however, was highly confounded and impeded by the simultaneous use of the exam wrapper.

References


26. National Academy of Sciences, Committee on Science, Engineering, and Public Policy, *Rising Above the
Gathering Storm: Energizing and Employing America for a Brighter Economic Future*, Washington, DC: The


Technology, Engineering, and Mathematics Education System*, Arlington, VA: National Science Foundation,
2007.

29. President’s Council of Advisors on Science and Technology, *Engage to Excel: Producing One Million
Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*,


Linking Engineering and Society*, 43(2), 16-23.

instructional practices: An analytic review of the literature,” *Journal of Research in Science Teaching*, 48(8),
952-984.


35. Schinske, J. and Tanner, K. (2014) “Teaching more by grading less (or differently),” *Cell Biology Education—
Life Sciences Education*, 13, 159-166.

36. Lovett, Marsha C. (2013) “Make exams worth more than the grade: Using exam wrappers to promote
metacognition,” In M. Kaplan, N. Silver, D. LaVague-Manty, & D. Meizlish (Eds.), *Using reflection and
metacognition to improve student learning: Across the disciplines, across the academy*, (pp. 18-52). Sterling,
VA: Stylus.

37. [http://www.cmu.edu/teaching/designteach/teach/examwrappers/](http://www.cmu.edu/teaching/designteach/teach/examwrappers/)