Integrating Cognitive Science with Innovative Teaching in STEM
Integrating COGNITIVE SCIENCE with INNOVATIVE TEACHING in STEM DISCIPLINES
Integrating Cognitive Science with Innovative Teaching in STEM Disciplines

Edited by
Mark A. McDaniel, Regina F. Frey, Susan M. Fitzpatrick, & Henry L. Roediger, III

Washington University Libraries
St. Louis, Missouri
2014
<table>
<thead>
<tr>
<th>Section</th>
<th>Subsections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contributors</td>
<td></td>
</tr>
<tr>
<td>Dedication</td>
<td></td>
</tr>
<tr>
<td>Preface</td>
<td></td>
</tr>
<tr>
<td>Acknowledgements</td>
<td></td>
</tr>
<tr>
<td>1. The Increasing Importance of Learning How to Learn</td>
<td></td>
</tr>
<tr>
<td>1.1. Abstract</td>
<td></td>
</tr>
<tr>
<td>1.2. Introduction</td>
<td></td>
</tr>
<tr>
<td>1.3. Components of Becoming Sophisticated as a Learner</td>
<td></td>
</tr>
<tr>
<td>1.4. Why Aren’t We Already Effective Learners?</td>
<td></td>
</tr>
<tr>
<td>1.5. What Constitute Effective Study Strategies and Do Learners Use them?</td>
<td></td>
</tr>
<tr>
<td>1.6. Why Do We Not Already Appreciate What Optimizes Learning?</td>
<td></td>
</tr>
<tr>
<td>1.7. Implications for Instructors and for Learners</td>
<td></td>
</tr>
<tr>
<td>1.9. Concluding Comment</td>
<td></td>
</tr>
<tr>
<td>1.10. References</td>
<td></td>
</tr>
<tr>
<td>2. Learning from the Test: Dos and Don’ts for Using Multiple-Choice Tests</td>
<td></td>
</tr>
<tr>
<td>2.1. Abstract</td>
<td></td>
</tr>
<tr>
<td>2.2. Introduction</td>
<td></td>
</tr>
<tr>
<td>2.3. What might be different about multiple-choice tests?</td>
<td></td>
</tr>
<tr>
<td>2.5. Evaluating Solutions</td>
<td></td>
</tr>
<tr>
<td>2.6. References</td>
<td></td>
</tr>
<tr>
<td>3. The Knowledge-Learning-Instruction (KLI) Dependency: How the Domain-Specific and Domain-General Interact in STEM Learning</td>
<td></td>
</tr>
<tr>
<td>3.1. Abstract</td>
<td></td>
</tr>
<tr>
<td>3.2. Introduction</td>
<td></td>
</tr>
</tbody>
</table>
3.3. Key Assertions of the Knowledge-Learning-Instruction (KLI) Framework
3.4. Cognitive Tutors: Illustrating Domain-Specific Knowledge Analysis and Domain-General Instructional Principles
3.5. KLI Dependency
3.6. Other KLI applications in STEM and recommendations
3.7. Summary and Conclusion
3.8. Acknowledgements
3.9. References

4. Bang for the Buck: Supporting Durable and Efficient Student Learning through Successive Relearning
4.1. Abstract
4.2. Introduction
4.3. Successive Relearning: Definition and Recommendations
4.4. Research Supporting these Recommendations
4.5. Does initial learning criterion matter?
4.6. Does relearning matter?
4.7. Recommendations for Students, Teachers, and Researchers
4.8. Acknowledgements
4.9. Appendix:
4.10. References

5. Understanding How to Teach Physics Understanding
5.1. Abstract
5.2. The Problem and the Goal
5.3. The Current Situation: Students and Schools
5.4. A (Partial) Solution: A Conceptual Problem-Solving Approach
5.5. Discussion of Steps in the Conceptual Problem-Solving Approach
5.6. An Empirical Study in High Schools
5.7. Summary of the High School Conceptual Problem-solving Study
5.8. Laboratory Empirical Studies: A Closer Examination of Processing Difficulties

5.9. Can short training increase principle-based problem categorization?

5.10. What do students learn from illustrative examples?

5.11. Thinking Back, Thinking Forward

5.12. Concluding Remarks

5.13. References

6. Recommendations for Instructors and Students: A Commentary

6.1. Recommendations for Instructors and Students

6.2. Recommendations for Instructors

6.3. Acquiring Conceptual Understanding

6.4. Learning Facts, Principles, and Categories

6.5. Closing Thoughts for Instructors

6.6. Recommendations for Students

6.7. Summary

6.8. References

7. Change from Within: The Science Education Initiative

7.1. Abstract

7.2. Introduction: Why the Push to Improve STEM Education Now?

7.3. A Vehicle for Reform: The SEI Program

7.4. The SEI Approach to Course Transformation

7.5. In Action: Transforming the Junior Level in Physics

7.6. Lessons Learned and Recommendations

7.7. Getting Started: Helpful Resources and Readings

7.8. Acknowledgements

7.9. References

8. Process Oriented Guided Inquiry Learning

8.1. Introduction

8.2. What is POGIL?
11.3. Origins
11.4. Paths to Student Understanding
11.5. A Course Ecosystem
11.6. Measures of Success
11.7. Conclusions
11.8. References

12. The Benefits of Cross-Talk: Cognitive Psychologists and Stem Educators from Multiple Disciplines Can Enrich Their Research and Enhance Stem Education through Shared Knowledge
12.1. Introduction
12.2. Pedagogical Methods (Curriculum)
12.3. Faculty development and implementation
12.4. Conclusion (Developing a unified research agenda)
12.5. References
Contributors

Robert A. Bjork
University of California, Los Angeles

Allison D. Cantor
Duke University

Stephanie V. Chasteen
University of Colorado, Boulder

Jennifer Y. Cole
Northwestern University

Terry Derting
Murray State University

Jennifer L. Docktor
University of Wisconsin, La Crosse

John Dunlosky
Kent State University

Diane Ebert-May
Michigan State University

Susan M. Fitzpatrick
James S. McDonnell Foundation

Regina F. Frey
Washington University in St. Louis

Matthew R. Glucksberg
Northwestern University

Janet Hodder
University of Oregon
Dedication

We dedicate this book to our colleague Brent Dolezalek for his steadfast enthusiasm for this project and his willingness to gain the e-publishing skills essential for its successful completion.
Preface

This volume collects the ideas and insights discussed at a novel conference, the Integrating Cognitive Science with Innovative Teaching in STEM Disciplines Conference, which was held September 27-28, 2012 at Washington University in St. Louis. With funding from the James S. McDonnell Foundation, the conference was hosted by Washington University’s Center for Integrative Research on Cognition, Learning, and Education (CIRCLE), a center established in 2011.

The conference organizers envisioned an event that would bring together researchers with expertise in discipline-based education research from diverse STEM fields with researchers in cognitive science and educational psychology who study learning, memory, and reasoning. The conference was designed to limit the number of presentations and to allow for the majority of the time to be devoted to topical roundtable or whole-group discussions.

Ultimately, 64 participants and 10 speakers gathered in St. Louis, representing the fields of biology, chemistry, engineering, physics, and psychology. Although the conference participants shared common goals and interests in improving student learning, collaboration among the groups they represented is unusual in two ways, especially in the context of STEM higher education. First, cognitive science researchers, who conduct laboratory research in order to derive general principles of effective instruction and student learning, rarely have opportunities to exchange knowledge with discipline-based STEM-education researchers, who develop, implement, and evaluate curricular innovations in the classroom. Second, the conference brought together in conversation faculty from different STEM disciplines and institutions, who rarely have occasions to discuss curricular challenges and innovations with colleagues from outside their fields. Hence, the conference encouraged and fostered discussions about STEM teaching that typically do not occur, especially in a structured and sustained manner focusing on topics in both cognitive science and STEM-education research.

The conference broadened the perspectives of both groups—the cognitive science researchers and the discipline-based STEM-education researchers, sparking a bi-directional mode of collaboration. Cognitive science researchers inform and reveal techniques that can enhance the curricular innovations developed by STEM-education researchers; discipline-based education research—grounded in specific learning and teaching issues observed in STEM classrooms—characterizes the authentic contexts within which learning occurs and develops interventions for implementation within these authentic contexts. By the end of the conference, we observed the cognitive scientists considering (in discussion with the STEM-education researchers) more authentic contexts within which to study learning and effective instruction, and we observed the STEM-education researchers considering (in discussion
with the cognitive scientists) interventions based on robust findings from the cognitive science laboratory.

By publishing the current book, we hope to foster similar interactions among a broader audience. The format of this volume provides a foundation for this bi-directional conversation. The first part of the volume contains chapters from the cognitive science presenters and their research and ideas on learning in higher education, and the second part contains chapters by the STEM discipline-based education researchers describing their innovations in pedagogy or faculty development. Both groups bring valuable knowledge to efforts to improve STEM education, and by working together, they can push these efforts in new directions that have the potential for enormous impact in the classroom and beyond. In addition, this book provides a synthesis of the current state of knowledge on the science of education, identifies outstanding issues yet to be addressed, and provides fruitful directions for continuing to create an evidence-based framework for STEM education.

The 2012 conference was the beginning of a conversation rather than the last word. Therefore, our intention is to encourage discussion and integration of ideas and to foster potential collaborative efforts between cognitive scientists and discipline experts. The decision to publish the conference proceedings as an e-book derived from a commitment to broaden the conversation by making the ideas presented at the conference widely and easily available. We hope the volume will stimulate conversation about how to merge the science of learning with the science we want students to learn.

Mark A. McDaniel, Regina F. Frey, Susan M. Fitzpatrick, & Henry L. Roediger, III
Acknowledgements

The editors thank the participants of the inaugural CIRCLE workshop and the author contributors to this e-book for their time, energy, and ample patience in seeing the project come to fruition. Thank you to Andrew Rouner, digital library director, for his work in learning the intricacies of e-publishing and creating Washington University Libraries’ first e-book. The success of the inaugural workshop was in large part due to the careful planning and organizational efforts of Mike Cahill, research scientist and project manager of CIRCLE, and Mary Stewart, office coordinator of CIRCLE. Thank you to Shannon Davis for lending her expertise in the thoughtful design of the cover. We would like to thank the Knight Center for providing a pleasant and supportive environment for the inaugural workshop. We thank the James S. McDonnell Foundation for its support of the workshop and the decision to publish the proceedings as a widely available e-book. Finally, thank you to the publisher Washington University Libraries for its support of this project.
1. The Increasing Importance of Learning How to Learn

doi:10.7936/K7QN64NR
Robert A. Bjork and Veronica X. Yan
University of California, Los Angeles

Correspondence:
Robert A. Bjork
Department of Psychology
University of California
Los Angeles CA 90095-1563
http://bjorklab.psych.ucla.edu/

Abstract

Increasingly, learning is happening outside of formal classroom instruction. As a consequence, learners need to make multiple decisions, such as what to study, when to study, and how to study, and computer-based technologies offer multiple options and opportunities for how to manage one’s own learning. Knowing how to learn effectively has never been more important, not only during the years of schooling, but across one’s lifetime—as careers change, new job skills are required, and hobbies and interests develop and change. Recent research suggests, however, that we are often prone to both mis-assessing and mis-managing our own learning. In this chapter we summarize the evidence that intuitions and standard practices are often unreliable guides to optimizing one’s learning and that there exists the potential for learners and instructors alike to make self-regulated and teacher-regulated learning more efficient and effective.

Introduction

For understandable reasons, instructors at all levels are interested in how they should teach their students, and considerable research attention has focused on how lectures and assignments can be structured to enhance students’ learning and comprehension. A topic that is arguably even more important, however, namely, what learners can do to enhance their own learning has received much less attention. Knowing how to manage and assess one’s
own learning has always been important, but learning how to learn has become increasingly crucial. In a world that is not only complex and rapidly changing, but also characterized by technologies, such as online courses and classrooms, podcasts, and the myriad learning opportunities afforded by the Internet, classroom-type learning is being pushed more and more outside of the classroom.

Knowing how to learn effectively outside the classroom becomes especially important during and after college. During the K-12 years, most of us sat, or sit, in class listening to teachers or engaging in class activities for up to seven hours a day and were, or are, assigned regular homework. In college, however, learning takes place largely outside of the classroom and is mostly in our own hands as learners. Furthermore, beyond the years of formal education, and increasingly across the lifespan, learning is almost exclusively the responsibility of the individual learner. Thus, not only is it important to be concerned with how instructors should teach, but it is also critical that we teach our students what Bjork, Dunlosky, and Kornell (2013) have recently labeled the “ultimate survival tool,” namely, how to learn.

**Components of Becoming Sophisticated as a Learner**

Becoming a metacognitively sophisticated learner is not a trivial matter, because an efficient and effective learner has to be able to both monitor and control his or her own learning effectively, which can require overcoming certain intuitions and impressions. Monitoring one’s own memory accurately involves knowing whether information or procedures have been learned to a degree that will support their later recall and transfer when needed, which could be in the context of an examination, or in the context of a job. It is important not only that additional study time be allocated when such learning goals have not yet been achieved, but also that study time not be wasted when those goals have already been achieved. While this sounds simple enough, recent research has shown that learners often are not accurate in monitoring their own learning.

That such monitoring can be faulty is illustrated by real-world experiences we have all had. We can, for example, imagine times where we have gone into an exam lacking confidence but then scored well, or, conversely, gone into an exam full of confidence and then scored poorly. Similarly, based on watching somebody else execute some to-be-learned procedure, we might experience a sense of complete understanding, but then, later, find out that we have no idea what to do next when it is required that we execute the procedure. All of these examples are instances of imperfect monitoring.

Achieving accurate monitoring is, however, only part of the battle. Even when we are able to monitor our own learning accurately, we must then understand how to control our learning activities effectively. If a learner identifies something as requiring more study, the next step is
to know how to go about gaining that knowledge. Study strategies are not made equal, however, and knowing how to go about scheduling one’s own learning, both effectively and efficiently, is critical, particularly when time is limited, as it often is in today’s busy lives.

**Why Aren’t We Already Effective Learners?**

Every one of us is a lifetime learner. We have been learning since birth and every day thereafter, through our schooling years and beyond. From that standpoint, one might expect that we would be educated by the “trials and errors of everyday living and learning” (Bjork, 2011) and become experts at managing the conditions of our own learning, but that appears, surprisingly, not to be the case. Instead, research findings have demonstrated that we are susceptible to both mis-assessing and mis-managing our own learning.

A major reason that we can be fooled as to whether we have learned and how we should learn is that one’s current performance and the subjective ease of processing are often poor indicators of long-term retention and transfer. In fact, as we describe below, there are conditions of instruction that make performance improve rapidly and thus may make it appear and feel as though we are learning, but which do not support long-term learning. Conversely, there are many conditions of instruction that appear to create difficulties for the learner and slow the rate of apparent learning but enhance long-term learning. These latter conditions of instruction may be considered “desirable difficulties” (Bjork, 1994) in that they engage learners in deeper, more elaborate, and more effortful processing. The mismatch between current performance and long-term learning has huge implications for how we as learners, as instructors, and as parents assess and guide learning.

In this chapter we focus on only a few of the different conditions of instruction that can introduce desirable difficulties for learners, namely, distributing practice, increasing contextual interference, and engaging in test-induced retrieval practice. Each of these manipulations, in our view, has important implications for the learning of science.

**What Constitute Effective Study Strategies and Do Learners Use them?**

**Distributing practice.**

True long-term learning requires repeated studying of information. The way in which repeated study opportunities are distributed, however, makes a large difference in whether information will be retained. If you are going to read a passage twice, should you study it twice consecutively or read it once and then wait before reading it a second time? Many learners may feel tempted to restudy the passage immediately after the first study in an effort to gain clarity about things that were not clear during the first reading. An immediate restudy
can also feel easier and can convey a sense of fluency (which, unfortunately, can be confused with comprehension and understanding), whereas, when a gap is introduced between study and re-study, learners can sense that they have forgotten information in the interval, making the restudy session feel less productive. Zechmeister and Shaughnessy (1980), for example, found that participants judged information to be less well learned after spaced repetitions than after massed repetitions. Figure 1 displays schematically an important pattern of results that has emerged from decades of research on the effects of spacing repeated study opportunities. As shown in the Figure, performance on tests administered after a very short delay often show a benefit of massed practice, as in the left side of Figure 1, whereas—and often contrary to learners’ judgments—spaced practice enhances long-term learning, often substantially, as revealed by superior retention on delayed tests (e.g., Estes, 1955).
Figure 1. The typical design of a spacing experiment and hypothetical results showing the typical effects of spaced and massed practice: Massing study may lead to better performance in the short-term, but spaced study yields better long-term retention.

The “spacing effect,” which refers to long-term benefits of spacing, rather than massing, repeated study sessions, is one of the most robust findings from the entire history of experimental psychology (reported as early as Ebbinghaus, 1885/1964), and has been repeatedly demonstrated across a number of time scales, from seconds (e.g., Peterson & Peterson, 1959) to months (e.g., Bahrick, Bahrick, Bahrick & Bahrick, 1993), and across a
variety of domains, from verbal learning (see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006, for a review) to advertisements (Appleton-Knapp, Bjork & Wickens, 2005) to motor skills learning (e.g., Shea, Lai, Black & Clark, 2000). It may feel more difficult to restudy information after an interval because access to information stored earlier can be lost across that interval (i.e., forgetting happens), but it is exactly this loss of access that can make spaced restudy more effective than immediate restudy. Immediately after an initial study, the to-be-learned information is highly accessible. Restudying the information at this time has a relatively small effect on overall learning, because such restudying does not engage any active retrieval processes. On the other hand, if learners are allowed to forget the information in between study and restudy, that restudy episode triggers retrieval of the first episode and thus constitutes a much more potent learning event (see the study-phase-retrieval interpretation of spacing effects; e.g., Appleton-Knapp, Bjork, & Wickens, 2005; Thios & D’Agostino, 1976).

More recently, this theory has been re-conceptualized in the framework of “reminding” (Benjamin & Ross, 2011): Restudying reminds learners of initial-study experiences, and the longer the lag, or the more dissimilar the restudy experience is from the initial study experience, the more potent the reminding as a learning event (with a caveat: If the restudy is delayed or dissimilar to the point that it does not trigger reminding, then learning does not profit). One might in fact state that forgetting leads to learning (given a restudy opportunity), rather than undoes learning. Said differently, conditions that lead to loss of access, such as increasing the length or difficulty of intervening activity (see, e.g., Bjork & Allen, 1970) or introducing a change of context (see, e.g., Smith, Glenberg, & Bjork, 1978), are also the conditions can lead to more effective relearning.

Despite the overwhelming evidence in favor of spacing, it appears that students do not appreciate the long-term benefits of spacing. One reason, perhaps, is that cramming (massed practice) right before an exam can produce good performance on the exam, as illustrated in the left side of Figure 1. The fact that such cramming will produce poor long-term retention is not, typically, something students become aware of, whereas they do experience and become aware that cramming can often, if not always, produce good exam scores. Hartwig and Dunlosky (2012), for example, found that 66 percent of surveyed students report cramming the night before exams, and Kornell and Bjork (2007) in a survey of 472 undergraduate students found that 64 percent of students surveyed said they did not restudy information if they feel like they had learned it.

Such survey findings provide some evidence that students fail to appreciate the benefits of spacing, rather than massing, repeated study sessions. Additional support comes from experiments in which learners are asked to make study decisions. In several experiments in which participants were asked whether they wanted to study individual vocabulary words
“sooner,” which meant during the first of two restudy periods, or “later,” during a second restudy period, Cohen, Yan, Halamish, & Bjork (2013) found that participants wanted to study high-value or difficult items sooner rather than later. Importantly, the authors’ experimental design made it clear to participants that the retention interval from either of the two restudy periods to the retention test on the items re-studied during that period would be the same. When the final retention interval is not controlled between the items that are studied “sooner” and “later”, learners might choose to restudy valuable or more difficult items later, not because they understand the benefits of spacing, but, instead, because they want to place the study of those items closer to the final test. Cohen et al.’s findings suggests that participants do not appreciate that longer spacing is more beneficial than is shorter spacing for long-term retention.

**Interleaving, rather than blocking, practice on different tasks.**

When one really wants to learn complex concepts or categories, rather than just memorizing simple words or facts, one common piece of advice is to sit down and really focus on it. This is, for example, how textbooks are often organized: We immerse ourselves in studying and practicing one concept so that we can make sure we understand it fully before moving onto the next. Why do we not mix our study up, alternating back and forth between different concepts? The answer seems obvious: That would be confusing!

In the face of this very compelling intuition, however, research has demonstrated that providing “contextual interference” (Battig, 1972)—that is, arranging the instruction or practice of separate to-be-learned tasks or topics in a way the maximizes, rather than minimizes, the possible interference between those tasks—can enhance long-term retention and transfer. Initially, dating back to a classic paper by Shea and Morgan (1979), the benefits of interleaving were demonstrated in the learning of motor tasks, such as interleaving, rather than blocking, practice trials on the three different types of badminton serves (for a review, see Lee, 2012). More recently, however, research has demonstrated the benefits of interleaving extend to learning categories and concepts from examples (e.g., Kornell & Bjork, 2008) and to the learning of procedures, such calculating the volumes of different solids (e.g., Rohrer & Taylor, 2007). Kornell and Bjork (2008) investigated whether the learning of individual artists’ styles was facilitated by presenting examples of a given artist’s paintings in succession or presenting those paintings interleaved among examples of paintings by the other to-be-learned artists. More specifically, participants were shown six paintings by each of twelve artists. The paintings were shown one by one, with the paintings of a given artist shown either in immediate succession (the blocked condition) or intermixed among the paintings of other artists (the interleaved condition). On the final test, participants were shown new paintings by the studied artists and had to identify which artist painted each new painting.
Thus, participants could not simply rely upon memory for any individual painting, but had to have abstracted some idea of each artist’s style.

Despite the intuitions of the researchers—and, as it turned out, the participants themselves—the interleaved condition, not the blocked condition, resulted in participants being better able to identify the artist responsible for a given new painting on the final test. Even though the participants were given feedback after each trial of the test, they overwhelmingly—when asked whether blocked or interleaved study helped them learn better—said that blocking was better. Kornell and Bjork speculated that whereas blocking exemplars of a given category might help learners notice commonalities within a category, interleaving juxtaposes exemplars of different categories, highlighting the differences between categories, which is apparently the more important consideration at the time of the final test. Subsequently, Kang and Pasher (2012) obtained evidence favoring the idea that interleaving enhances, in their term, “discriminative contrasts.” Rohrer and Taylor (2007) have also demonstrated the benefit of interleaving over blocking with mathematics learning. College students learned to compute the volumes of different geometric solids in two practice sessions (spaced one week apart). In each of these sessions, students either practiced four problems of each of four different geometric solids in a blocked or interleaved manner. Throughout these two practice sessions, those in the blocked practice condition were consistently more accurate than those in the interleaved practice condition. As illustrated in Figure 2, however, this pattern was reversed on the criterion test one week later: Those who had practiced the problems interleaved had learned the formulas significantly better than those who practiced the problems blocked. Rohrer and Taylor argued that students who had undergone interleaved practice not only learned how to solve each kind of problem, but also learned to discriminate which formula was appropriate for each problem, a claim supported by the fact that the majority of errors made by the blockers took the form of applying the incorrect formula.
Figure 2. The accuracy in solving math problems at the end of practice or on a delayed criterion test by participants who learned and practiced volume formulas blocked by type of solid or intermixed (from Rohrer and Taylor, 2007). Error bars represent one standard error of the mean.

It is easy to understand how learners might underappreciate the benefits of interleaving study: Not only does blocked practice feel easier, it can also produce better performance during the learning process. In fact, it appears, based on recent research by Tauber, Dunlosky, Rawson, Wahlheim, and Jacoby (2012), that prior to any laboratory testing, learners hold an intuition that blocking practice is optimal. Tauber et al. gave participants the task of learning 12 bird families from examples of each family but let the participants make their own study decisions as to what example of what family they would like to see next. Overwhelmingly, the participants chose to block—that is, to see another example of a given family, rather than to see an example of a different family.

Questions still remain, however, as to how universally applicable interleaving benefits are: There are fewer studies on interleaving than there are on spacing, and existing literature has focused particularly on the interleaving of related concepts. There is now some evidence, using artificial rule-based categories, that the interleaving benefit is eliminated in cases where to-be-learned categories are relatively dissimilar and thus, highlighting commonalities within a category may be relatively more beneficial for learning (Carvalho & Goldstone, 2014). Also,
it is less clear how the mechanisms of interleaving will work with unrelated materials. Should a student, for instance, interleave study of history, psychology, chemistry and literature? If the benefit of interleaving arises from enhanced discrimination between topics, it may not make sense to juxtapose completely unrelated topics, as there is unlikely anything in chemistry that will enhance the learning of history. On the other hand, given that interleaving introduces spacing—i.e., time elapses between when one first studies a topic and next returns to it—it may be that even in the absence of the useful contrasts between unrelated topics, interleaving the study of unrelated topics may still be advantageous. Recent findings reported by Birnbaum, Kornell, Bjork, and Bjork (2013) support the notion that when learning related topics, the opportunity to contrast these different topics and the spacing between examples within a given topic both contribute to the benefit of interleaving over blocking, so it may well be that interleaving unrelated as well as related topics is a good study strategy.

Finally, one practical benefit of interleaving relative to conventional spacing—given the reality of the time constraints in everyday life—is that implementing interleaving does not require a longer total period of instruction or time on task. That is, whereas spacing takes more total elapsed time than massing, interleaving takes the same amount of time as blocking. In both blocked and interleaved schedules, the same amount of information is studied across the same period of time; simply rearranging the order in which that information is studied can yield greater long-term learning.

Using tests as learning events.

By the end of their college years (and often to their chagrin), students have become experienced test-takers. From pop quizzes, to midterms, to final exams, tests in education are overwhelmingly used as high-stakes tests of assessment. The utility of tests, however, reaches far beyond simple assessment. Rather, tests are potent learning tools, and confer learning benefits that are often substantially greater than the benefits the result from re-reading. The results of an experiment by Roediger and Karpicke (2006) provide a good example of the point. Participants were asked to learn the content of a prose passage (either about “The Sun” or “Sea Otters”). Some participants studied this passage four times; others studied it three times and then took one free-recall test; and still others studied the passage only once and took three successive free-recall tests. Importantly, when participants took these free-recall tests, no feedback was provided. The results of the final criterion test, which was administered either five minutes or one week later, are displayed in Figure 3. On an immediate criterion test, those who studied more, recalled more. The pattern completely reversed, however, after one week: Participants who received more tests during the initial study session recalled more “idea units” from the passage.
Again, though, as in the case of interleaved versus blocked practice, the metacognitive judgments of the learners did not reflect the observed pattern. When asked at the end of the study phase how much they would remember in one week’s time, those who had studied more gave higher ratings than did those who were tested more. Similarly, in Kornell and Bjork’s (2007) survey of undergraduates’ study habits, a majority 68 percent of the undergraduates reported using tests to assess their own learning, whereas only 18 percent indicated that they would learn more through tests than through rereading. Even more worryingly, nine percent of the respondents said they did not use testing for any reason.

Figure 3. The results of the criterion free recall test in Roediger & Karpicke (2006), experiment 2. Mean proportion of idea units from the prose passage (and standard errors of the mean) show that while those who studied more recalled more at a short delay, those who took more tests retained more of the information after a longer delay.

The testing effect is incredibly robust and has been demonstrated across a variety of stimuli and types of tests, both in cases where feedback is and is not provided (although, in cases where accuracy is very low, the benefit of testing without feedback can be absent or replaced by a benefit of rereading; for a review, see Roediger & Karpicke, 2006). One reason testing is
thought to benefit learning is that the active act of retrieving information is a “memory modifier” (Bjork, 1975), strengthening what we have retrieved and weakening access to information that is in competition with the retrieved information (see Anderson, Bjork, & Bjork, 1994). If what is retrieved becomes strengthened, however, maybe it follows that testing is bad when people get the answers wrong? Maybe the wrong answer then becomes strengthened and persists, and/or interferes with subsequent learning of the correct answer (e.g., Roediger & Marsh, 2005). Concern over the negative effects of testing led, in fact, to a movement known as “errorless learning” (Skinner, 1958; Terrace, 1963).

Such fears are legitimate, but recent research has demonstrated that the benefits of testing go beyond simply strengthening what is retrieved. A good example is a study by Richland, Kornell, and Kao (2009). Across five experiments, participants were asked to study an essay about vision. In the experimental, pretest condition, participants were tested on concepts embedded in the passage, before reading the passage. Their posttest performance was significantly better than that of participants in an extended study condition, who did not answer test questions before reading, but instead used that time for extra study. This benefit was maintained even when analyses was restricted to only those items that the pretested group had answered incorrectly prior to study. In other words, taking time out of study to generate wrong, competing answers enhanced future learning. Richland et al. further eliminated the possibility that those who took the pretest were better able to attend to the important information by emphasizing the tested concepts through the use of bolded and italicized keywords in the extended study condition and allowing the extended study participants to read (but not answer) the same pretest questions that the experimental condition answered.

One interpretation of these findings is that activating the semantic network associated (through pretesting) with the to-be-learned topic allows for the subsequently studied information to be more elaborately encoded. Support for this view comes from research employing a simplified procedure introduced by Kornell, Hays, and Bjork (2009), one in which participants are asked to learn a list of weakly associated word pairs (such as Frog: Pond). For half of such word pairs, participants have to first try to predict the target word before being shown the intact cue-target pair; for the other pairs, participants simply study the intact cue-target pair. Despite the fact that the pairs are selected so that participants’ predictions are virtually always wrong, and despite the fact that the time available to study a given intact pair is reduced by virtue of having to first try to guess the upcoming to-be-remembered target word, later performance on a cued-recall test is reliably better when participants try to predict the to-be-learned response.

It is crucial, though, and consistent with the semantic-activation idea, that the to-be-learned response bears a semantic relationship to the cue. Huelser and Metcalfe (2012) found that
when the pairs are unrelated (e.g., Frog-Bench, where any semantic activation generated by “frog” would be unhelpful for encoding “bench”) the benefit of trying to predict the response disappears. Importantly, Huelser and Metcalfe (2012) also asked participants at the end of their experiment using related pairs whether they thought having to first predict the to-be-remembered response before seeing that response helped or hindered their remembering that response on the final test, versus simply being shown the intact pair. Even though their own performance exhibited a benefit of the prediction condition, participants judged the pure study condition to be better for learning.

The benefit of making errors has furthermore been demonstrated in the classroom. In 7th grade mathematics classrooms from three Singapore public schools, Kapur and Bielaczyc (2012) demonstrated the benefit of what they called “productive failure.” Half the classes were taught in the traditional way (“directed instruction”), cycling through seven, six, or four periods in schools A, B, and C, respectively (the variation was a result of what was afforded by each school’s structure) of classroom instruction, practice, homework and feedback. The other half of the classes spent the first six, four, and two periods, respectively, working in triads on complex problems without any instruction from the teachers. In this “productive failure” condition, very few of the groups (16%, 7% and 0%) ever reached the correct solutions. Thus, these students in fact spent the majority of the class time generating errors. Compare that to the homework performance of the directed instruction condition, which averaged 91-93 percent. In the “productive failure” classrooms, teachers stepped in to provide answers in only the last one or two periods and did so by first eliciting students’ failed methods and then drawing attention to the critical features of each failed solution before presenting the correct solution. On the final post-intervention test, students in the productive failure condition performed significantly better than students in the directed instruction condition on complex and graphically represented questions (i.e., demonstrating greater transfer of knowledge).

It is important to note, however, that in the productive failure classrooms, teachers did not simply tell students they were incorrect and then tell them the correct answer. Rather, they spent time discussing why incorrect solutions might have been chosen and why the correct solution was more appropriate. Similarly, laboratory experiments investigating the benefits of pretesting with text passages and simple word pairs indicate that errors should be related to the to-be-learned topic in order for there to be a benefit of making errors (e.g., Huelser & Metcalfe, 2012; Kornell, 2014; but see Potts & Shanks, 2014, for an example of a pretesting benefit even when errors are unrelated). One might argue, therefore, that is it not making errors per se that benefits learning, but rather, learning is benefited by the elaborate processing (for example, pretesting helps relate new knowledge to prior knowledge) that is involved when errors (and correct responses) are generated.
In addition to the more elaborated processing that may be possible, or even contingent on, having made a prediction error, it may be that there is, in fact, something special about the errors themselves. Making errors may be very useful, for example, in helping the learners understand what conceptual mistakes they may be prone to making in the future. If a student is simply told the answer, it may appear obvious when that answer or explanation is in front of them. At a later time when that information is needed, however, prior misconceptions that were not highlighted during learning may re-emerge.

**Why Do We Not Already Appreciate What Optimizes Learning?**

As mentioned earlier, a basic impediment to accurate metacognition is that current performance is an unreliable index of learning. As we have illustrated, there are many cases where current performance is not only unreliable, but also misleading. Both massing (studying and re-studying the same information without breaks) and blocking (completing study of one concept before moving onto the next) practice, for example, can lead to better performance in the short-term (and make learning feel easier) as compared to spacing (taking breaks) and interleaving practice (mixing up concepts), which can optimize long-term retention and transfer.

In addition to the fact that current performance can be misleading, it is also the case that our subjective experiences (especially that of *ease*) as a learner can be misleading. The sense of perceptual fluency we gain when reading something multiple times, for example, can be mistaken for understanding and assumed to be a reliable measure for how recallable the information will be later. Similarly, how readily information is recalled at one point in time, in the presence of certain cues, can be interpreted as a measure of how recallable that information will be at a later time, in the presence of different cues.

In addition to the fact that current performance can be misleading, it is also the case that our subjective experiences as a learner can be misleading. The sense of perceptual fluency we gain with reading something multiple times, for example, can be mistaken for understanding and a measure for how recallable the information will be later, and how readily information is recalled at one point in time, in the presence of certain cues, can be taken as a measure of how recallable that information will be at a later time, in the presence of different cues. Also, taking tests (as compared to simply studying material), particularly when one makes mistakes, can make it feel as though little learning is occurring.

There are good reasons to use indices such as perceptual fluency or retrieval fluency as guides for learning—because fluency is often an indicator of learning. If something feels easy, it may be because we really do know it! Thus, fluency can be a useful heuristic. Issues arise however, when the experience of fluency is caused by factors other than learning. Rereading
the same passage 10 times, for example, will definitely lead to feeling of perceptual fluency, but the gain in learning and understanding may be minimal. Fluency, when it arises from factors unrelated to learning, can lead to illusions of competence, which then have cascading effects on students’ monitoring, control, and ultimately, learning.

A good example of learners being influenced by perceptual fluency is a study by Rhodes and Castel (2008), who presented participants with a list of words, half of which were written in small font, and half of which were written in a large font. Even though actual recall did not differ between large and small font words, participants judged the large font words to be more memorable than the small font words. Reder and Ritter (1992) demonstrated that priming can also lead to the experience of fluency and increase participants’ “feeling of knowing.” Pre-exposing participants to certain key terms (e.g., golf, par) that subsequently appeared in a set of general-knowledge questions (e.g., What term in golf refers to a score of one under par on a particular hole?) led to a “feeling of knowing;” Participants were faster at estimating whether they could answer the question and more likely to judge that they could answer the question, even though such priming did not change their actual ability to answer a given question.

Retrieval fluency (how readily information “comes to mind”) can be another misleading indicator of learning. In a study by Benjamin, Bjork and Schwartz (1998), for example, participants were asked a series of 20 very easy trivia questions, such as “Who was the first president of the United States?” and asked to hit the Enter button as soon as the answer came to mind. After answering each question, the participants were asked to judge the likelihood that they would be able on a later free-recall test to recall having provided that answer (“George Washington” in this example). Critically, the participants were told that they would not get the questions again but would simply be given a blank sheet of paper on which they were to write down as many of the 20 answers they provided as they could. The results were that the faster an answer came to mind, the more likely participants thought they would later be able to free recall having given that answer, whereas the actual likelihood of recalling an answer was the opposite: The longer an answer took to come to mind, the more likely they were to recall that answer on the final test. That is, the more effort that was put into generating an answer, the more likely that answer was to be later recalled in the absence of the trivia question itself. Participants, however, apparently relied on a heuristic that what is more readily recalled now will be more readily recalled in the future.

In fact, in the presence of the correct answers, one might experience an illusion of competence. Most studying takes place with the textbook and notes open, and in the presence of the answers, it can be hard to appreciate how difficult it will be to retrieve the information during the test when those notes and textbooks are not present. In the classroom, too, when students are passively listening to the lecturer, with notes in front of them and on the
projector screen, they become prone to over-estimating how much they actually understand (that is, they can experience an illusion of competence). When students are asked to explain back what they learned in class, for example, they typically struggle to explain the concepts that have just been presented.

This under-appreciation of the difficulty of later being able to retrieve an answer that is present now, but will be absent and required at test, has been labeled “foresight bias” by Koriat and Bjork (2005). In one of their experiments, for example, Koriat and Bjork (2005) asked participants to learn a number of cue-target word pairs so that, later, when presented the cue word, they would be able to retrieve the target word. Some of the pairs had a very strong forward association, from cue to target, but only a weak backward association (e.g., lamp-light), whereas the opposite was true for other pairs. The participants were asked to judge, pair by pair, the likelihood that they would be able, on a later test, to recall the target word, given the cue word. Actual recall was significantly higher for forward pairs, because in the case of backwards pairs, such as light-lamp, the cue on the final test (light-?) triggers many other possible responses, such as dark or heavy. The participants, however, made much the same prediction for forward and backward pairs. That is, they suffered from a foresight bias—the inability to think ahead in time to the point where the correct response would be absent and in competition with other words associated to the cue.

In subsequent research, Koriat and Bjork (2006) explored ways in which students might be taught to avoid foresight bias. They found that giving learners the experience that backward-associated pairs are more difficult than forward-associated pairs enabled them to make more accurate judgments about those specific pairs; only the combination of experience- and theory-based (receiving an explanation of the asymmetric relationship between cue and target words) de-biasing techniques transferred learning to new word pairs.

The “curse of knowledge.”

Finally, instructors themselves are also prone to problems of fluency, not just in judging students’ learning from performance, but also in their own teaching. Instructors have to be mindful of the difference between what they know and what their students know. One might think that an expert should really make the best instructor: The expert knows the material inside out, has a strong conceptual understanding of the information, and therefore, should be able explain ideas and concepts in the most eloquent way. What is obvious to an expert, however, is not necessarily going to be obvious to a learner. The expert may not understand the misconceptions and barriers that novices face. Indeed, Piaget (1962) remarked that ‘Every beginning instructor discovers sooner or later that his first lectures were incomprehensible because he was talking to himself, so to say, mindful only of his point of view. He realizes
only gradually and with difficulty that it is not easy to place one’s self in the shoes of students who do not yet know about the subject matter of the course’ (p. 5).

One might try Newton’s (1990) study as a thought experiment to illustrate the “curse of knowledge.” First, tap out the rhythm of a well-known song (e.g. “London Bridge”) to a listener. How likely do you think that the listener would correctly identify the song? Newton (1990) found that while the tappers (who chose from a list of 25 well-known tunes) estimated that roughly half of the listeners would successfully identify the song, the reality was that only 2.5 percent of the listeners were able to do so. What was obvious to the tappers—who could hear the tune in their heads very clearly (and perhaps even hear the lyrics and full orchestration)—was not at all apparent to the listeners who heard only a series of atonal and irregular taps. As teachers, it is crucial to understand that we are the tappers and our students are the listeners.

**Implications for Instructors and for Learners**

What then are the implications of cognitive psychology research for instructors and for learners? There is a certain societal assumption, particularly in college classrooms, that an instructor transfers information to students who then store that information, in ways akin to a recording device. One of the most important principles of effective learning, however—perhaps the single most important principle—is that learners must be active participants in the learning process.

If information is to be well learned, spacing out study—for example, returning to old material every so often—is necessary. Spacing engages retrieval processes, and retrieval is a powerful memory modifier—the act of retrieval strengthens what we retrieve. By the same token, low-stakes or no-stakes tests should be frequently used as pedagogical tools, rather than simply for assessment. Testing is something that both instructors and learners can implement: Instructors can introduce low-stakes tests in their courses, while learners can engage in self-testing. There are a myriad of ways in which learners can self-test: Attempting practice tests offered by the instructor, using flashcards (to test themselves, not to simply read), using end-of-chapter questions, creating their own questions, posing questions in study groups, and so on.

Another technique that is especially beneficial for inductive learning is intermixing study or practice of related concepts (i.e., interleaving, rather than blocking study). At the end of chapter, learners might practice questions both related to that chapter as well as related to preceding chapters. Interleaving practice enhances discrimination between related concepts. Additionally, especially when we consider that in exams and in real life, students need to be
able to know when to apply which concepts (without the aid of having just read the relevant textbook chapter), the value of discrimination becomes clear.

In addition to understanding specific learning techniques they should use, learners should also be cognizant of the discrepancy between current performance and long-term retention and mindful of their interpretation of fluency. For example, learners should be aware that getting all the practice questions correct immediately after studying a chapter might not be an accurate measure of learning. They should, instead, wait for a period of time, perhaps study some other information, and then attempt those practice questions (without consulting the textbook or notes). This strategy could be a way of avoiding misattributions of fluency and illusions of competence, as well as incorporating beneficial learning strategies.

**Efficient Learning Is Not Easy Learning: Challenging Counterproductive Assumptions**

Beyond the evidence that people have faulty mental models about how they learn, and thus do not know how best to manage their own learning, there are also some societal assumptions that can be barriers to effective learning. One such assumption, which we have already mentioned, is that an instructor’s responsibility is to transmit information and a student’s responsibility is to record that information: That is, the instructor talks and students are expected to remain silent and absorb knowledge. As we have emphasized, however, one key to effective learn is that students must be active participants in their own learning.

**The role of errors.**

Another counter-productive assumption has to do with the meaning and role of errors. Errors play a critical role in our learning, yet errors and mistakes, rather than being viewed as an important component of effective learning, are often assume to reflect inadequacies of the instructor, the student, or both. The assumption that errors are to be avoided is related to another counter-productive assumption: that learning should be easy. Testing, spacing and interleaving (as opposed to rereading, massing and blocking, respectively) do not, however, make learning easy. In fact, they make learning more effortful by virtue of requiring more active processing. It is important to note here, however, that not all forms of effort are productive and not all difficulties are “desirable.” Difficulties that engage deeper processing are good; difficulties that simply allow learners to wallow in frustration are not.

**Over-attributing differences in performance to innate abilities.**
In general, in our view, differences in performance across individuals tend to be over-attributed to differences in innate abilities and under-attributed to differences in experience, effort, and practice. When we are struggling to learn it is easy to turn blame on our own aptitude and innate abilities, or turn the blame on to an instructor’s teaching style. In fact, the “styles of learning” idea—that one is a visual learner and learns best visually, for example—is very appealing: “If only teaching was presented just the right way for me, then I would learn.” There is little evidence, however, to support the existence of individual learning styles (Pashler, McDaniel, Rohrer, & Bjork, 2009), and, in our view, the styles idea can be counterproductive with respect to learning. Individual differences are, of course, important: The knowledge and assumptions we bring to new learning, as well as our motivations, aspiration and expectations of learning, do matter—and matter greatly. What is critical to appreciate, however, is what we all share: a remarkable capacity to learn.

**Concluding Comment**

The most important message is that learners should break away from the misconception that the most effective ways of learning are those that make learning easy. The experience of having to expend effort, generate errors, or work hard to achieve understanding should not be interpreted as evidence of one’s inadequacy as a learning, but, instead, as important steps towards actual long-term learning and comprehension. In short, the good news is that there exists a great potential to upgrade self-regulated learning. Far from the notion of individual “styles of learning” that pose challenges of requiring individually tailored learning methods, there are learning techniques and principles that cognitive psychology has demonstrated can benefit all of us.

**References**


Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don’t. *Bulletin of the Psychonomic Society, 15*, 41-44.
2. Learning from the Test: Dos and Don’ts for Using Multiple-Choice Tests

doi:10.7936/K7Z60KZK
Elizabeth J. Marsh and Allison D. Cantor

Duke University

Correspondence:
Elizabeth J. Marsh
Psychology & Neuroscience
Duke University
417 Chapel Drive
Durham NC 27708-0086
(919) 660-5796
dmarsh@psych.duke.edu
http://marshlab.psych.duke.edu/

Abstract

Multiple-choice tests are ubiquitous in the classroom; while typically used for assessment, our focus in this chapter is on how such tests can also serve as learning opportunities for students. We review evidence from cognitive psychology that multiple-choice tests can change what students know, helping them to remember forgotten information, boosting retention of recently learned information, and even promoting new learning. However, the educator needs to exercise care when using multiple-choice tests, because by definition multiple-choice questions pair correct answers with plausible but incorrect lures. That is, multiple-choice testing can also yield a negative testing effect, whereby prior exposure to multiple-choice questions boosts the likelihood that students will use multiple-choice lures to answer later general-knowledge questions. We evaluate a number of solutions to this problem, with the goal of maximizing the benefits and minimizing any costs of multiple-choice testing. While a number of possible solutions involve changes to test construction (e.g., changes to the plausibility and number of lures), the best solution turns out to be a simple one: educators should make sure to tell students the correct answers after they complete the multiple-choice test. We conclude with a discussion of future directions for research, with an emphasis on the need for additional studies in classroom settings.
Introduction

Scantron® answer sheets are a common sight in classrooms across America, albeit sometimes an unwelcome one. The drill is a familiar one: bring one’s number two pencil and fill in the bubbles corresponding to one’s answers to a series of multiple-choice questions. Such multiple-choice tests are commonly used to evaluate student knowledge, with educators choosing them over other types of tests because multiple-choice questions are readily available (e.g., through test banks) and easy to grade. Furthermore, both teachers and students perceive the grading of multiple-choice tests as relatively objective. For these and other reasons, multiple-choice tests are unlikely to disappear from our classrooms. Consequently, much research has been focused on how to construct multiple-choice tests to best measure what students do versus don’t know, without introducing bias, and how to create questions that go beyond fact knowledge to tap higher-level cognitive skills like applying and reasoning from knowledge (e.g., Frederiksen, 1984). While these issues are important ones, our focus in this chapter is different. Instead of focusing on multiple-choice tests as assessment devices, our focus is on multiple-choice testing as a learning tool. That is, while we are all familiar with the use of tests to determine how much students have learned from a text, lecture, or other resource, less familiar is the perspective that tests not only measure learning but also—importantly—change learning.

Known as the testing effect in the psychology literature, numerous empirical studies have demonstrated that the act of retrieving something from memory is a powerful way to promote its later retention (see Roediger & Butler, 2011, for a review). This phenomenon is best explained with an example, so we begin with a description of how testing improved performance in a web-based university course on Brain and Behavior (McDaniel, Anderson, Derbish, & Morrisette, 2007). As would be expected in such a course, students were assigned reading from the course textbook each week. The key manipulation involved whether or not students practiced the course material after (allegedly) reading it, and if so what form that practice took. Of interest was how well students learned facts like “All preganglionic axons, whether sympathetic or parasympathetic, release acetylcholine as a neurotransmitter.” For facts in the short-answer practice condition, students were quizzed with questions like “All preganglionic axons, whether sympathetic or parasympathetic, release __________ as a neurotransmitter” and had to fill in the blanks. For facts in the multiple-choice practice condition, students were quizzed with questions like “All preganglionic axons, whether sympathetic or parasympathetic, release __________ as a neurotransmitter” and selected a response from four options (e.g., “acetylcholine,” “epinephrine,” “norepinephrine,” and “adenosine”). In both quiz conditions, students received feedback about their answers. Additional facts were assigned to a re-exposure control condition involving re-reading the facts or to a no-exposure control where the facts were not practiced (neither quizzed nor re-read).
Our focus is on how these practice conditions affected performance on two multiple-choice unit tests, each of which covered three weeks’ worth of material. Figure 1 depicts performance collapsed across the two unit tests. Overall, there was no benefit on unit test performance from re-reading the facts, as compared to the no-exposure control. However, there was a benefit of prior quizzing, with students later remembering more facts previously tested on multiple-choice or short-answer quizzes, as compared to the no-exposure control. This testing effect cannot be solely attributed to re-exposure of material, since the re-reading control did not yield a benefit over the no-exposure control. Briefly, testing benefits memory because it is a good match to what students later will be asked to do: retrieve information from memory (no teacher asks students to simply read on exams). Returning to our earlier example, answering the question “All preganglionic axons, whether sympathetic or parasympathetic, release __________ as a neurotransmitter” requires the student to retrieve from memory the meanings of the terms as well as the answer. In the multiple-choice version of the question, the learner still needs to retrieve the meanings of the terms in the question prompt as well as those in the response alternatives. The student might also reason her way to the correct answer to the multiple-choice question, retrieving related knowledge that allows her to reject some of the lures. Consistent with these ideas, many studies have shown that learners benefit from answering both short-answer (Butler & Roediger, 2007; Kang, McDermott, & Roediger, 2007) and multiple-choice questions, retaining more information on later tests (e.g., Roediger & Marsh, 2005; Butler, Karpicke, & Roediger, 2007).

![Figure 1](image_url)

**Figure 1.** Unit exam performance for practiced items (which were read, tested in multiple-choice format, or tested on a short-answer test), as compared to no-exposure control items. Adapted from McDaniel et al. (2007).

A second result in Figure 1 highlights a puzzle in the literature. McDaniel et al. (2007) found that the benefit of prior quizzing was not as great for facts previously tested in multiple-choice format than in short-answer format. This result is consistent with the traditional view
of multiple-choice tests as requiring less effort from the student, which in turn would translate into less benefit to the learner (as much research has shown that difficult practice pays off in the long run; the concept of desirable difficulties; Schmidt & Bjork, 1992). Consistent with this perspective, a number of studies have shown better long-term retention following short-answer testing than multiple-choice testing (e.g., Butler & Roediger, 2007; Kang et al., 2007; McDaniel et al., 2007). However, more recent work has shown the opposite pattern, with better retention following multiple-choice testing than short-answer testing (e.g., Little, Bjork, Bjork, & Angello, 2012; McDermott, Agarwal, D’Antonio, Roediger, & McDaniel, 2014). What is safe to say is that testing benefits the learner; less clear is whether it matters as much what format the questions take, so long as they require retrieval practice.

**What might be different about multiple-choice tests?**

As already alluded to, multiple-choice tests may sometimes involve less extensive retrieval practice. The structure of multiple-choice tests implies additional differences in learning may result, as compared to using other types of tests. Consider a question like the following: *The maintenance of a constant internal salt concentration by brine shrimp is called: phagocytosis, spermatogenesis, parthenogenesis, homeostasis, or peristalsis.* The provision of response alternatives raises the possibility of several scenarios unlikely with short-answer questions.

**Possible Benefits.**

First, the provision of multiple-choice alternatives suggests some unique benefits to the learner. There is the possibility of learning new information from the test: for example, consider the possibility that the learner does not know the answer to the above question and would not be able to generate the correct answer “homeostasis” in response to the prompt but might be able to reason her way to the answer based on knowledge of the other response alternatives (e.g., she knows what spermatogenesis is and thus can eliminate that response option).

A related situation likely occurs when the learner does in fact know something and yet is unable to call it to mind. Most students are familiar with the tip-of-the-tongue state and the feeling that they know something but cannot produce it at that moment in time (this experience has often been compared to being on the brink of a sneeze). Cognitive psychologists use the term *marginal knowledge* to refer more generally to knowledge that is stored in memory but is momentarily inaccessible, regardless of whether or not the learner is in a tip-of-the-tongue state (Berger, Hall, & Bahrick, 1999). While cognitive psychologists may often focus on the acquisition of new knowledge, it remains a major goal of education to retain learned information and prevent forgetting (in other words, to prevent knowledge from becoming marginal). Returning to our focus on multiple-choice testing, answering
multiple-choice questions is a powerful way to stabilize access to marginal knowledge (Cantor, Eslick, Marsh, Bjork, & Bjork, in press). This work draws on elegant work by Berger and colleagues (1999), who showed that a simple five-second re-exposure to information was enough to stabilize access to marginal knowledge after nine days. In other words, given that a student failed to answer a question like “What is the last name of the person who proposed the Theory of Relativity?” a five-second exposure to “Einstein” was sufficient to help students retrieve that fact nine days later. How did the researchers know that they were stabilizing access to previously stored knowledge, as opposed to teaching students new facts? To tease these possibilities apart, they examined the effects of giving feedback in response to failures to retrieve true facts (like the example just given) versus false facts (e.g., a failure to answer “What was the last name of the person who proposed the theory of maladaptability?”). The key logic is that students can only have marginal knowledge for true facts, whereas any reproduction of false facts must represent new learning. After nine days, retention of the false facts was at zero; the feedback only helped retention of true facts for which marginal knowledge was possible. The key point here is that similar effects are found when students answer multiple-choice questions, even if they never receive feedback on their answers. This benefit is driven by the student’s ability to correctly select the right multiple-choice option, which serves to expose them to the correct answer. Multiple-choice testing was just as powerful as studying in fact, helping students to produce facts one week later that they had failed to produce on an initial short-answer test (Cantor et al., in press).

Possible Costs.

Thus far we have outlined two benefits that may be unique to multiple-choice tests, but multiple-choice tests may also pose unique hazards to the student. Aside from the fact that multiple-choice tests may sometimes provide less extensive retrieval practice than short-answer tests, the concern is that by definition a multiple-choice question exposes the learner to incorrect answers. In almost no other situation would an educator expose students to wrong information, but multiple-choice testing requires the student to read and consider plausible wrong information. In other words, returning to our example, what happens when the student decides that “phagocytosis” is the process by which a brine shrimp maintains a constant internal salt concentration—does he or she leave the testing situation with a false belief? This chapter focuses on evaluating this possible negative consequence of multiple-choice testing, both because we think it is a relatively unfamiliar concern to educators (but one, as discussed below, that may be avoided with a few precautions) and because the likely persistence of multiple-choice testing in our educational system means it is important to understand how to minimize any problems.
Unfortunately, several experiments have demonstrated that prior multiple-choice testing increases the likelihood that students will incorrectly answer later questions with multiple-choice lures; we refer to this effect as a negative testing effect. Consider a sample experiment, where students answered multiple-choice questions in a laboratory setting; of interest were the effects of answering (retired) SAT II questions on later performance (Marsh, Agarwal, & Roediger, 2009). On an initial multiple-choice test, Duke University undergraduates answered biology, chemistry, U. S. History, and World History questions. Each multiple-choice question was paired with the correct answer and four distractors, and included a “don’t know” option. Students were instructed to treat the experiment as if they were actually taking an SAT II test and were informed they would receive one point for correct answers, lose one-fourth of a point for incorrect answers, and lose no points for skipping questions. After completing the initial multiple-choice test, subjects completed a filler task for about five minutes (to ensure no answers were held in short-term memory) and then took a final short-answer test. In this test, questions were not paired with answer choices; answering a question required subjects to produce a response. The questions included the forty items from the initial test as well as forty new questions that served as a baseline measure of performance without prior testing.

The left side of Figure 2a shows the benefits of prior multiple-choice testing: students answered more questions correctly on the final test if the questions had appeared on the initial multiple-choice test (a positive testing effect). This effect was very robust—students answered many more questions correctly after testing. For example, students answered 26 percent more short-answer questions correctly when the questions had appeared on the initial multiple-choice test than when they had not.
Figure 2. Production of correct answers and multiple-choice lures on the short-answer final test as a function of whether the questions appeared on the initial multiple-choice test. Panel A shows the data from undergraduates and Panel B shows the data from high school juniors. Data from Marsh et al. (2009).

However, a negative testing effect was also observed, as shown in the right side of Figure 2a. Final short-answer questions were more likely to be answered with the multiple-choice lures if the questions had appeared on the initial test. Students were more likely to answer “What is the term that describes how brine shrimp maintain a constant internal salt concentration?” with one of the multiple-choice lures (e.g., phagocytosis) if the question had had appeared on the initial multiple-choice test than if it had not.

An examination of Figure 2a reveals an important point: the positive testing effect was much larger than the negative testing effect. Overall, even though undergraduates picked up some multiple-choice lures from the initial test, multiple-choice testing benefited them much more than it impaired performance.

However, Figure 2b reveals a very different conclusion. These data come from a very similar experiment, using the same materials and very similar methods, with the crucial difference that the experiment was conducted with juniors in an Illinois public high school. The high school students again showed both positive and negative testing effects: having answered the questions on the initial multiple-choice test increased both correct and multiple-choice lure answers on the final short-answer test. However, compared to the undergraduates, the high school students showed a smaller benefit from prior testing and a larger cost. Therefore,
while multiple-choice testing was overall beneficial for the undergraduates, it was much less helpful for these high school students.

To understand these group differences, we need to consider how the students performed on the initial multiple-choice test. Critically, the undergraduates answered 55 percent of the initial multiple-choice questions correctly and endorsed multiple-choice lures 22 percent of the time (they chose “I don’t know” for the remaining 23 percent of questions). In contrast, the high school students only answered 34 percent of initial questions correctly, endorsing multiple-choice lures for a whopping 56 percent of questions (choosing “I don’t know” for only 9 percent of questions). This difference is crucial because it turns out that across studies, students only reproduce multiple-choice lures that they selected when answering the multiple-choice questions (Roediger & Marsh, 2005; Butler, Marsh, Goode, & Roediger, 2006). Merely reading multiple-choice lures does not appear to have any consequences, given that students reject them. The problem is these high school students had less knowledge and also appear to have less metacognitive knowledge about what they do versus do not know (and thus take less advantage of the “don’t know” option). Returning to the ideas discussed earlier, the problem for the high school sample was that the multiple-choice test represented an undesirable difficulty for them; it is not useful to make practice so hard that students cannot succeed. It is important to note that the results shown in Figure 2 are likely more about ability to succeed on the multiple-choice test than about undergraduates vs. high school students per se. A study comparing students from an elite high school to students from a junior college might even find the reverse pattern, if the elite high school students scored better on the multiple-choice test.

The negative testing effect has been observed in numerous studies (e.g., Roediger & Marsh, 2005; Odegard & Koen, 2007). It holds even when only the highest confidence responses are observed or when students are warned against guessing on the final test (Fazio, Agarwal, Marsh, & Roediger, 2010; Roediger & Marsh, 2005; Roediger, Agarwal, Kang, & Marsh, 2010). There is some evidence that students are not simply memorizing a response, but rather are picking up a false belief (Marsh, Roediger, Bjork, & Bjork, 2007). That is, after students answer multiple-choice questions like “What biological term describes fish slowly adjusting to water temperature in a new tank?” they are more likely to use a multiple-choice lure to answer a question about a new application of the same concept (e.g., Animals that thicken their fur during winter are exhibiting what scientific phenomena? Correct Answer: acclimation). Overall, the negative testing effect appears to be real and the challenge becomes minimizing it while keeping the benefits of multiple-choice testing.

Evaluating Solutions
Because of the ubiquity of multiple-choice testing as well as the fact that (in some circumstances) it can benefit the learner, it is important to evaluate possible solutions for eliminating the negative testing effect. We have already alluded to one solution: test difficulty needs to be properly calibrated, so that it is difficult enough for students to have to retrieve information but not so difficult that students endorse a large number of multiple-choice lures (as was the case with the high school students whose data is depicted in Figure 2b). This solution may not be feasible, however—instructors may not have a sense of how difficult a test will be until they administer it. Furthermore, given that tests typically serve an assessment function, an instructor is unlikely to want to limit the range of performance.

**Passive Solutions.**

The educator will be relieved to hear that the easiest solution is a very passive one—simply wait for time to pass. Research has found that both the positive and negative testing effects are largest immediately after testing and that both effects diminish over time (Fazio et al., 2010). However, the catch is that time will also diminish the positive effects of testing, so what are really needed are solutions that keep the positives but lose the negatives of multiple-choice testing.

**Active Solutions: Changing the Test.**

We begin our evaluation of active strategies with a discussion of solutions that involve changes in test construction, some of which may be more feasible to the educator than others. One issue involves the nature of the multiple-choice lures, which of course can vary considerably—the same question could be very easy or very hard depending on the answer choices. Little and colleagues (2012) argued that multiple-choice questions can be designed to promote even greater learning than short-answer questions if questions are designed to have plausible lures that require students to consider both why each correct answer is the right choice and why each lure is incorrect. In other words, Little *et al.* argued that good multiple-choice questions require a lot of retrieval practice, as opposed to simple recognition of the correct answer. While intriguing, we believe this research is still in the early stages and is not yet ready for classroom implementation. One issue is that there is no “roadmap” for how to make these types of items; how does an educator know (other than relying on intuition) that an item requires evaluation of both the correct answer and the lures? Furthermore, this benefit depends upon students having the requisite knowledge stored in memory. In the example given earlier, if the student knows what spermatogenesis is she can reject it. However, the concern is that plausible distractors may increase the likelihood that students select multiple-choice lures as their answers, which in turn would increase the likelihood of yielding a negative testing effect. In the Little *et al.* study, performance on the initial multiple-
choice test was quite high, but one can easily imagine a situation where the lures are so plausible that performance on the multiple-choice test plummets, reducing the positive and increasing the negative testing effect (yielding performance like that depicted in Figure 2b). Clearly, however, understanding how to design effective multiple-choice questions is an important direction for future research.

Because instructors may not have the time or ability to write new multiple-choice distractors, we must consider whether simpler modifications might help. We will evaluate whether there are any benefits from changing the number of lures (as opposed to their content), including a “none of the above” option, or allowing students to skip questions. All of these options have some appeal because they would be easy for instructors to implement and potentially may reduce the likelihood that multiple-choice lures will be endorsed.

We begin by considering whether it matters how many lures are included for each multiple-choice question. That is, should a multiple-choice prompt be paired with 2, 3, 4, or 5 answer choices (or some other magic number)? On the one hand, with fewer answer choices, the student will be less likely to endorse a lure (with consequent reductions in the negative testing effect). But on the other hand, an increased number of lures might be conceptualized as a desirable difficulty, which should boost learning. Both of these positions have support in the literature: sometimes increasing the number of lures boosts memory, and sometimes it hurts memory (see Butler et al., 2006, for a review). The key determinant is how well students do on the initial multiple-choice test and hits upon a familiar theme: difficulties are desirable if and only if the learner can overcome them. That is, memory benefits when students face extra multiple-choice lures and still select the correct answer. But when an increased number of multiple-choice lures is accompanied with an increase in errors, memory suffers. Because there is no clear general answer to the question “How many alternatives is best?” we recommend sticking with what students are used to (normally 4 or 5 response options).

A similar issue arises when we evaluate whether or not to recommend including a “none-of-the-above” choice. The answer depends on whether that choice is a correct or incorrect answer (Odegard & Koen, 2007). When none-of-the-above is incorrect (meaning that the correct answer is present in the list of response options), a positive testing effect emerges. However, it is problematic when “none-of-the-above” is the correct answer (meaning that the correct answer is not present in the list of response options). In this situation, no benefits of initial testing are observed on a later short-answer test, presumably because students are less likely to engage in retrieval practice of the correct answer when it was not in the list of response options. Furthermore, students are more likely to later answer short-answer questions with multiple-choice lures, presumably because they are more likely to endorse lures when the correct answer was not present. Because it is hard to envision a situation
where “none-of-the-above” is never right, it is not clear that this is the best modification to make to a multiple-choice exam.

The final simple test modification we will evaluate involves giving students the option to skip questions. In the studies depicted in Figure 2, students were given the option to skip; students lost no points for skipping questions but lost one-fourth of a point for each multiple-choice lure endorsed (Marsh et al., 2009). In another experiment, only one-half of the undergraduates were given the option to skip questions whereas other students were required to answer every question (inevitably increasing errors). On a later test, the negative testing effect was smaller in the group given the option to skip, although this reduction, while statistically significant, was relatively small (a drop of 3 percent). In addition, it is not clear how general any benefit would be since it depends upon people’s ability to effectively determine what they do versus do not know. If people don’t know when to say “don’t know,” they will not show any benefits; for example, it is unlikely that the option to skip helped the high school students whose data are depicted in Figure 2b, since they rarely took advantage of it.

**Active Solutions: Providing Feedback.**

We conclude this section with a discussion of the best-known strategy for avoiding the negative testing effect, namely the provision of feedback. By feedback, we mean giving the student information about the correctness of their responses (this information can take several forms, as described below). Providing feedback does not require the instructor to modify his or her test and sometimes can be provided automatically (e.g., it is an option that can be selected on most Scantron® machines). Figure 3 shows the power of feedback in eliminating the negative testing effect, drawing on the relevant conditions from an experiment by Butler and Roediger (2008). Students read a series of passages, and some took a multiple-choice test on the content they had just studied. Tested subjects received no feedback, immediate feedback, or delayed feedback on their responses. Immediate feedback took the form of the correct answer and was delivered immediately after each response, whereas delayed feedback also consisted of the correct answer but was delivered at the end of the test. A week later subjects returned and took a final short-answer test on the material; the data presented here come from a condition where subjects were required to answer all questions on that final test (as is often the norm in education, where students are typically better off guessing than not responding). Students used more multiple-choice lures to answer final questions in the Tested – No Feedback condition than in the not tested control (the negative testing effect), but this effect was eliminated when feedback was presented immediately or after a delay. Figure 3 only shows the production of multiple-choice lures as answers on the final test, but it should be noted that feedback also boosted correct answers on
the final test. Similar results have been obtained with learners as young as seven years of age (Marsh, Fazio, & Goswick, 2012). Overall, feedback is a powerful way for educators to both increase the size of the positive testing effect and decrease the size of the negative testing effect.

**Figure 3.** Proportion of final short-answer questions answered with multiple-choice lures, as a function of initial learning condition. Data from Butler and Roediger (2008).

When administering feedback, there are a few key decisions an educator must make. We have already alluded to one such decision, namely the choice of when to provide the feedback. In the Butler and Roediger (2008) study just described, immediate and delayed feedback were similarly effective at reducing the negative testing effect (see also Butler et al., 2007). Delaying the feedback may, however, boost retention of correct responses (e.g., Butler & Roediger, 2008; Butler et al., 2007). Benefits of delayed feedback are likely due to the fact that it provides another spaced presentation of the to-be-learned material; spacing is known to promote superior long-term retention of material relative to massed presentation (See Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006 for a review). Thus on balance, delaying the feedback may be advised—with the important caveat that the educator must ensure that students process the feedback carefully (which may be less likely if feedback is delayed).

A second decision concerns the content of the feedback, which could range from indicating whether an answer is right or wrong (verification feedback) to providing the correct answer (answer feedback) or an explanation of the answer (elaborative feedback). Our example earlier (Figure 3) involved answer feedback; of particular relevance for present purposes is the distinction between verification and answer feedback since many Scantron® machines provide instructors with both of these options. While much research suggests that verification feedback is not particularly effective for correcting errors (e.g., Pashler, Cepeda, Wixted, & Rohrer, 2005), most of this work has examined correcting errors made on short-answer tests.
Verification feedback provides more information to the learner after a multiple-choice selection than after a response to an open-ended question. Labeling an open-ended response as incorrect does nothing to winnow down the possible alternatives, whereas labeling a multiple-choice selection as incorrect constrains the possible choices. These ideas were tested in a study where students answered multiple-choice questions such as “What is the capital of Belize?” (Correct answer: Belmopan) and received no feedback for one-third of responses, verification feedback for another third, and answer feedback for the rest of their responses (Marsh, Lozito, Umanath, Bjork, & Bjork, 2012). Of interest was students’ ability to correct errors made on the multiple-choice test on a later general-knowledge test. Both verification and answer feedback increased the retention of correct responses, as compared to no feedback. However, when it came to the correction of errors, both types of feedback yielded improvement, but answer feedback was much more beneficial than verification feedback. In short, verification feedback may be more useful for multiple-choice tests than for short-answer tests, but the greatest benefits are still obtained with answer feedback.

Conclusions and Future Directions

• Our belief is that testing is a powerful tool for the educator, which can be used to promote learning and retention in students. Multiple-choice tests may be particularly appealing given the existence of test banks and the relative ease of grading them. The point of the present chapter was to point out that multiple-choice tests should be used with care, since they do have the potential to teach students wrong information. This problem may be particularly challenging when tests are hard and/or students are underprepared. Fortunately for the educator, there is a simple solution: provide the students with the correct answers after taking the test. When thinking about implementing this tool in the classroom, it is important to keep in mind that there are many different ways to do so. That is, the instructor could give low-stakes multiple-choice quizzes, use clickers to collect in-class responses, or assign multiple-choice questions for homework using a classroom management platform such as Sakai. Multiple-choice questions do not need to be embedded in high-stakes tests or exams to yield benefits to the learner; regardless, the key is to provide learners feedback on their answers.

• It is important to note that many of the experiments discussed in this chapter were conducted in the laboratory. A burgeoning area of research involves extending such laboratory findings to the classroom. Laboratory studies are certainly valuable—they demonstrate basic cognitive phenomena and allow us to more clearly determine mechanism. Additionally, many laboratory studies use the same materials as those utilized in the classroom, increasing the similarities between the lab and the classroom. However, it is crucial to examine these issues in real classrooms, where there may be
much variability in students, longer retention intervals, and limited time to master the material. To date the most impressive attempt to generalize the research is a project called the Columbia Middle School project, where 1,400 middle school students participated over a five-year period (see Agarwal, Bain, & Chamberlain, 2012, for a review). Consider a prototypical experiment: students studied science textbook chapters; the classroom teacher gave lessons; students took multiple-choice quizzes using clickers; and retention was measured on exams that were weeks or months later. Overall, there were long-term benefits of the low-stakes multiple-choice quizzing, even when the students quizzed themselves at home (Roediger, Agarwal, McDaniel, & McDermott, 2011). The results held with history materials (Roediger et al., 2011) and with science materials (McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011), and benefits were observed as much as eight months later (McDaniel et al., 2011).

• Much of the concern about testing in the classroom is that it leads to “teaching to the test,” which may promote relatively rote learning of tested concepts and may be at the expense of material that is not tested (see Frederiksen, 1984, for a discussion of some of these concerns). It is a very fair concern that much of the laboratory work on testing has focused on relatively simple materials (that are easily tested) and retention rather than complex materials and higher-level cognition. However, most of the classroom studies do not use simple materials—they use real classroom concepts, quizzes, and exams. One Columbia Middle School teacher noted that one advantage of multiple-choice quizzing as a learning activity was that it did not require her to change her teaching style. And more importantly, she reflected that implementing retrieval practice helped her understand what makes students succeed in the classroom (Agarwal et al., 2012, p. 445).

• In closing, we would like to point out that the early evidence is promising: tests can promote transfer of learning and performance on inference questions (e.g., Butler, 2010; Little et al., 2012; Rohrer, Taylor, & Sholar, 2010) and can promote learning of complex material like functions and skills like CPR (e.g., Kang, McDaniel, & Pashler, 2011; Kromann, Bohnstedt, Jensen, & Ringsted, 2010; Larsen, Butler, & Roediger, 2008). Initial classroom evidence finds that multiple-choice tests also promote transfer of learning (McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013). An exciting direction for future research is further investigating the conditions where multiple-choice testing promotes not only retention but also a deeper understanding of material.

References


3. The Knowledge-Learning-Instruction (KLI) Dependency: How the Domain-Specific and Domain-General Interact in STEM Learning

doi:10.7936/K76Q1V59
Kenneth R. Koedinger and Elizabeth A. McLaughlin

Carnegie Mellon University

Correspondence:
Kenneth R. Koedinger
Pittsburgh Science of Learning Center
Human Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh PA 15213-3891
koedinger@cmu.edu

Abstract

Student learning from a course can be improved by first identifying the needed concepts and skills and second, selecting the instructional methods that best support the kind of learning needed for these concepts and skills. This hypothesis suggests that the choice of a domain-general instructional approach depends on the domain-specific nature of the target knowledge. To produce theoretically motivated, successful educational interventions, consideration must be given to domain-specific details as well as domain-general principles of instruction. Successful use of general principles depends on a careful domain analysis such that an instructional principle in one domain context may work completely different in another. We illustrate effective instruction that incorporates both knowledge analysis and instructional principles and offer recommendations to (re)design instruction and improve student learning.
**Introduction**

Two key goals for enhancing student learning are to identify what concepts and skills need to be acquired and what instructional methods best support learning. Toward achieving these goals, we advocate bringing together two groups of educational specialists, subject-matter experts in specific STEM domains (e.g., chemistry, physics, mathematics) and cognitive scientists striving for domain-general principles of effective instruction and student learning. In the Knowledge-Learning-Instruction (KLI) framework, Koedinger, Corbett and Perfetti (2012) hypothesized that the effectiveness of domain-general principles of instruction changes depending on domain-specific characteristics. For example, in subject-matter domains where facts are the target knowledge (e.g., definitions in anatomy), general instructional principles that support memory, such as retrieval practice (e.g., Karpicke, 2012) and spaced practice (e.g., Cepeda et al., 2009), may be more effective than competing general principles, respectively example study (e.g., Sweller & Cooper, 1985; Renkl & Atkinson, 2010) and comparison (e.g., Gick & Holyoak, 1983; Rittle-Johnson & Star, 2007). The latter instructional principles support induction and may be more effective in subject-matter domains where general skills are the target knowledge (e.g., schemas and procedures for solving STEM problems). We call this hypothesis that suggests the choice domain-general instructional approach depends on the domain-specific nature of the target knowledge the “KLI Dependency.”

We emphasize that to produce theoretically motivated, successful educational interventions, consideration must be given to domain-specific details as well as domain-general principles of instruction. Successful use of general principles depends on a careful domain analysis such that a principle in one domain context may work completely different in another. This point will be illustrated later. More broadly, our goal is to foster further collaboration between STEM discipline experts and cognitive scientists toward research that tackles both the domain-specific and domain-general sides of the very complex problem of effective STEM instruction.

This chapter starts with a summary of three assertions of the KLI Framework related to the dependency between domain-specific knowledge analysis and domain-general instructional principles. Then, we describe prior research on Cognitive Tutors that illustrates knowledge analysis and application of instructional principles. In this prior research, we recognized a synergy between knowledge analysis and instructional principles but had not yet established the dependency connecting them. Next, we illustrate the KLI Dependency in the context of contradictory results found from two contrasting theories of effective instruction. Finally, we provide an example of effective instruction that incorporates both knowledge analysis and
instructional principles and offer recommendations for applying KLI to (re)design instruction and improve student learning.

**Key Assertions of the Knowledge-Learning-Instruction (KLI) Framework**

We begin by stating three key assertions of the KLI framework. First, human learning is substantially influenced by existing knowledge. As cognitive scientists, we think this notion has been under-investigated and we will elaborate later on that. Second, optimal instructional choices are knowledge dependent, not domain-general (this restates the KLI dependency). Thus, we need cognitive theories of domains not just of the learning in general (e.g., *Anderson & Lebiere, 1998*). Finally, STEM learning mostly happens below our conscious awareness. As a result, the task of designing instruction is more challenging than simply relying on reflection on what we know and on our experiences of learning. If a large amount of knowledge acquisition in the STEM disciplines is implicit and instructional design is dependent on an accurate understanding of expert knowledge acquisition, it is reasonable to expect such intuitive design might be faulty.

The Knowledge-Learning-Instruction or KLI (*Koedinger, Corbett, et al., 2012*) framework (see Figure 1) tries to clarify some of the conceptual and terminological quagmire that can be found in a wide variety of writings on learning. The framework makes a clear distinction between instruction and learning and between knowledge and assessments. The ovals in Figure 1 represent observable actions or events. We can observe instruction, and we can observe assessment, but we cannot observe learning. Knowledge and learning events have to be inferred from those other things (i.e., the observable events of instruction and assessment). It is important to maintain these distinctions; otherwise, our conclusions about learning can be inaccurate. For example, productive teacher-student classroom discussions (e.g., where difficult questions are answered correctly) are often taken as evidence of learning—one may get the impression of observing learning directly in these discussions. However, the inference that individual students are learning and making lasting changes in their knowledge from these discussions is an uncertain one—some students may have already known answers, others may be echoing teacher cues, etc. — contrasting assessment results prior to and after this discussion would be a better way to infer whether or not learning occurred.
Figure 1. The Knowledge-Learning-Instruction (KLI) framework distinguishes observable Instructional and Assessment Events from unobservable mental Learning Events and the Knowledge Components they change. What mental process changes occur in a student’s brain can only be inferred from associated observable Instructional and Assessment Events. [*ITS = Intelligent Tutoring System].

The rectangles at the bottom of Figure 1 depict the learning and knowledge that are inferred from concrete instructional and assessment events. The two arrows connecting these rectangles indicate a bi-directional causal link between learning and knowledge, that is, the idea that learning produces knowledge (see blue arrow) and the knowledge we have influences our learning (see green arrow). The direction of the blue arrow is suggestive of a blank slate view of learning (e.g., an hypothesis of memory) while the green arrow suggests a knowledge-based view of learning where our current state of knowledge helps us acquire new knowledge.

The KLI framework (Koedinger, Corbett, et al., 2012) is grounded in prior cognitive theory and experience in elaborating such theory in educational contexts. Using such theories, Koedinger and colleagues (1997) built a high school Algebra course based on an intelligent tutoring system that employed a cognitive model of student learning. The course is technology-based but also includes text materials and teacher professional development whose design was influenced by cognitive theory. The Cognitive Tutor Algebra course has been in classroom use for some 20 years. Today, about a half million students a year use this course and interact with the computer-based Cognitive Tutor for nearly 80 minutes per week through the school year.
Cognitive Tutors have not only been demonstrated to be effective (e.g., Ritter et al., 2007; Pane et al., 2013), but now provide a research platform, both generating vast data on learning and providing a context for experiments on alternative instructional approaches. Technology-based experiments have been run not only in mathematics (e.g., Aleven & Koedinger, 2002; Salden et al., 2008), but also in chemistry (e.g., McLaren et al., 2006), physics (e.g., Hausmann & VanLehn, 2007), thermodynamics (e.g., Rose et al., 2004), engineering statics (e.g., Steif & Dollar, 2009), and statistics (e.g., Lovett, Meyer & Thille, 2008).

Figure 2 provides an example of an activity from the Algebra Cognitive Tutor. Many activities involve students solving authentic problems such as comparing two different cell phone plans. In this problem, as students work in different representational tools, they get feedback on the steps they are making. The tutoring system has a model of correct and incorrect student solution strategies and follows each individual student approach. The top right of Figure 2 shows one student’s approach to creating a table of the quantities involved in the problem. It illustrates how the tutor tracks her particular variation on this complex solution (e.g., including that she has used “t” to represent Time in minutes) and gives feedback that the expression she has entered for the cost of the current cell phone plan is incorrect (note the red box around this cell) and is missing the fixed charge of $14.95 (see contextual feedback in the small box in upper right-hand corner of Figure 2).

An example of one of the challenging questions offered is shown in the lower left of Figure 2, and the student can use the Grapher tool (lower middle) to help find the time use at which the two plans have the same cost. At any point, a student can receive personalized instruction by clicking on the Hint button (upper right in Figure 2). One form of personalization is that the instruction provided (see the example hint message suggesting the use of an equation) depends on an individual student’s existing solution and what concept or skill is needed next. A second form of personalization involves tracking each student’s progress with respect to the target components of knowledge to be acquired (see the skill bars under “...individualization” in the lower right of Figure 2). A student’s progress on these skills is used to select new activities so that each student can get more practice and instruction on just the components of knowledge he needs and not waste time with practice on what he has already mastered.
Cognitive Tutors provide interactive support for learning by doing.

Cognitive Tutors: Illustrating Domain-Specific Knowledge Analysis and Domain-General Instructional Principles

We have selected three cognitive tutor principles (see Anderson, Corbett, Koedinger & Pelletier, 1995) for eight principles to illustrate the importance of both domain-general and domain-specific cognitive analysis. First, the primary instruction of Cognitive Tutors occurs in a learning-by-doing problem-solving context where students get practice in retrieving and applying concepts and skills (much as recommended by the testing effect, Roediger & Karpicke, 2006). Second, to increase students’ learning potential, Cognitive Tutors are designed to minimize working memory or cognitive load (cf., Sweller, 1994). These first two principles are domain-general as they apply to learning activities in all domains. The third Cognitive Tutor principle, to base instruction on a “production system model,” is domain-specific. This model is a detailed expression (typically in computer code) of the knowledge components (skills and concepts) students need to know to be successful in this domain. In early cognitive tutor work the importance of both domain-general principles and discipline-specific analysis was recognized. However, at that point the notion, discussed below, that one was dependent on the other was not considered.
The Cognitive Tutor research program has been successful both in widespread dissemination and in producing positive evaluation results (Ritter et al., 2007). However, much work remains. Student achievement is still far from ideal, particularly in urban settings, and field studies still encounter the same difficulties (e.g., non-compliance, implementation issues, null results). Most importantly, existing cognitive science is insufficient to guide the vast number of design decisions that must be made to develop a complete instructional program. Most of these decisions get made through team discussions, ideally including both domain experts and cognitive scientists, but are ultimately driven by intuition and conscious self-reflection. These are ingredients for error in judgment and suboptimal instructional design.

Such reflection led to the creation of LearnLab (see learnlab.org) and the idea of using educational technologies as a kind of scientific instrumentation to run experiments, collect data, and improve the cognitive science, and, in so doing, promote a scientific and data-driven foundation for making instructional design decisions. LearnLab research has involved a wide variety of domains, including math, science, and second language learning, and a wide variety of technologies, including tutoring systems, online courses, science simulations, and educational games. Because these technologies are embedded in functioning courses that are in widespread use, we can do basic research studies at scale, that is, sampling from populations of real students in real courses. LearnLab supports researchers in running these course-based “in vivo experiments,” and more than 200 such experiments have been completed (cf., Koedinger, Corbett, et al., 2012). In vivo experiments are “principle-testing” whereby a treatment condition is different from a control condition as guided by a single instructional principle under investigation. They typically involve a single curricular unit and run for a week or so.
Figure 3. The KLI dependency predicts a relationship between the complexity of knowledge (along the bottom), the complexity of the learning processes needed (far left) and associated kinds of instruction (middle left) that best facilitate that learning. The “+” indicates a LearnLab in vivo study with a positive result and the “0” represents a null effect. Note the absence of successful studies in the upper left corner.

The KLI framework emerged from reflecting across these many studies in many domains and noticing the many different kinds of instructional principles tested and the different kinds of knowledge goals involved (see Figure 3). We found that in some studies the target knowledge is simple facts, involving a fixed (or constant) relationship between the retrieval conditions and the correct response. In other studies, the target knowledge is more complex general skills or rules that involve a changing (or variable) relationship between the retrieval conditions and the correct response that is constructed based on these conditions. As tacit rule knowledge is increasingly elaborated, verbally expressed, and associated with a rationale, we get more complex principles. Across the bottom of Figure 3, we see examples of kinds of knowledge addressed in LearnLab studies and differences in complexity (increasing from left to right). The left side in Figure 3 shows some of the instructional principles that have been experimented with and each one is associated with a different kind of learning process (memory, induction, or sense-making) ordered from simpler to more complex (bottom to top).
Figure 3 summarizes a number of in vivo experiments that have been run whereby the instructional principle on the left is used in a treatment condition, but not in the control condition. All other instruction, content, and content variables are the same. The knowledge content targeted in each experiment is indicated at the bottom. The “+” indicates a significant enhancement of robust student learning due to the treatment, the “0” indicates no difference, and an empty cell indicates that an experiment with the associated principle and knowledge goal has not been run (to our knowledge). As can be seen in Figure 3, the results of these studies seem to indicate a positive correlation between the complexity of instruction and the difficulty of knowledge components and vice versa (i.e., simpler KCs are associated with less complex instruction manipulation). A few noteworthy observations can be made from the results of the studies pictured in Figure 3. First, the lack of successful studies in the upper left corner indicates that the application of complex instructional principles to simpler knowledge components has not been beneficial or considered a worthwhile endeavor. For example, no one has tried accountable talk with Chinese vocabulary, but given the arbitrary association of why lao shi is the Chinese word for teacher, it probably does not make sense to have a long dialog about it. Second, on the opposite spectrum, note the reluctance (or at the very least, lack of time and motivation) of researchers to run studies that incorporate simpler instructional principles on more difficult concepts (as indicated by the blanks in the bottom right quadrant of Figure 3).

**KLI Dependency**

The key claim being made by the knowledge-dependency hypothesis is that instructional choices depend on which of the many possible learning processes are needed to achieve which of many possible knowledge acquisition goals. To be more succinct, you have to know what you are targeting and what learning processes are needed to make good decisions about what kind of instruction will be optimal. A meta-analysis on spacing effect by Donovan and Radosevich (1999) underscores this idea of knowledge dependency. They found that the more conceptually difficult a task, the sharper the decline in a spacing effect. The implication of their findings suggests that the distribution of practice is dependent on the type of task with more difficult tasks having different requirements than simpler tasks, hence domain-specific and domain-general. A different example focuses on the requisite cognitive processes to learn complex motor skills. In a review of that work, Wulf and Shea (2002) conclude that “situations with low processing demands benefit from practice conditions that increase the load...whereas practice conditions that result in extremely high load should benefit from conditions that reduce the load.”

**Domain-Specific Cognitive Task Analysis**
To help provide a sense of what we mean by domain-specific and its potential impact on learning, we give a brief description of a few domain-specific techniques. A Cognitive Task Analysis (CTA) is a proven set of methods for obtaining domain-specific details and improving instruction within that domain. Different CTA methods include structured interviews of experts, think alouds of experts and novices, difficulty factors assessments (see discussion of Koedinger & Nathan, 2004 below), and mining data from educational technology. Clark et al. (2007) have demonstrated how a structured interview approach to cognitive task analysis can be used to successfully modify instruction. A number of studies have compared traditional instruction with instruction that has been re-designed based on a cognitive task analysis. For example, the traditional observe-then-do technical training medical school interns receive on catheterization was redesigned after a Cognitive Task Analysis involving structured interviews of multiple experts. The interns who received instruction in the newly designed course performed significantly better on post assessments (both on paper and as field observations of surgical performance) than traditionally trained interns (Velmahos et al., 2004).

The predominant message here is that even in high stakes domains with instructors with years of experience, such as medical professionals performing surgical procedures, real leverage can occur from this kind of analysis. In Lee’s (2004) meta-analysis, a large effect
Figure 4. A difficulty factors assessment, using three representations of the same problem (a), indicates equations (not story problems as anticipated) have the lowest percent correct for high school algebra students (b). Moreover, a survey of teachers shows they are unlikely to make the correct prediction, especially the high school teachers closest to the content, a phenomenon referred to as an “expert blind spot” (c).

size of 1.7 (Cohen’s d) was reported for benefits of CTA-based instruction in comparison to existing instruction. Why the huge success with Cognitive Task Analysis in fairly well established domains? One of the reasons goes to this notion that much of what we know we don’t know we know. Indeed, it is estimated (Clark et al., 2007) that experts can report directly only about 30 percent of what they know.

One might observe that these domains seem more like training domains than academic domains and thus question whether CTA and CTA-based course design is relevant to STEM domains. Koedinger and Nathan (2004) performed a difficulty factors assessment using algebra as a subject. They were interested in understanding why story problems were considered to be notoriously difficult and how student performance on story problems compared to equation solving. Thus they created matched versions (story problems vs. word problems vs. equations) of the same core mathematical problem content (see Figure 4a for an
example). Look at the three problem types listed in Figure 4a, and reflect on which is the hardest for algebra students to solve correctly (i.e., which has the lowest percent correct). The overwhelming response to this question is that either the story or word problem is the most difficult. But, Koedinger and Nathan’s (2004) results tell a different story. Students actually perform worse on equations (42% correct) and better on word and story problems (61% and 70%, respectively; see Figure 4b). Similar results have been replicated in a number of studies, but Koedinger, Alibali, & Nathan, 2008 also found that some kinds of more difficult problems flip this relationship around, so that equations are easier than matched story problems.

Another interesting discovery came from the surveys of math educators and teachers (Nathan & Koedinger, 2000). It seems the greater one’s expertise, the more likely an incorrect prediction error. As seen in Figure 4c, less than 10 percent of the teachers who teach the course (i.e., high school teachers) rank equations as the type of problem in which students would have the most trouble. This “expert blind spot” suggests that as we develop expertise our ability to recognize what is difficult for the novice learner can be clouded and can lead to erroneous instructional decisions (e.g., more practice with word problems and less with equations). The point here is that Cognitive Task Analysis methods, such as structured interviews or difficulty factors assessments, help to combat an expert blind spot. With them, course developers can identify elements of tacit expertise that are not well addressed in current instruction.

For example, based on the algebra analysis above, Koedinger and Anderson (1998) developed an instructional strategy for helping students learn the difficult task of translating story problems to algebraic notation. This “inductive support” strategy has students first using arithmetic to solve a few specific instances of a general algebra story problem (e.g., A company charges $42 per hour plus $35 for the service call. How much would you pay for a 3 hour service call?) before being directed to formulate an (analogous) algebraic expression (e.g., write an expression for the number of dollars you must pay). Their experiment demonstrated that students learned this strategy better than those in a control condition who were instructed with a typical textbook approach where the specific instance questions came after the request for an expression.

Other researchers have applied Cognitive Task Analysis methods to improve instruction in STEM domains. For example, Feldon et al. (2010) redesigned a biology lab course (i.e., used CTA-based instructional changes) to better reflect the process of scientific reasoning by scaffolding the steps, those branching points and relevant cues used by experts when making decisions about conducting biology research. The results of a double-blind experiment showed that students using CTA-based instruction wrote significantly better discussion sections than students taught with traditional best practice methods, suggesting the
The redesigned course fostered a deeper understanding of the connection between prediction and data.

**Domain-General Instructional Principles**

The scientific literature offers many principles (often conflicting) on how to advance robust student learning (e.g., Clark & Mayer’s, 2003 book *e-Learning and the Science of Instruction*; the IES practice guide by Pashler *et al.*, *Organizing Instruction and Study to Improve Student Learning*, 2007). The more generally publicized debate about what is the best form of instruction seems to imply that there are only two choices: traditional vs. informal. Cognitive scientists may prefer the simplicity of this dichotomy, but in reality there are a myriad of possibilities for delivering effective instruction. We mention just a few to provide a sense of the vast number of instructional variations available: massed practice vs. distributed practice (e.g., Cepeda, 2009), study vs. test (e.g., Roediger & Karpicke, 2006), and examples vs. problem solving (e.g., Sweller & Cooper, 1985; Renkl & Atkinson, 2010). Each strategy can then be combined with others (e.g., distribution of practice with worked examples and distribution of practice with problem solving). Should the examples be concrete or abstract? Should the feedback be immediate, delayed, or mixed? And so forth. Recently, we did a coalescing of existing instructional principles and after accounting for redundancies, we came up with a list of 30 (Koedinger, Booth, & Klahr, 2013). Recognizing that what is ideal for an early learner is often different for a later learner, and basing our estimate on 15 principles instead of 30, the combinatorics indicate there are over 205 trillion instructional options.

Another example of a Cognitive Task Analysis used in a STEM domain involves the use of an expert/novice think-aloud study to derive improvements to an introductory college chemistry course. This example also illustrates how the effect of a domain-general instructional principle is dependent on domain-specific analysis. Davenport *et al.* (2007) were surprised to find that the addition of molecular level diagrams to an online text on equilibrium chemistry did not enhance student learning in contrast to ample empirical evidence supporting the multimedia principle (i.e., that words plus pictures favor learning more than either words or graphics alone; Mayer, 2001). In a subsequent Cognitive Task Analysis, Davenport *et al.* (2008a) established from a think-aloud analysis that some of the steps experts employ when reasoning about equilibrium are not well-addressed in existing instruction. In particular, the Cognitive Task Analysis results led to instructional redesign to focus on the concept of progress of reaction. Davenport *et al.* (2008b) returned to test the multimedia principle, but this time using diagrams that emphasize the progress of reaction. The students who received instruction with diagrams (i.e., CTA based) had better results on open-ended transfer items than those who received instruction only from text. This instructional format led to statistically better learning for low performing students, thus supporting the
domain-general multimedia principle, but only after the influence of domain-specific Cognitive Task Analysis.

**KLI Dependency Example**

Both the domain-specific knowledge analysis and domain-general instructional principles are important to effective course development. But, can they be performed/applied independently, as we did in the original Cognitive Tutor project, or should the result of one influence the other?

Consider two contrasting theories of effective instruction, Sweller’s Cognitive Load Theory (1994) and Bjork’s Desirable Difficulties (Bjork & Linn, 2006). Cognitive Load Theory recommends reducing extraneous and meaningless influences that could impede learning while Bjork’s Desirable Difficulties intentionally inserts complexity during learning with the aim of improving student learning. The direct instruction used in worked examples and multimedia principles is an illustration of Cognitive Load Theory while features such as spacing, delayed feedback, and testing effects characterize desirable difficulties. Although these two recommendations are similar in that both want to find the appropriate balance of instructional difficulty to produce the best outcome of student learning, they are essentially at odds with each other. We explore this difference with an example (see Figure 5) that shows how the testing effect (study vs. test) and worked example (example vs. problem solve) paradigms are similar yet have contradictory results (see Figure 6). Then, we will use KLI to help explain this discrepancy.
Instructor assistance and student load (see upper portion of figure) are two instructional properties used to demonstrate the similarity between the testing effect and worked examples paradigms. In both the Study and Example trials, assistance is given (high assistance) and students study the instruction (low load) while in the Test and Problem trials, responses are elicited (low assistance and students must do (solve) the problem (high load) (see bottom of figure).

Two complementary characteristics of instructional methods are (1) how much assistance is offered and (2) how much difficulty or load the instruction demands of students. For instance, when an instructor gives help, we say that is high assistance, whereas if he elicits responses it would be considered low assistance. On the student side, to study is lower difficulty or load, whereas to do an activity is higher difficulty or load. These characteristics and how they relate to each other can be seen in the upper portion of Figure 5 while the bottom part of the figure provides a specific example that shows how the testing effect and worked example paradigms are similar with respect to these two dimensions. In the case of the testing effect, a study trial gives the Chinese word and its translation, but in a test trial only the Chinese word is given and the learner needs to provide (or do) the English translation. An analogous situation occurs in the worked example paradigm where the example trial gives a problem with its solution steps worked out, but in the problem trial only
the problem is given and the learner needs to solve (or do) the problem. In other words, the study trial from the testing effect and the example trial from worked examples provide high assistance and require less load, whereas the test trial from the testing effect and the problem trial from worked examples provide less assistance and require greater load (are more difficult).

**Figure 6.** Results of two studies by different researchers (Testing effect by Roediger & Karpicke, 2006; Worked examples by Sweller & Cooper, 1985) showing a discrepancy in whether offering high instructional assistance with a low difficulty load on students has a better learning outcome than low instructional assistance with a high difficulty load.

Roediger and Karpicke (2006) ran an experiment to see the effect of testing vs. studying (see “Testing effect” row in Figure 6) on student learning. In the Test condition, students were given one study trial (analogous to an example) followed by three test trials (analogous to a problem). In the Study condition, students were given three study trials followed by one test trial. On a delayed post-test, students in the Test condition performed significantly better (64%) than students in the Study condition (57%). The outcome of an analogous study of worked examples (see “Worked examples” row of Figure 6) was quite different (Sweller and Cooper, 1985). In the Worked Example condition, students were given alternating examples (analogous to study trials) and problems (analogous to test trials). In the Problem Solving condition, students were given the same scenarios but all as problems (analogous to test trials), with correct solution feedback after a student attempt. On a test of learning transfer, students in the Worked Example condition performed significantly better (88%) than students in the Problem Solving condition (analogous to test trials, 42%). On the surface these contradictory results may seem baffling; however, there are some subtleties about the differences between testing effects and worked examples. Studies of the testing effect have focused on facts (knowledge components with a constant-to-constant mapping), while worked example studies have focused on skills or schemas (knowledge components with a variable-to-variable mapping). Therein is one key example of the KLI Dependency, that is, the knowledge goals are different. As can be seen in Figure 7, although the goal for the testing effect and worked examples is the same (best possible instruction to improve student learning), these paradigms differ on all three of the KLI elements. They differ in the kind of knowledge that is targeted (facts vs. skills) and the kind of learning that is required (facts
require memory, skills require induction). To achieve these different learning functions, then, different instructional choices appear optimal (eliciting via tests best enhances fact memory, more example study best enhances skill induction).

<table>
<thead>
<tr>
<th>Knowledge</th>
<th>Learning</th>
<th>Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Testing effect</strong></td>
<td>Constant-constant facts</td>
<td>Memory</td>
</tr>
<tr>
<td><strong>Worked examples</strong></td>
<td>Variable-variable skills</td>
<td>Induction</td>
</tr>
</tbody>
</table>

**Figure 7.** KLI framework can help to explain the contradictory results found in the testing effect and worked examples studies (see Figure 6) and other such discrepancies by recognizing that different knowledge goals require different learning processes leading to different optimal instruction.

**Other KLI applications in STEM and recommendations**

We have begun to see some powerful examples of using data and cognitive science to demonstrably improve instruction (e.g., Koedinger & McLaughlin, 2010; Koedinger et al., 2013; Lovett et al., 2008). A case in point is the Open Learning Initiative at Carnegie Mellon University where the application of these types of analyses has resulted in instructional modifications to a number of college level courses (e.g., chemistry, physics, engineering, statistics). Such scientifically-based open online learning environments are improving the quality, efficiency, and productivity in higher education. Educational technology is opening a new pathway by providing data to enhance and automate cognitive task analysis, which has already been shown to produce better student learning outcomes.

Indeed, cognitive science principles are being combined with technology in ways that are making instruction more relevant, increasing learning gains, and reducing time to mastery. For example, Lovett et al. (2008) provide evidence of such improvements: In a typical semester of an introductory statistics course, they found CMU students spend about 100 hours and demonstrate about a three percent gain on a statistics concept assessment. In contrast, using the OLI statistics course designed using data and cognitive science, students spent about 50 hours—that is, half the time of the traditional course—while achieving an 18 percent gain on the same statistics concept assessment.
The goal of all educators is to advance learning and improve student outcomes. One means for improving instruction is to employ both the domain knowledge analysis methods (e.g., Difficulty Factors Assessment) and the general instructional principles illustrated in this chapter (e.g., spacing, testing, worked examples). It is important to lay out specific knowledge goals that a STEM course should address and then link them to appropriate instructional activities. However doing this specification and linkage is not enough. We recommend collecting data on student performance on those instructional activities, through cognitive task analysis techniques. Such techniques can be implemented in a classroom by creating and administering quizzes that make contrasts between tasks as illustrated in the algebra example above (cf., Heckler, 2010). Such data will often reveal needed changes to the knowledge goal specifications and will have implications for improved instruction as illustrated in the algebra, biology, and chemistry examples above. Such data will often contradict intuition as was illustrated above in the contradiction of domain expert predictions (algebra stories are harder than equations) with data (equations are actually harder).

Not only can data be collected through quizzes, student interviews, or think alouds, but educational technology use provides a new and powerful source for data to inform Cognitive Task Analysis. We have recently developed educational data mining methods for using data from educational technologies to make domain-specific discoveries about student learning (cf., Koedinger, McLaughlin & Stamper, 2012). As in the examples above, these data-driven discoveries can be used to recommend instructional improvements. Koedinger et al. (2013) demonstrated experimentally that such improvements can enhance student learning. More specifically, they created new instructional tasks for a Cognitive Tutor unit on Geometry area after discovering (from mining student data) requisite planning skills when finding the area of composite shapes (e.g., the remaining area when one shape is inscribed in another). Hence, the redesign involved more time on new tasks that isolate problem-decomposition planning. Results indicate treatment students spent more time on these new CTA-based problems, yet they achieved mastery faster than control students and performed better on posttest problems requiring planning.

### Summary and Conclusion

The Knowledge-Learning-Instruction (KLI) framework was developed from three fundamental assertions about human learning and derives its hypotheses from these assumptions. KLI asserts human learning is substantially influenced by existing knowledge. To extract the mostly implicit learning that occurs in STEM disciplines and to attain optimal instruction, research needs to focus on knowledge-dependent features, not just domain-general instructional strategies. That is, we need to develop and refine cognitive theories of domains. To do so well requires deep collaboration between cognitive scientists and STEM
researchers. Such collaborations are a great opportunity to both advance our understanding of how people learn and to improve education.

**Acknowledgements**

This work was supported by LearnLab and the Pittsburgh Science of Learning Center by a National Science Foundation award (SBE-0836012).

**References**


4. Bang for the Buck: Supporting Durable and Efficient Student Learning through Successive Relearning

doi:10.7936/K7F769GZ
Katherine A. Rawson and John Dunlosky

Kent State University

Correspondence:
Katherine A. Rawson
Department of Psychology
Kent State University
Kent OH 44242
krawson1@kent.edu
http://www.kent.edu/CAS/psychology/people/rawsonLab/

Abstract

As students progress from primary and secondary school to college, they are increasingly expected to learn foundational information (facts, terminology, formulae, concepts, etc.) on their own outside of class. Doing so effectively is no small feat, considering the amount of material students are expected to learn within and across classes, the limited amount of time students have to spend, and that the goal is to learn information well enough to retain it across time. Unfortunately, students are often not effective at regulating their own learning, and thus educators can further support student learning by teaching students how to effectively regulate their own learning outside of the classroom. Herein lies an important challenge for both educators and researchers: What strategies should students use to get the biggest bang for their buck (i.e., the most learning out of limited time)? This chapter describes and prescribes successive relearning as a highly potent strategy for efficiently achieving durable learning.

Introduction

In principle, learning is easy. Here’s a trivial thought experiment to illustrate this point: Suppose the only thing a student needed to learn was that “maison” is French for “house.”
All she would need to do to retain this information is simply to repeat it over and over in her mind (“maison means house, maison means house...”). What makes this “in principle” illustration trivial—even ridiculous—is the one fact that makes learning difficult in practice: Students never have only one thing to learn. In addition to “maison-house,” our student is also likely to have hundreds (if not thousands) of other vocabulary words and grammatical rules she needs to learn for her French class, a hundred or so key concepts to learn for her Introductory Psychology class, a hefty list of key term definitions for Cell Biology, and so on. Simultaneously keeping all this information in mind is obviously impossible, and thus it must be learned effectively enough that it can later be accessed from long-term memory.

One might argue that at least some information need not be “learned” in this manner (i.e., learned in a way that affords reliable access from long-term memory) and that students can refer back to textbooks or look up information online as needed. However, as the amount of information to be learned increases, this strategy quickly becomes impractical and in many situations would arguably interfere with higher-level comprehension and critical thinking. One can quickly come to appreciate this point by attempting to read a book in an unfamiliar foreign language in which most words have to be looked up in a dictionary or textbook—this process is laborious and likely to yield a cursory understanding of the content at best. At some level, this notion applies in other domains and disciplines as well. For example, imagine reading a journal article in a domain outside one’s own area of expertise (e.g., a psychologist reading an article in a physics journal)—a reader lacking the requisite background knowledge is unlikely to come away with much understanding of the content, even if the reader could look up the meaning of key terms and concepts in a textbook or on the web. Likewise, a student who has failed to learn the foundational terminology and concepts within a domain will likely struggle to engage in higher-order comprehension, integration, and application of advanced concepts within a domain. Thus, much of the foundational information that students encounter in their coursework does need to be relatively well learned (i.e., readily accessible from long-term memory) to be useful.

As students progress from primary school to secondary school to college, more and more of the acquisition of foundational knowledge (facts, terminology, formulae, concepts, etc.) is offloaded to students to learn on their own outside of class, for several reasons. First, both the number of topics and the informational density within topics tends to increase with grade level, without a concomitant increase in the amount of time students spend in class. Thus, the amount of class time that can be spent to support learning of any given unit of foundational information necessarily decreases. Second, many teachers are increasingly investing time in class on active processing activities (e.g., flipped classrooms) that will promote critical thinking, problem solving, and comprehension and application of core concepts. (Note that achieving these ends will often depend on students having sufficient facility with
foundational facts, definitions, and concepts, as outlined above.) Third, students differ widely in learning abilities in general and in learning rates for particular items within a unit, but use of classroom time cannot easily be tailored for instruction of particular items to particular students.

Thus, students are expected to do a great deal of learning on their own, but a constraining factor is that students have a limited amount of time and energy that they can (or will) spend studying. The problem gets even trickier when we factor in that we do not want students to learn important concepts just well enough to squeak by on an exam but then quickly forget everything they have learned—among other reasons, students need to retain what they have learned as they move forward from introductory to advanced coursework if they are to achieve a cumulative knowledge base within a domain. In sum, one major reason learning is difficult in practice is that students typically have relatively little time to spend learning a lot of information that they need to retain for a long time.

Thus, if students are expected to achieve durable learning of large amounts of information in a limited amount of time on their own outside of class, how well are they equipped to achieve this goal? Unfortunately, a sizeable amount of research in cognitive and educational psychology indicates that students are often not particularly effective at regulating their own learning. Extended discussion of the many factors that limit the effectiveness of students’ self-regulated learning is beyond the scope of this chapter, but one factor of particular interest for present purposes is that students often have misconceptions about what study strategies and study schedules are effective. For example, research has shown that students often underutilize and underestimate the efficacy of good learning strategies and unfortunately overutilize and overestimate the efficacy of relatively poor strategies (for reviews, see Bjork, Dunlosky, & Kornell, 2013).

Here is the important implication: In addition to educators supporting student learning of content inside the classroom, another critical role that educators can play involves teaching students how to effectively regulate their own learning of content outside of the classroom to achieve durable and efficient learning. But what should teachers be telling students to do? Answering this question highlights what we believe to be an important challenge for both educators and researchers: What prescriptive conclusions can we offer to students about how to get the biggest bang for their buck (i.e., the most learning out of limited time), particularly when the goal is durable learning of foundational information?

Concerning study strategies and study schedules that are particularly effective for promoting durable learning, more than a century of research points to two clear winners: spacing and self-testing (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Simply put, spacing concerns the extent to which studying is spread out across time rather than completed en
masse. For example, suppose a student has three hours to spend preparing for an upcoming exam. One option would be to “cram” by spending all three hours consecutively the night before the exam. Alternatively, our student could choose to spend one hour a night on three different nights. A wealth of research has shown that when study sessions are spaced versus massed, long-term retention is substantially enhanced (for reviews, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dunlosky et al., 2013).

Note that spacing concerns when time is spent, not what that time is spent on. Concerning what students should spend their time doing, self-testing is arguably the most effective strategy known to date for producing durable learning. Hundreds of experiments have shown that when the same amount of time is spent on practice testing versus restudying, long-term retention is significantly enhanced after practice testing versus restudy (for reviews, see Dunlosky et al., 2013; Roediger & Karpicke, 2006; Roediger, Putnam, & Smith, 2011). Self-testing has the added advantage of being a strategy that students can easily employ on their own, via flashcards or similar techniques. Indeed, when used properly, flashcards support a particularly effective kind of self-testing—namely, self-testing that involves practicing retrieval of target information from long-term memory. Furthermore, testing followed by restudy is generally more effective than either testing or restudy alone, and flashcards also afford an opportunity to restudy target information after each recall attempt.

Putting testing and spacing together, self-testing (with restudy) is particularly effective for long-term retention when implemented on a spaced schedule, as opposed to completing all practice testing in close succession (e.g., Bahrick & Hall, 2005; Kornell, 2009; Sobel, Cepeda, & Kapler, 2011). Thus, we can confidently recommend that students should include spacing and self-testing in their self-regulated learning. Note, however, that simply telling students to engage in spaced self-testing is likely not sufficient. Although this general qualitative prescription includes basic information about what and when, a very important piece to the puzzle is missing—how much? This critical question arises from the conflicting demands outlined above. To revisit, students only have a limited amount of time to learn a lot of information, and thus they must be strategic about how much time they spend learning a given piece of information if they are to accomplish their learning goals for all of the to-be-learned material.

Given the sheer volume of research that has been conducted on testing effects in general, it is perhaps surprising that this literature does not offer much by way of specific prescriptions for how much self-testing students should do to achieve durable learning without wasting time. Concerning quantity of practice, the literature tends to support two basic qualitative conclusions. First, students would be wise to continue practicing until they can correctly recall each item at least once. That is, the benefits of testing for long-term retention are greater
following correct practice-test responses than incorrect practice-test responses (e.g., Kang, Pashler, Cepeda, Rohrer, Carpenter, & Mozer, 2011). Second, research also indicates that once is not enough—i.e., correctly recalling an item once and then never engaging in any additional practice will typically not be sufficient for long-term retention (e.g., Vaughn & Rawson, 2011). But how much is enough, and when is additional practice a waste of time?

In sum, the general qualitative prescriptions afforded by most of the literature on spaced self-testing still fall short of the mark of providing students with specific recommendations for how best to accomplish the goal of durable learning as efficiently as possible. Fortunately, research is emerging that does point to specific prescriptive conclusions for how to optimize spaced self-testing for both durability and efficiency of learning—that is, to increase the likelihood that time spent on spaced self-testing is time well spent. In the remainder of this chapter, we describe a particularly potent combination of spacing and self-testing, make some specific prescriptive recommendations for how students should invest their time, briefly summarize the outcomes of research that support these specific prescriptive conclusions, and then conclude with some recommendations for both teachers (concerning training students to support their effective use of the strategy) and researchers (concerning fruitful directions for further research).

**Successive Relearning: Definition and Recommendations**

A particularly potent combination of spacing and self-testing involves what Bahrick (1979) referred to as successive relearning. Successive relearning is relatively straightforward: During initial learning, engage in retrieval practice with restudy until all items have been learned to criterion (i.e., target information can be correctly recalled from memory when given a cue). Then, in one or more subsequent sessions, engage in additional retrieval practice with restudy until all items are successfully relearned (i.e., correctly recalled from memory). (Although our focus here is on spreading out practice across sessions on different days, spacing practice trials for a given item within each session is also recommended, e.g., by practicing other items in between trials for a given item.) Returning to the notion that flashcards provide an easy-to-use tool for students to engage in self-testing, consider how a student might engage in successive relearning using a stack of flashcards. For each card, students would use the information on the front of card (e.g., “maison”) as a cue to try to remember the target information on the back (e.g., “house”). After attempting retrieval, they would then flip the card over to check their retrieved response against the correct answer on the back. If their response was wrong, they would restudy the correct answer and then put the card at the back of stack to try again later. If their response was right, they would remove it from the stack for the rest of that session. Once all cards had been dropped from the stack, they would set the stack aside and stop practicing for that day. Importantly, successive
relearning would entail that students picked the stack back up on one or more subsequent (and not necessarily consecutive) days and resumed self-testing to see which items they could still correctly recall, with additional test-restudy practice as needed until they relearned the items they missed.

This particular implementation of successive relearning is straightforward enough, but students still have many specific decisions to make during successive relearning. Concerning the question of how much is enough, the two key decisions of greatest interest here are as follows: First, during the initial learning session, how many times should students get an item right before taking that card out of the stack? Is one correct response enough, or would it be time well spent to keep cards in the stack until each item has been correctly recalled more than once? Second, on how many other days should students pick the stack back up for relearning? Is relearning items in one other session enough, or is it time well spent to relearn items on more than one other day?

To cut to the chase, here are our specific prescriptive recommendations: Concerning how many times a student should correctly recall an item during the initial learning session, once will usually be enough, perhaps twice to be on the safe side, but anything more than that will largely be a waste of time. Concerning how many days the stack should be picked back up for relearning, one is clearly not enough. Up to 3-4 relearning sessions produce meaningful gains in long-term retention for minimal cost with respect to time. Thus, additional relearning sessions are indeed time well spent. Next, we summarize the outcomes from recent research that support these specific recommendations.

Research Supporting these Recommendations

Does initial learning criterion matter?

The first part of our prescription concerns initial learning, or the first time students sit down to practice with a new stack of flashcards—how many times should they correctly recall an item before dropping that item from the stack for that first day? This question is particularly important, given that the initial learning session tends to be costly with respect to time (students often require several unsuccessful self-testing trials before they are able to successfully recall a target). Is the pay-off in terms of long-term retention sufficient to warrant the additional costs associated with continuing to practice beyond the first correct recall for an item during initial learning? The answer emerging from an increasing amount of research appears to be no, as long as students engage in relearning in later sessions.

For example, in one of our initial experiments on this issue (Rawson & Dunlosky, 2011, Experiment 1), we had college undergraduates who were enrolled in an Introductory
Psychology course come to the lab to learn key concept definitions that were relevant to their course (e.g., “What is declarative memory? Memory for specific facts and events that can be stated verbally” or “What is the self-serving bias? The tendency to attribute positive outcomes to our own traits or characteristics but negative outcomes to factors beyond our control”). Students learned these concepts via a computerized flashcard program that presented the concepts one at a time with a prompt to recall the definition (“What is declarative memory?”). Students were told that they need not recall the definition verbatim and that they could use their own words, as long as they came up with the correct meaning of the definition. Students typed in their response and then were shown the correct answer for restudy. If the student’s response did not correctly capture the meaning of the definition, that concept was placed at the back of the virtual stack so they could try again later in practice. (If you are wondering how the program determined whether a student’s response correctly captured the meaning of the definition, we will say a bit about that in the concluding section of the chapter.)

The key manipulation concerned what happened after a student correctly recalled a concept’s definition. For one group of students, each concept was removed from the stack after one correct response. For another group of students, each concept was placed at the back of the virtual stack until it was correctly recalled a second time. In two other groups, concepts remained in the virtual stack until they were correctly recalled either three or four times.

All students returned to the lab two days later for a relearning session. Relearning began with one practice trial for each item. Practice trials were the same as during initial learning—i.e., each concept was presented as a cue and students were asked to type in the definition of each concept, followed by restudy. On this first practice test trial during relearning, performance was significantly lower for students who had only practiced to one correct recall during initial learning versus two, three, or four correct recalls (55% versus 65%, 73%, and 72%, respectively—see Figure 1). Importantly however, because this session was a relearning session, any concept that a student did not correctly recall was restudied and then placed at the back of the virtual stack for another try later until every student had correctly recalled every concept once that day.

Roughly six weeks later, all students returned to the lab one more time for a final test of long-term retention. Each concept was presented as a cue, and students were prompted to recall the definition for each one. The outcome was striking: With just one relearning session intervening between initial learning and the final test, the degree of initial learning no longer mattered—performance was just about as good for the group who only practiced to one correct recall during initial learning versus two, three, or four (see Figure 1). After students had engaged in relearning, initial learning criterion just didn’t seem to matter much anymore for long-term retention.
Whereas initial learning criterion didn’t matter for long-term retention after relearning, it certainly did matter for the amount of time students spent learning overall. During the initial learning session, it took students much more time to practice concepts until they were correctly recalled three or four times versus only once (see the left set of bars in Figure 2). The good news is that students who practiced to a higher criterion during initial learning recouped some of this additional time via faster relearning in the next session (see the right set of bars in Figure 2). The bad news is that they did not recoup all of the additional time they had spent during initial learning. When the amount of time spent during initial learning and during relearning is combined, students spent less time overall for each concept when initial learning involved one or two correct recalls versus three or four (3.2, 2.9, 3.9, and 4.5
minutes per concept, respectively).

**Figure 2.** Mean number of minutes spent practicing each concept during the initial learning session and during the relearning session in Experiment 1 of Rawson & Dunlosky (2011). Error bars represent standard error of the mean.

In sum, practicing to a higher criterion during initial learning yielded no bang for long-term retention but required more buck with respect to time spent. We have shown similar outcomes in several subsequent experiments but should note that the particular experiment summarized above produced the strongest effects so far. In most of our subsequent experiments, more than one relearning session was needed to wipe out the initial criterion effect on long-term retention, and in one experiment the effect was not completely wiped out. Nonetheless, all experiments point to the same conclusion—as long as students engage in
relearning, the criterion achieved during initial learning does not seem to matter much for long-term retention, whereas it does matter for time spent studying.

The important part of the conclusion above is the conditional noted in italics—namely, as long as students engage in relearning. If students are unable or unwilling to engage in relearning sessions after the initial learning session, then we would definitely recommend they practice to a higher criterion during initial learning. However, students should be strongly encouraged to engage in successive relearning, because the benefits of relearning sessions far outweigh the relatively minimal costs associated with these additional practice sessions, to which we turn next.

**Does relearning matter?**

The experiment summarized above only involved one relearning session. Is one enough, or should students spend time on additional relearning sessions? Outcomes from several experiments suggest that time spent on additional relearning is definitely time well spent.

For example, another experiment ([Rawson & Dunlosky, 2011](#), Experiment 3) involved college undergraduates enrolled in an Introductory Psychology course who came to the lab for an initial learning session involving test-restudy practice for key concept definitions. Concepts were presented via the computerized flashcard program until all concepts had been correctly recalled an average of two times. (Some students practiced until concepts were correctly recalled once whereas other students practiced until concepts were correctly recalled three times, but in the results described here we have averaged across these groups to simplify the discussion.) All students returned two days later for a relearning session, in which all concepts were again practiced until correctly recalled once. For one group of students, this session was the only relearning session they completed. Another group of students was assigned to return to the lab the following week for a second relearning session. Another group was assigned to return the following week for a second relearning session and then the week after that for a third relearning session. Other groups of students were assigned to return for a fourth and even a fifth relearning session. All students returned to the lab approximately one month after their last relearning session for a test of long-term retention. Each concept was presented as a cue, and students were prompted to recall the definition for each one. We also asked students to return to the lab again for a final long-term retention test that took place four months after their last relearning session. We should mention that students were tested on only half of the concepts on the one-month test and were tested on all of the concepts on the four-month test. This design allowed us to examine whether taking the one-month test provided a “booster” (much like a cumulative final exam would do) for the four-month test. The short answer is yes: performance on the four-month test was better for concepts that were tested versus were not tested at one month. In the results described here
we have averaged across these conditions to simplify the discussion (for more detail, see Rawson & Dunlosky, 2011).

As shown in Figure 3, the number of relearning sessions that students completed definitely mattered for long-term retention after one month and after four months. For example, as compared to performance after only one relearning session, four relearning sessions produced a 76 percent relative improvement on the one-month test and a 62 percent relative improvement on the four-month test. However, Figure 3 also shows that the benefit from increasing the number of relearning sessions tapers off—the gain in long-term retention from each next relearning session is smaller and smaller, and students showed minimal gains from completing a fifth relearning session.
Figure 3. Mean percent correctly recalled on long-term retention tests that were administered one month and four months after the last relearning session for each group in Experiment 3 of Rawson & Dunlosky (2011). Error bars represent standard error of the mean.

Given that our participants were students enrolled in an Introductory Psychology course and that our experimental materials involved key concepts from Introductory Psychology, one might wonder how much the successive relearning they completed in the context of the experiment actually enhanced their learning above what they might have learned about these concepts on their own from exposure in their class. One part of the method that we have not yet mentioned is that our materials actually included two sets of concepts. Students practiced one set in the context of the experiment, whereas the other set was not practiced but was nonetheless tested during the long-term retention tests (with counterbalanced assignment of which set was practiced). When prompted to come up with definitions for the unpracticed concepts on the four-month test, students could only produce around 11 percent of the correct definitional information for the unpracticed concepts. This relatively dismal level of knowledge could be due to not having encountered these concepts in class. However, when we asked students to report how many of the concepts they had encountered in their Introductory Psychology class, students reported having encountered about half of the unpracticed concepts in their class. Thus, lack of exposure cannot completely explain the low performance for unpracticed items. Successive relearning significantly enhanced the durability of learning beyond what students were accomplishing via “business as usual” for their class, and we have observed similar outcomes in several other experiments (including a recent demonstration of the benefits of successive relearning over business as usual in a classroom study; Rawson, Dunlosky, & Sciartelli, 2013).

What was the cost associated with achieving these impressive improvements in long-term retention? Fortunately, the answer is “not much.” Figure 4 shows the mean amount of time in minutes that students spent on each concept during initial learning (to practice until each concept was correctly recalled twice on average, as noted above) and on relearning each concept in the subsequent relearning sessions. Note that these values represent “start to finish” times that run from the time taken to type in a sentence-length response on each trial to the end of restudying the correct answer as needed. Importantly, relearning concepts a second time only required about 1.5 minutes on average per concept, and subsequent relearning sessions took even less time. The relatively minimal amount of time required to relearn concepts more than once was arguably time well spent, given the non-trivial gains produced in long-term retention.
Recommendations for Students, Teachers, and Researchers

For students.

We highly recommend that students incorporate successive relearning into their self-regulated learning, because (a) it yields durable learning, (b) relearning is relatively low-cost with respect to time, and (c) the strategy can be applied widely to many different kinds of information from many different content domains (e.g., foundational facts, terminology, formulae, concepts, and so on). Concerning more specific recommendations about how much is enough, the available research currently suggests that students will make best use of their time if they practice until items are correctly recalled once (or at most twice) during initial learning and then save the rest of their study time to spend on relearning items again in 3-4 subsequent sessions. The third and fourth relearning sessions will have significant bang for

Figure 4. Mean number of minutes spent practicing each concept during initial learning and during each relearning session in Experiment 3 of Rawson & Dunlosky (2011). Error bars represent standard error of the mean.
Concerning the potential benefit of more than 3-4 relearning sessions, the study described above suggested that the benefit of additional relearning sessions may be minimal. However, we believe it is premature to discourage students from engaging in additional relearning, because additional relearning sessions may be advantageous under some conditions. For example, further research may reveal that the benefits of additional relearning sessions are enhanced if the amount of time between later sessions is increased (cf. students engaging in additional relearning near the end of a semester in preparation for a cumulative final exam).

For teachers.

We offer three recommendations for teachers. First, we believe that spending some classroom time to teach students about what successive relearning involves and how to implement it successfully will be a good investment of time in terms of pay-off for enhanced learning. As part of this training, teachers may want to consider presenting students with one or two outcomes from research. Showing students the meaningful gains that can be made in learning from using the strategy and the minimal time involved for relearning may facilitate buy-in on the part of students. Second, in addition to teaching students about what successive relearning involves, many students will likely also need some explicit guidance and support for time management, in that successfully implementing a schedule of successive relearning will require planning in advance. For example, about two weeks in advance of an upcoming exam, the first author has recently started providing her students with a simple, one-page reminder about recommended study strategies along with a recommended schedule for studying prior to the exam (a sample is provided in Appendix).

Note that in addition to a recommended schedule for retrieval practice, the handout also includes recommendations for other study strategies that students can use to accompany successive relearning. We point this out to also make clear that we are not suggesting that successive relearning is the only strategy that students should be using during self-regulated learning, and teachers may also want to recommend other strategies that students can use to supplement successive relearning, depending on the nature of the to-be-learned material and the learning goals.

Third, for simple materials (e.g., foreign language vocabulary such as “maison – house”), students can easily judge whether they have correctly recalled a response during retrieval practice, which is important for knowing when to remove a flashcard from the stack. However, for more complex material (such as the key concept definitions we have used in much of our prior research), students often struggle to evaluate whether they have correctly
captured the meaning of a definition. In the computerized flashcard program that we use to conduct our experiments on successive relearning, the computer does not actually score the accuracy of the students’ responses. How then does the program know when a student’s response is correct? Answer: The program counts a response as correct when the student says that the response is correct. Specifically, after each recall attempt, the program shows students their response along with the correct definition and prompts them to compare their response to the answer to decide if the response is completely correct. Students enter their judgment, and trials in which they judge a response as completely correct are treated as a correct recall.

Of course, if students incorrectly judge that a response is correct when it is not, that concept will be prematurely dropped from practice, and learning will suffer. Thus, it is important for students to make their judgments as accurately as possible. The bad news is that students are not always very good at judging the accuracy of their responses. Recent survey data suggest that students do not always check their recall responses against the right answer to evaluate the accuracy of their response (Wissman, Rawson, & Pyc, 2012), and judgments tend to be very inaccurate when they do not (Rawson & Dunlosky, 2007). The good news is that checking responses against the right answer does improve judgment accuracy. Even better, we have found that students’ judgments are even more accurate if they break a complex definition down into its key idea units and then compare their response to the definition one idea at a time (Dunlosky, Hartwig, Rawson, & Lipko, 2011). More generally, the important point is that when students engage in successive relearning, they will likely also profit from some coaching on the importance of accurately judging when they are recalling correct information.

For researchers.

As noted earlier, a student attempting to implement successive relearning will have several different decisions to make. We have only considered two of those here (initial learning criterion and number of relearning sessions). Other decisions we have not considered concern issues such as how many flashcards to practice at a time (which functionally determines the spacing between trials within a session; Rawson, Dunlosky & Sciartelli, 2013), how many days between relearning sessions (which in turn may depend on the targeted retention interval; e.g., Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008), and whether different sets of flashcards should be practiced in blocks versus interleaved with one another. These and other decision points each raise a bang-for-the-buck question—how much does each of these possible approaches influence the effectiveness of successive relearning, both with respect to gain in learning and cost in time? Currently, not enough research is available to support prescriptive conclusions about these decision points. These gaps highlight important directions for future research that will support even finer-grained recommendations for how
students can capitalize on the potency of successive relearning. Nonetheless, sufficient research is available to strongly recommend that successive relearning is an important tool for students to have in their self-regulated-learning toolbox and one that they can (and should) use widely and regularly to promote efficient and durable learning.

Acknowledgements

The research reported here was supported by a James S. McDonnell Foundation 21st Century Science Initiative in Bridging Brain, Mind and Behavior Collaborative Award, and by the Institute of Education Sciences, U. S. Department of Education, through Grants #R305H050038 and #R305A080316 to Kent State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U. S. Department of Education.

Appendix:

Sample handout for students

WHAT TOOLS TO USE?

FOR MEMORY:

• Retrieval Practice (cued recall – use concepts on study guide as cues to recall definitions)
• Spaced Practice (Short sessions spread over days better than one long session night before)
• Retrieval practice + spaced practice = successive relearning
• Go until you recall each answer correctly once (or twice)
• Make sure you carefully check accuracy of your recall
• Relearn on 3-4 other days after that

FOR COMPREHENSION:

• Practice applying the concepts via online quizzes (in online supplement to textbook)
• Example generation – come up with your own real-world examples
• Compare-contrast – for sets of concepts that are related but confusable
• Explanatory questioning – ask yourself, “What is new here that I didn’t know before? How does it relate to what I already know? Why might this fact/concept be true?”

PLANNING AND TIME MANAGEMENT
You can split information into smaller sets and have a slightly more sophisticated schedule if you want, but here’s a sample schedule if we split our content into two halves:

SET A ⇒ survey methods, observational methods, bivariate correlations

SET B ⇒ multivariate correlational designs, experimental designs

<table>
<thead>
<tr>
<th>DATE</th>
<th>SET A</th>
<th>SET B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wed 10/17</td>
<td>retrieval practice</td>
<td>---</td>
</tr>
<tr>
<td>Fri 10/19</td>
<td>retrieval practice</td>
<td>---</td>
</tr>
<tr>
<td>Mon 10/22</td>
<td>retrieval practice, comprehension practice</td>
<td>retrieval practice</td>
</tr>
<tr>
<td>Wed 10/24</td>
<td>retrieval practice, comprehension practice</td>
<td>---</td>
</tr>
<tr>
<td>Fri 10/26</td>
<td>retrieval practice</td>
<td>retrieval practice</td>
</tr>
<tr>
<td>Sun 10/28</td>
<td>retrieval practice, comprehension practice</td>
<td>retrieval practice, comprehension practice</td>
</tr>
<tr>
<td>Mon 10/29</td>
<td>retrieval practice, comprehension practice</td>
<td>---</td>
</tr>
<tr>
<td>Tue 10/30</td>
<td>COURSE EXAM (will cover both Set A and Set B)</td>
<td>retrieval practice</td>
</tr>
</tbody>
</table>

This schedule is intended to illustrate how to plan your practice schedule. You can tailor it with different dates and times given your schedule constraints and availabilities. But to maximize learning, I’d suggest shooting for retrieval practice on 4-5 different days for each set, and comprehension practice on 2-3 different days for each set.

References


5. Understanding How to Teach Physics Understanding

doi:10.7936/K79G5JR7
Brian H. Ross¹
http://orcid.org/0000-0001-8936-1037,
José P. Mestre²
http://orcid.org/0000-0003-2110-1954 and
Jennifer L. Docktor³

¹Department of Psychology, University of Illinois; ²Department of Educational Psychology, Department of Physics, University of Illinois; ³Department of Physics, University of Wisconsin, La Crosse

Correspondence:
Brian H. Ross
Department of Psychology
University of Illinois
bhross@illinois.edu
http://beckman.illinois.edu/directory/person/bhross/

Abstract

Physics courses often assess understanding in terms of problem-solving performance. However, many students who do well in a physics course do not understand the underlying concepts and principles, leading to poor retention and transfer. The students’ problem solving relies on superficial aspects of the problem and chaining of possible equations. We propose an intervention to integrate the conceptual understanding with the problem solving, forcing the problem solving to be guided by the underlying principle(s). More specifically, the approach requires students to identify the principle, justify why this principle is appropriate, plan how to solve the problem, and, finally, implement the plan in terms of equations. Initial results with this intervention in high schools are encouraging: Students gain greater understanding and also perform better on the usual problem-solving tasks. We suggest some improvements that might be made to the approach and consider its generality for other STEM disciplines.
The Problem and the Goal

Learning physics is hard. Evidence for this difficulty comes not only from testimonies of generations of students, but also from the existence of a field in physics, Physics Education Research (PER), that conducts research on the learning of physics and on applying findings to improve physics teaching. However, the problem is not just that learning physics is hard, but that even many students getting good grades at good schools do not really understand the physics they supposedly learned. Yes, they can solve problems, but they don’t understand the concepts involved, nor can they explain the important aspects of why the problem is solved in this way.

In the late 1980s the Force Concept Inventory (Hestenes, Wells & Swackhamer, 1992), which tests basic physics concepts in introductory physics that all instructors expect students to learn, began circulating among physics instructors. Much to the surprise and dismay of instructors, student performance showed that conceptual understanding following well-taught introductory courses was much less than the instructors had assumed. Physics at this elementary level (mechanics) is elegant—a small number of principles and concepts can be used to generate solutions to a large variety of problems. Instead, many of the students do well by focusing on the surface aspects of the problems (e.g., inclined planes) and then finding/manipulating equations until an answer is derived.

Teachers hate this—they hope the students come away with a deep understanding and appreciation of the elegance of the field. However, it is not just instructors’ aesthetics that suffer; from the work done as early as the 1930s in Gestalt Psychology, we know that poor conceptual understanding leads to poor transfer to new problems and poor retention. Students are not learning in a way to use their knowledge in new situations, to remember enough to use it in similar situations, or to build upon it in future courses.

The goal of our research was to devise a better way of instruction that emphasized conceptual understanding of the students. We had two major guiding principles. First, the method was to be built on principles from Cognitive Science. There is always some judgment in how any idea is implemented, but we used well-tested ideas as the foundation for our method. Second, it had to be easy to adopt. We decided to avoid any major restructuring of the curriculum and focused instead on how to structure problem-solving instruction to get students to solve problems with a conceptual approach.

Although we do not have the space for a long discussion of other programs, there are a number of other physics curricula that aim to improve conceptual understanding (see Docktor & Mestre, 2011, for an NRC-commissioned white paper containing a review of these
programs). Unlike many other programs, we did not redesign the course but rather devised an intervention that could be easily adopted in a typical course.

**The Current Situation: Students and Schools**

Although there is general agreement that students in physics classes often solve problems without a deep understanding of the concepts, it is important to be explicit about how they do solve problems. We can identify two important parts to their non-conceptual approach. First, when faced with a problem, they do not look for the underlying principles, but instead rely on the surface objects, quantities, and terms mentioned in the problem statement to generate potentially-relevant equations (e.g., Chi, Feltovich & Glaser, 1981, Docktor, Mestre, & Ross, 2012). Students will often use surface terms (e.g., inclined plane problems) to talk and think about problems and use physics terms (e.g., kinetic energy; acceleration) to think of equations. Second, they often use equation-chaining strategies, using variable terms they have computed and ones they desire to help generate the next equation.

It is important to understand that this surface- and equation-oriented problem-solving approach often allows students to solve the problem, though it does not lead to learning much for future problem solving (e.g., Sweller, 1988). In many domains, surface properties are correlated with deep properties, so relying on surface terms provides a good chance of generating appropriate equations even if concepts are not understood (Blessing & Ross, 1996). For example, in algebra if a boat is moving on a river, some version of rate-time-distance if often needed. In physics, inclined plane problems may often require Newton’s Second Law. In addition, for homework problems, the section in the textbook often helps to provide clues as to the relevant equations, narrowing the search considerably. As physics instructors already know, smart students can do remarkably well in problem solving without much understanding of the underlying principles (Hake, 1998).

Given the importance of understanding the underlying principles, why are conceptual approaches not eagerly embraced by students and instructors? There are probably many reasons, but we mention four here. First, despite a general acknowledgement of the importance of deep analysis, assessments in many physics courses are primarily problem solving. Homework typically consists almost exclusively of problem solving, as do many exams. Even problem-solving instruction often inadvertently downplays the importance of concepts by including a short conceptual analysis followed by a long application showcasing the equations and mathematical procedures used to solve the problem. The result is that good students can learn to solve problems without conceptual understanding. Second, understanding is hard work and may lead to some resistance from students. Earlier work using strategy writing to force conceptual analysis (Leonard, Dufresne & Mestre, 1996) led to much resistance from the students, and important conceptual tasks, such as categorization,
are often difficult for students (Hardiman, Dufresne & Mestre, 1989). Third, instructors are often under time pressure to cover lots of topics, and conceptual approaches often require extra time and discussion. Finally, we tend to teach the way we were taught, and to learn a new approach that we are unfamiliar with takes considerable time and energy.

In the main work to be reported here, we focused on high schools, where students are first learning physics. These classes move at a “slowish” pace given the difficulty of the material for high school students, and teachers are often under pressure to cover more material. In addition, not only is the variability among students great, but so is the variability among teachers. An amazing two-thirds of high school physics teachers in the US are not trained in physics (White & Tesfaye, 2010), so it is unclear how deep a conceptual understanding the teachers all have. We will come back to this point later, but any useful approach needs to be robust against varying teacher and student abilities.

Before we go through an example that illustrates our approach, it may be helpful to give a high-level characterization of conceptual problem solving. Although students often rely on surface terms and are very equation-oriented, experts instead use the presented problem to trigger possible principles, check to ensure that the appropriate conditions for the use of such principles holds, and then use the principle with the problem to generate a solution plan with relevant equations. This conceptually based problem-solving approach helps to decompose the problem into parts that are visible to the student and to have the conceptual analysis integrated with the problem solving, not as some separate task.

**A (Partial) Solution: A Conceptual Problem-Solving Approach**

We illustrate the approach with the problem of a skateboarder moving up a curved ramp where the solver needs to find the height of the curved ramp. Figure 1 is provided as a model for the teacher to use in teaching the approach to students, but we surely do not expect students to write such complete solutions. There are four main parts that capture our conceptual problem-solving intervention. First, the solver is asked to **identify** the underlying principle (usually students simply write “conservation of energy,” or “energy”). Second, the solver is asked to **justify** the choice of this principle; in this case the solver needs to state how the problem context allows the application of conservation of mechanical energy. Third, the solver is asked to **plan** how the problem will be solved (without listing any equations). For example, in addition to drawing a diagram and labeling variables, the plan here includes writing an equation for conservation of mechanical energy, expanding each component of the initial and final energy into their potential and kinetic components, writing equations for each of these components, and finally filling in numbers from the problem and solving for the height of the ramp. Fourth, the solver **implements** the plan in a two-column solution format: each step of the plan is put in the left-hand column, and then it generates parts of the solution
in the right-hand column, as can be seen in the figure. (Also see Mestre, Docktor, Strand, & Ross, 2011, for a broader perspective on this approach.)

**Figure 1. Sample problem and conceptual problem-solving approach**

A skateboarder enters a curved ramp moving horizontally with a speed of 6.5 m/s, and leaves the ramp moving vertically with a speed of 4.1 m/s. The skateboarder and the skateboard have a combined mass of 55 kg. Find the height of the ramp, assuming no energy loss to frictional forces.

**Principle:**

Conservation of energy: the total mechanical energy (sum of kinetic and potential energies) of an isolated system is the same in the initial and final states.

**Justification:**

Mechanical energy is conserved if there are no non-conservative forces that do net work on the system. The normal force exerted on the skateboarder is a non-conservative force, but the work that the normal force does is 0 because its direction is always perpendicular to the displacement. The gravitational force is conservative (it is already included in the potential energy term), and we are ignoring non-conservative frictional forces. Therefore, mechanical energy is conserved.

**Plan:**

1. Draw a picture and assign symbols for quantities in the problem. Choose a coordinate system.
2. Write an equation for conservation of mechanical energy. Expand the equation to include the initial and final kinetic and potential energy terms.
3. Solve for the height of the ramp. Substitute values to get an answer.
### Two-Column Solution

<table>
<thead>
<tr>
<th>Plan Step</th>
<th>Equation(s) used in step</th>
</tr>
</thead>
</table>
| 1. Draw a picture and assign symbols for quantities in the problem. Choose a coordinate system. | Initial state: \[ v_i = 6.5 \text{m/s} \]  
Final state: \[ v_f = 4.1 \text{m/s} \]  
\[ \begin{align*}
\Delta E &= 0 
\Rightarrow E_i &= E_f \\
KE_i + PE_i &= KE_f + PE_f \\
\frac{1}{2}mv_i^2 + mgh_i &= \frac{1}{2}mv_f^2 + mgh_f \\
\frac{1}{2}mv_i^2 + 0 &= \frac{1}{2}mv_f^2 + mgh_f
\end{align*} \]  
\[ m = 55 \text{kg} \]  
Mass of the skateboarder and skateboard combined  
\[ v_i = 6.5 \text{m/s} \]  
Initial speed of skateboarder  
\[ v_f = 4.1 \text{m/s} \]  
Final speed of skateboarder  
\[ h_i = 0 \text{m} \]  
Initial height of skateboarder  
\[ h_f \]  
Height of the ramp (final height of the skateboarder) |
| 2. Write an equation for conservation of mechanical energy. Expand this equation to include the initial and final kinetic and potential energy terms. | \[ \Delta E = 0 \Rightarrow E_i = E_f \]  
\[ KE_i + PE_i = KE_f + PE_f \]  
\[ \frac{1}{2}mv_i^2 + mgh_i = \frac{1}{2}mv_f^2 + mgh_f \]  
\[ \frac{1}{2}mv_i^2 + 0 = \frac{1}{2}mv_f^2 + mgh_f \] |
| 3. Solve for the height of the ramp. Substitute values to get an answer. | \[ mgh_f = \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2 \]  
\[ h_f = \frac{\frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2}{mg} \]  
\[ h_f = \frac{\frac{1}{2}v_f^2 - \frac{1}{2}v_i^2}{g} \]  
\[ = \frac{1}{2}(4.1 \text{m/s})^2 - \frac{1}{2}(6.5 \text{m/s})^2 \]  
\[ = \frac{1}{2}(9.8 \text{m/s}^2) \]  
\[ = 1.3 \text{m} \] |
**Discussion of Steps in the Conceptual Problem-Solving Approach**

**Identify the underlying principle**

A critical part of conceptual problem solving is for the solver to learn to identify what type of problem it is, that is, to categorize it based on the principle needed for solution. The goal is to have the students do this before trying to generate any equations, based on the problem context and features. Good instructors often do present the underlying principle, but then spend much time on the equations in the problem solving. Although they may assume students understand the importance of the principles and that the problem solving is meant as an illustration of how to solve problems of this type, there is much evidence that students learn in a much more specific way than was intended (e.g., Brookes, Ross, & Mestre, 2011; see later in chapter for a fuller discussion). Finally, most instructors present a problem as an example of a principle to be applied, but the categorization skill in need of being learned is to recognize the principle given a problem. Although these might seem to be very similar, the cognitive processing is very different. Categorization is very difficult not only because very different looking problems can be solved by the same principle, but also because very similar looking problems can be solved by different principles.

Categorization of problems is very difficult for learners in a complex domain and may take much practice, but starting with it as a requirement for problem solving forces an integration of the conceptual analysis and the equations. In addition, it can be made easier by starting when the students have few choices to select among near the beginning of a course. They can also begin to see how the categorization allows them to generate equations in the context of a plan. For example, they might begin to learn that conservation of energy allows them to set initial and final energies equal and realize that helps them to solve a number of problems with which they might otherwise have difficulty. This may also allow them to begin to see the commonalities across problems of the same type.

**Justify the principle selected.**

A critical part of expertise, which is often not explicit, is the checking of the applicability conditions for potential principles. The solver needs to use the problem content to explain why a principle is or is not appropriate in this situation. When presenting problems in class, instructors may present a problem and offer a principle to use, but they may not make clear why they are using this principle or may do so in a way that does not highlight the importance of providing a justification.
Notice that a result of not emphasizing this justification leaves the learner in a poor position for later problem solving. She may learn how a conservation of energy problem is solved, setting equal initial and final energies, but not learn how to tell whether or not an initial selection of the conservation of energy principle is appropriate. We view a major strength of the current approach as making explicit the often tacit knowledge for justifying principle selection in the context of problem solving. Not only does it help allow the learner to appreciate the underlying rationale, but making it explicit allows a student an opportunity to process the information, use it in later situations, and reflect upon it.

**Plan a solution.**

In this step, students are required to generate a plan for how to solve the problem in a general way, without referring to any specific equations. Much of the plan can be determined from the categorization, again emphasizing to solvers the importance of the principle in solving the problem. As they gain experience with the approach, they will learn to generate plan parts that are effective for decomposing the problem into sections for which meaningful sets of equations may be retrieved. For example, if they select conservation of energy, they can include early in their plan setting the initial and final energies equal to each other.

**Implement the plan**

(*Two-Column Solution*). In this part, the right-hand column of the two-column solutions is filled in with equations to correspond to the plan elements in the left-hand column. By decomposing the plan into meaningful steps, the approach further reinforces the idea that the equations follow from the principle, rather than from the problem’s surface features (objects and terms). The first time any equation is written, it is a high level one that captures an essential aspect of the principle, such as equating initial and final energy for the conservation of energy principle. This is a huge difference from the usual way novices generate equations. If such an approach leads solvers to better solve and understand solutions, it will reinforce the importance of conceptual analysis even for students whose goals are to better solve problems.

A second aspect of the implementation is a heavy reliance on worked-out examples during learning, rather than having students solve the problems themselves. Although many of us may feel that the active processing required in generating a solution is critical in learning, a large body of research shows that worked-out examples provide a much more efficient and effective learning process, particularly early in learning (*Paas, 1992; Sweller & Cooper, 1985*). Worked-out examples avoid students spending time and effort on incorrect paths and reduce memory load to allow for better learning of the correct solution path (*Sweller, 1988*). The tight link between the steps of the plan and the equations allows students to better understand
why these steps were taken and how they lead to the next step, which are essential components of building a forward-working approach to the problem solving (similar to ideas of the importance of self-explanations in learning; e.g., Chi, et al., 1989).

**An Empirical Study in High Schools**

This conceptual approach has been used in two ways in our work. First, we have conducted multi-session laboratory experiments with university students. Because of the background knowledge needed before new material can be taught and the difficulty of learning new principles, we were limited in how much could be taught in laboratory sessions. We opted for testing students who had earlier completed the basic physics course, so the sessions were a review of a number of earlier-learned topics but with an emphasis in the experimental group on this conceptual problem-solving approach. We did find very modest advantages for the conceptual group (Nokes-Malach, et al., under review).

Second, we have conducted a study in high schools over the course of a semester. We concentrate on this study here. High school classes have the advantage that students are learning the material for the first time and do not have prior approaches for solving these physics problems. However, they pose a set of challenges to an approach that requires a conceptual analysis of a problem. High school students vary tremendously on their facility with quantitative problems. There are also strong time pressures in high school classes to finish the mandated materials, yet much time must be spent on material not directly relevant to the topic, such as algebraic manipulations. In addition, as mentioned, many of the high school physics teachers were not trained in physics, and they, too, vary greatly on their understanding.

A full presentation of the study’s methods and results are included elsewhere (Docktor, Strand, Mestre, and Ross, under review), so we provide a short summary here. Teachers from three local high schools participated. There was great variation in teacher background and type of high school. The teachers included two physics-trained instructors, one with many years of experience and one with a few years of experience who mainly taught math classes. The other teacher was chemistry-trained in his second year of teaching physics. In addition the high schools varied with ones in a small city (32% of students classified as low income), suburban (12% low income), and rural (53% low income). One class in each school was taught using the conceptual approach, and one class was taught using the traditional approach. The traditional class depended on the school, with one being a different teacher and two having the same teacher (the same year or the previous year).

Our method was designed to help teachers teach problem solving more conceptually without changing their courses very much and without using additional time, while also allowing for
a variety of different classroom adaptations. That is, we provided a set of guidelines and materials, but allowed that the exact way in which they would be used would differ as a function of teacher expertise and preferences. Although the method was used throughout a semester for elementary mechanics, in each 2-3 week unit (e.g., conservation of energy) it might be used in only part of two days of instruction plus as one or two homework problems. The other days during a unit were spent in such activities as laboratory experiments, quizzes/tests, and the teacher covering content and explaining new material being covered.

At the end of the semester, a final assessment was given consisting of five types of tests to examine different aspects of what was learned, but we focus here on just two, the conceptual and problem-solving tests. These tests relate to the two main issues of concern here, understanding and problem solving (and were the only two with clear significant results). Table 1 shows sample items.

Table 1
Sample questions for the conceptual and problem-solving tests

**Conceptual test example:**

Elevators have a safety device to save occupants in the unlikely event that the cable breaks. The emergency friction brake presses against the sides of the elevator shaft if and when free fall is detected. Choose the statement that best describes the types of energy transformation processes (i.e., the exchange of energy from one type to another) that occur when the friction brakes are activated as the elevator is falling

- a. Kinetic energy is being transformed into potential energy.
- b. Kinetic energy is being transformed into potential energy and some energy is dissipated in heat/sound.
- c. Potential energy is being transformed into kinetic energy.
- d. Potential energy is being transformed into kinetic energy and some energy is dissipated in heat/sound.

**Problem solving test example:**

A 0.5-kg croquet ball is initially at rest on the grass. When the ball is struck by a mallet, the average force exerted on it is 230 N. If the ball’s speed after being struck is 3.2 m/s, how long was the mallet in contact with the ball?

The details of the procedure and results are presented elsewhere (Docktor et al., under review), but the simple summary is that there was a moderate and consistent advantage for the conceptual group in these tests, though it was not significant in every school. The conceptual test ranged from a 10-20 percent advantage and was statistically significant in two
of the three schools. The problem-solving test showed a 10-16 percent advantage but was significant in only the third school.

The teachers loved the conceptual approach and all planned on continuing it in future years. (One planned to adapt it for his math classes.) In one school, there is a district-wide test each year in a variety of subjects, including physics. The class having the conceptual approach performed better on 9 of the 12 conceptual questions (an average of 9% better overall, marginally significant) than the same instructor’s class did the year before.

**Summary of the High School Conceptual Problem-solving Study**

The goal of conceptual problem-solving approaches is to teach students to apply conceptual knowledge in solving problems. Rather than generate equations from the variables, terms, and objects in the problem statement and then use equation-oriented chaining to get an answer, instructors want students to think about the possible type of problem they are facing, identify the principle and justify their choice, and then use that knowledge to help determine how to solve the problem. To accomplish this goal, we think it is essential to integrate the concepts being taught with the problem-solving instruction and make explicit some tacit knowledge, particularly the justification for the selection of a principle.

Our particular approach to conceptual problem solving was to eschew major curricular changes. Instead, we focus on how to use a problem that might be presented in the class or homework and simply add some structure to the problem solving. As such, it can be blended into many courses. Although problem solving in this approach may take longer for a particular problem, at least initially, it may also lower the probability of students getting stuck with no idea of what to do. In classrooms, the number of problems that may be fully presented may be less, but we believe the extra benefit of conceptual understanding from this approach for each problem more than makes up for the smaller number.

The early results are suggestive that this approach can improve students’ conceptual understanding and even their problem solving. Although the sample sizes were small and the effects were modest, it is important to remember three points about the intervention. First, it involved no redesign of the course, just simple changes to the instruction. Second, the intervention was a small part of the instruction, often just parts of two classes in a 2-3 week section, with a little on the homework. Third, this was a first attempt at applying it in high school classes. The classroom instruction was not optimized to make good use of the approach—teachers used it as they wanted to with only minor guidelines from us. There are a variety of minor changes (some mentioned in a later section) that we think would help, plus ways of supplementing this approach to help with the difficult selection/justification parts.
Laboratory Empirical Studies: A Closer Examination of Processing Difficulties

Although addressing cognitive issues in education ultimately relies on finding effects in real educational contexts, a better understanding of the issues often requires examining them in a more controlled environment. The research on the overall conceptual problem-solving approach is the main focus of the work we have done, but we have supplemented it throughout with finer investigations of particular difficulties. Due to space limitations, we mention only two here: whether categorization on the basis of principles can be improved with a short intervention, and the effect of illustrative examples on the transfer of abstract principles to new contexts. In this section, we briefly review the rationale and findings of each of these projects and provide some suggestions for how to incorporate these ideas into teaching physics.

Can short training increase principle-based problem categorization?

Problem categorization (the identification of the relevant principle) is a critical part of problem solving for experienced solvers because it helps the expert access the appropriate plan and equations. Although novices do not categorize problems on the basis of principles (e.g., Chi, et al., 1981), we asked here whether we might begin to influence the basis of novice categorization with a short training program. While the knowledge base that experts use for principle-based categorization may take years to acquire, there may be ways to train people to focus on the important aspects for categorization even in complex domains (e.g., Kellman & Massey, 2013). Docktor et al. (2012) examined whether novices can be taught to categorize problems on the basis of the underlying principle rather than objects and terms in the problem statement. On each of the 36 trials, two physics problems were presented on a computer screen and the task was to respond whether the problems would be solved in the same way or not. The problems could match on the underlying principle, surface features, both, or neither. For half the subjects, each trial was followed by standard feedback informing the subjects of the correctness of the response and the reason. For example, subjects would read, “(In)Correct. These problems would (NOT) be solved similarly.” The other half of the subjects were given elaborated feedback to focus them on the relationship between the problem and its underlying principle. For example, they might read, “Correct, they would not be solved similarly. Problem 1 uses force and acceleration so it would be solved using Newton’s Second Law. Problem 2 includes mass and change in speed with no external forces on the system, so it would be solved using conservation of momentum.” Going back to the conceptual problem-solving approach, this elaborated feedback is analogous to the selection...
and justification steps. Eight times during the training, the subjects were asked to explain their response: why did they believe the two problems should be solved similarly or not?

The main dependent variable was the type of explanation subjects gave on these trials, and the results are straightforward. Those students getting elaborated feedback increased their use of principle-based explanations and those students getting standard feedback did not (see Docktor et al., 2012). However, the categorization accuracy was no better in the elaborated explanation group, because these students would often cite inappropriate physics principles, resulting in an incorrect categorization. Thus, this short training did not turn novices into experts. The main point is that with the appropriate training, even novices can begin to learn that the rationale for problem solving needs to be in terms of the underlying principles, not the problem statements.

**Instructional implications.**

This technique could easily be adapted into curricula. For example, discussion sections or homework might include a short set of problems to categorize, with new problem types added throughout the course. By the end of the course, students would have had substantial practice categorizing problems, a main difficulty for students in current instruction. We used pairs of problems with same/different responses, to avoid spending much time on teaching the names of the underlying principles, but in a course naming is already taught, and students can be given a single problem and asked to choose the appropriate underlying principle.

One could also integrate this into the conceptual problem solving we have presented here (as one of our teachers did). Currently, each problem requires going through all the steps, but the planning and implementation steps take far more time than the selection and justification steps. However, the selection and justification are often more difficult to learn. The elaborated feedback training can be viewed as allowing students to get much more practice on the critical selection/justification parts, and yet allow them to integrate it with the full approach by asking them to solve a problem every once in a while.

*What do students learn from illustrative examples?*

A common instructional technique is to begin with an abstract principle and then apply the principle to concrete examples. These illustrative examples are meant to help students develop a deeper understanding of the principles. However, there is good evidence that students may learn in a much more restricted way than intended (e.g., Ross, 1987). This issue is a critical one for problem-solving instruction, because it suggests instructors may misunderstand the generality of what students are learning. We outline here what the
underlying difficulty is and mention some research that examines this issue in physics learning.

People have difficulty thinking abstractly. Without a deep understanding of a domain, the meaning of an abstract principle is not at all clear. The students use the illustrative example to try to figure out what the principle might mean. In fact, it is common for students of all levels to quickly read a principle and jump to the example to help figure out what is really being taught. These examples are necessary for providing understanding, but they have an unexpected consequence—the understanding of the principle may be bound to some of the particulars of the examples used to illustrate the principle. That is, the principle may initially be learned more specifically than intended. Even when students seem to be able to apply the principle correctly in some cases, past research has shown that when a test problem is similar enough to an illustrating example, students may end up being influenced by the specifics of that earlier example. Brookes, Ross, and Mestre (2011) showed this effect occurs in physics learning even when students have a strong background in the domain (most of the earlier work had examined learning in a new domain). Students in a calculus-based introductory physics course were taught refraction with a short lecture including the teaching of wave fronts and a laser demonstration showing the bending of light at an air-plastic interface. They were then given an example of light bending using a particular prism to illustrate refraction of light both entering and exiting of the prism. On the test, however, even students who showed evidence of having learned the principle when tested on a single interface were influenced by the specifics of the illustrative example when given a test example similar to it. When shown a test example similar to the illustrative example, their refraction rays were biased in the direction that the illustrative example rays were (see Brookes, et al., 2011, for specifics). Although they appeared to have learned the principle, their understanding was not completely separated from the learning example.

**Instructional implications.**

The effect of these examples shows that students may not be learning what was intended, but there are ways of overcoming this specificity effect through further instruction. One common means is to compare multiple examples. In most instruction, multiple examples are given to illustrate a principle. Although the instructors’ intuition that this may be helpful is correct, it may not be sufficient. It may also be necessary to require a comparison of multiple examples to allow students to understand what are common aspects of the various examples illustrating a principle (e.g., Loewenstein, Thompson, & Gentner, 2003).
**Thinking Back, Thinking Forward**

In this section, we consider the research program more broadly and ask what we wish we had known before, what we see as potential future projects, and how this work can be generalized to other STEM disciplines.

**What we wish we had known**

It would be impossible to look back on a research project and not wish one had done something differently. There are a number of decisions we might have changed given what we know now, but we focus on two that we think have some general interest.

First, we believe that when teaching a new problem-solving approach, such as conceptual problem solving, it is important to get students as they are first learning the domain. Our earlier work targeted university students after they had had a college course, with the experiment reviewing partially forgotten principles and teaching a new approach. Although we did see some effects, they were small, and for some students it appeared that extra practice (we equated overall time) was at least as beneficial as introduction to a new approach. Further, those students had already developed some problem-solving technique that they used in the course, and they may have been less likely to abandon it in favor of learning an entirely new approach.

Second, we did not take into account that the specifics of problem-solving instructional times may not fit well with the instructional objectives. In particular, parts that students may have the most difficulty with (identification, justification) may be a small proportion of overall problem-solving time. The result is that much of the time spent on conceptual problem solving in class and homework was spent on the planning and equation generating, not on what may be the more conceptually difficult parts. We address this below.

**Future directions**

Although many questions remain unanswered, two areas within this conceptual problem-solving approach that we see as holding much promise are training identification/justification and use of comparisons.

**Identification/justification**

If we want students to not rely on superficial aspects of the problem statement followed by equation manipulation (and we do want that), they need to have an alternative approach with which they are confident and which they can apply to generate a correct solution. It is
difficult to tout the conceptual problem-solving approach to a student who has no idea what the correct principle is or why it applies. We assumed that as students received instruction in class and were given practice on their own in which the possible principles were limited to just a couple, that they would progress rapidly enough to see the advantages of this approach. However, they actually received very little practice on identification of the principle (categorization) and it was, time-wise, a small part of their training.

We believe now that they need more practice and more scaffolding to make them able to confidently classify, as well as more help on justification. We would provide separate categorization practice in three ways. One, as one of the teachers did on his own, we would include collaborative groups who would practice categorizing and justifying problems without going on to solve them. This provides both practice and instruction, as they can see how others think about the problems. Two, we would have students take a short amount of time each week to practice categorizing problems with principle explanations linking the problems’ surface attributes to the principles, as in the Docktor et al. (2012) study. Kellman (e.g., Kellman & Massey, 2013) has shown that even difficult structural categorizations can be learned through extended practice of this kind. Three, we would provide students with scaffolding as they learn to categorize problems and justify their decisions. As an example, we constructed a very rudimentary tutor program that breaks the decision down into a series of simple steps, providing immediate feedback on each step. Such a tutoring program might be implemented in many different ways and used in weekly problem-solving labs or homework exercises.

In our first attempt, modeled a bit after the Hierarchical Analysis Tool (Dufresne, et al., 1992; Mestre, et al., 1993), the students choose from a set of choices at each step. They are first asked to categorize the problem at a high principle level (e.g., conservation of energy) and then told if correct or not. If incorrect, the student chooses again, until correct. The tutor could be set to give simple or more elaborated feedback, explaining why in fact this is an example of conservation of energy. They are then asked a series of related questions in multiple-choice format to help them both see why the specific principle applies (justification) and to better understand problems of that type. For example, they would have to decide what type of energy is present (potential and/or kinetic) in the initial and final states, without using the specific numbers in the problem. If there is potential energy present in either the initial or final states, they would be asked to specify whether it is gravitational or spring potential energy. Although we recognize that some research might show better ways of setting up and integrating this tutor, the main idea is to allow students to practice on their own and to be given both feedback and further instruction as they do. For both types of categorization training, new principles can be added as the course progresses, giving students a greater number of principles to choose among as they gain further proficiency. These three types of
practices (group discussions, categorization training, and step-by-step scaffolding) might all be profitably integrated into this, or other, conceptual problem-solving approaches.

**Comparison.**

The second direction we think important is to provide guided comparisons in learning. We did not implement this in the high school study due to a lack of class time, but we have in other studies *(Nokes-Malach et al, under review)*. Multiple examples help students understand what is common among problems of the same type (and different between problems of different types). However, as already known by many teachers, it is not sufficient to provide multiple examples—students must be helped in comparing both problems and problem solutions if they are to see the underlying similarities and differences (e.g., *Loewenstein, et al., 2003*). We know comparisons help, but there is much still not known about the best type of comparisons when teaching complex domains (see *Rittle-Johnson & Star, 2009*, for a thoughtful extensive treatment).

**Generalizability to other STEM disciplines**

How well might this approach work in other STEM instruction? The critical issues that we think make it successful in physics are the following: there are a small set of underlying principles into which problems can be categorized, what principle can be used to solve a problem is not obvious from the problem’s surface features, and the correct categorization of a problem by principle provides a huge advantage in solving the problem since it helps lead to the appropriate equations.

One STEM-related domain with these features is statistics (some other mathematics domains also have these features, though some do not). One of the teachers in our study who also taught math was planning to implement a similar approach in his math classes. The first author taught introductory statistics for the behavioral sciences for almost 10 years. The greatest difficulty students had in the course was correctly categorizing the nature of the problem. If he were still teaching that class, he would adapt this approach. In fact, the course TA who received one of the highest student satisfaction ratings implemented categorization practice during discussion sections in reviewing for the final exam. The students would be given a problem and have to say what type of statistical test would be appropriate; the students would then discuss any possibilities and why it might or might not be appropriate. Once there was agreement (and the TA agreed it was correct), another problem would be given. (Although we feel less confident of further extensions, queries of domain experts suggest some chemistry topics may be amenable to this instruction, whereas biological domains are less likely to have these features.)
Concluding Remarks

Good students usually learn in a way that is consistent with how the course material is taught and assessed. Introductory physics courses often rely heavily on problem solving in both instruction and assessment. The goal of the conceptual problem-solving intervention presented here is to integrate the problem solving with the conceptual understanding so that students learn to solve problems from a conceptual analysis. The results are suggestive that such an intervention may improve both conceptual understanding and problem solving. It is important to understand that these results occurred with little change in the course, little time spent on the intervention, and without earlier attempts to optimize its use.

A goal of the workshop in which these chapters were first presented is to improve the communication and help develop general principles for useful intervention. We end with a mention of three points that were brought up in the workshop, with which our research is consistent and that we think are important for other interventions as well. First, a suggestion was to “pick one thing” for an intervention and focus on that (Perkins). Second, consider an intervention that does not involve a redesign of the whole curriculum (McDaniel). Third, much knowledge in these domains is implicit, and understanding what it is and how to better teach it is crucial for improving instruction (Koedinger). Teaching students to make explicit their justification is a critical part of the conceptual problem-solving intervention.

References


6. Recommendations for Instructors and Students: A Commentary

doi:10.7936/K7TD9V7Q
Mark A. McDaniel and Henry L. Roediger, III
http://orcid.org/0000-0002-3314-2895

Washington University in St. Louis

Correspondence:
Mark A. McDaniel
Washington University in St. Louis
St. Louis MO
mmcdanie@artsci.wustl.edu

Recommendations for Instructors and Students

The chapters in our purview have many important implications for instruction and learning, for both faculty and students. In our commentary, we pull out and summarize some key points for both faculty and students to keep in mind when teaching and when studying, respectively.

Recommendations for Instructors

A key organizing theme provided by Koedinger and McLaughlin is that effective teaching relies on clarifying two goals: (a) identifying the concepts and skills that the student needs to acquire and (b) identifying instructional methods that best support student learning of those skills. With regard to (a), two components are highlighted in the chapters. Here we focus on each in turn and, while doing so, highlight instructional methods that the chapters reveal as promising innovations.

Acquiring Conceptual Understanding

One major component of what students need to learn is reflected in science courses that are oriented toward problem solving, such as physics, chemistry, and engineering. For these disciplines, a strong case can be made that the student needs to acquire a conceptual understanding of the underlying principles that are brought to bear to solve problems and that they approach problem solving from this conceptual basis. In some sense, this may seem an obvious claim. After all, it seems likely that in most all college level courses, instructors are providing the conceptual underpinnings of the science and then stimulating students to
practice the concepts in problem solving exercises. A key point, however, arises from research in cognitive science: Although instructors feel that they are approaching the topic from a conceptual basis, the students can be acquiring something very different, especially in early stages of learning. As Ross et al. note, students’ approach to solving their homework and exam problems can be based on surface features of the problem—such as the particular cover story in which the problem is embedded—and heavy reliance on the equations that have been taught in class. They may miss the conceptual forest by focusing on the trees that enable problem solving without deep understanding. Indeed, work in one of our labs suggests that a sizeable proportion of students have a preference toward an exemplar-learning approach, extrapolating from an example problem to a test problem. That is, these learners focus on memorizing the particular examples presented to illustrate concepts, and they approach new examples based on how similar these new examples are to the surface features of learned examples (rather than in terms of conceptual relations; McDaniel, Cahill, Robbins, & Wiener, 2014). Further, our preliminary results show that this learning preference (displayed on laboratory conceptual tasks) predicts how students will fare in introductory chemistry at the university level. Students with exemplar-learning preferences achieve lower performance on exam problems that require generalization and synthesis than students with conceptual-learning preferences.

The frustration for the instructors is that their pedagogical goal is to promote conceptual understanding of underlying chemistry principles, yet a sizable proportion of students seem to persist in equation-oriented problem solving. Thus, the lesson here is that instruction needs to be strategically and explicitly formulated to assist students in achieving conceptual understanding, because such understanding may not be easily achieved for at least some students who prefer to learn from examples. One such technique advanced in this volume is Process Oriented Guided Inquiry Learning (POGIL; see Moog’s chapter). POGIL is based on a three-phase learning cycle of inquiry learning (Lawson, 1999; 2001). In the first phase, students explore to attempt to find a pattern of regularity in a laboratory experiment, a demonstration, or graph/diagram; next, a concept is developed from the presented information; and finally the concept is applied in new situations. The learning cycle experience guides students to construct a conceptual understanding of the key principles and their application in solving problems. Several practical limitations of POGIL, however, are that instructor development is relatively extensive and POGIL classes can take 50 percent more time than standard classes. Accordingly, because of resource limitations at our institution, POGIL recitation sections can be offered to about one half of the general chemistry students. An intriguing solution to optimizing the benefits of this instructional technique might be to identify students’ conceptual learning tendencies at the outset of a course (using a laboratory conceptual-learning task). Those who display an exemplar-learning orientation would be assigned to the POGIL sections to help guide these students to construct a conceptual understanding they might otherwise not achieve. Our Center for Integrative Research on Cognition, Learning, and Education (CIRCLE) is currently exploring the fruitfulness of this approach to optimizing limited instructional resources (for POGIL).

A second instructional technique, emerging from Ross et al.’s chapter, is to teach students to categorize different problem types depending on the principle(s) needed to solve the problem. This skill provides a conceptual basis for students to generate a solution plan and retrieve appropriate sets of equations for new problems. Interestingly, the standard classroom
practice of blocking instruction on one type of problem and then proceeding to instruction on another type of problem likely undermines opportunities for students to acquire skill in identifying and categorizing different problem types, as is illustrated by the research on category learning presented in Bjork and Yan. That work clearly showed that category learning is impaired when all instructional instances of one category are presented, then the instances for the next category are learned, and so on.

The ability to identify the critical features that discriminate among categories is fostered when instruction interleaves examples from different categories. This was shown in compelling fashion by Rohrer and Taylor (2007), who found that blocked instruction on how to find the volume of different kinds of shapes produced quite low performance on new problems a week later, whereas interleaving practice across the different shapes supported relatively good performance a week later. A key aspect of this effect was that the students in the blocked instructional condition did not forget the equations—rather they made errors in applying the appropriate equation to the test problems. Combining ideas illuminated in the Ross et al. and Bjork and Yan chapters, we suggest that incorporating instruction on how to categorize different problem types—to provide a more conceptual approach to problem solving—will be most effective if the instruction is also modified to include some interleaving across problem types. For instance, in the chemistry classes at Washington University, instructors are considering modifying the homework so that problem types from several weeks prior are included, not just the problem type that is the focus of the current week’s instruction.

**Learning Facts, Principles, and Categories**

A frequently displayed banner of contemporary educational reform is that instruction should replace a focus on learning facts and principles with a focus on learning conceptual skills necessary for problem solving, inference, and creativity. Cognitive scientists, however, emphasize that “One cannot analyze what one knows if one knows nothing. And one cannot apply what one knows in a practical manner if one does not know anything to apply” (Sternberg, Grigorenko, & Zhang, 2008, p. 487; see also Willingham, 2009). Similarly, in education, Bloom’s (1956) theory assumes that remembering and understanding facts and principles are requisite for application, synthesis, and evaluation. We agree that acquiring conceptual approaches to problem solving in STEM and learning facts, principles, and foundational categories are interdependent companions. Several of the chapters in this volume from cognitive scientists reveal techniques to promote robust learning of key facts, principles, and categories that depart from standard practice.

One converging recommendation is that instructors expand their view of tests as only summative assessments to also view tests as potent learning techniques. From our perspective, integrating low or no-stakes tests into STEM instruction is especially useful for STEM areas that are heavily fact and category laden and that do not afford ready problem-solving exercises. Two clear examples are biology and psychology, which unlike physics, chemistry, mathematics, and engineering, do not necessarily require finding solutions to computational problems. Instead, students are faced with having to learn components of theories, central findings, categories of processes and principles, and so on (to which many
biology and psychology students can attest). Tests as learning exercises provide students with practice in retrieving and applying concepts and facts (just as homework problems in physics and math provide students with practice in generating problem solving solutions and retrieving appropriate equations), and as several of the chapters report, this retrieval practice produces robust gains in learning and retention.

In the public forum and from some teachers, a not uncommon reaction to reading about the use of tests to promote retention of content is met with the objection, “Once again, another author confuses learning with recalling information” (online comment reported in Brown, Roediger, & McDaniel, 2014, p. 29). We strongly emphasize that tests can provide practice on a range of uses of knowledge that dovetail with many instructors’ educational objectives. For instance in university biology and psychology courses, instructors are generating questions designed to require application, synthesis, and evaluation of target concepts. For example, after learning about operant conditioning in introductory psychology, a quiz could ask students to describe how they would train a dog to roll over on command. The student would have to use concepts of discriminative stimuli, reinforcement, shaping, and others in constructing her answer. Thus, tests can demand more of students than simply retrieving target information—they can provide rich practice in application and generalization of concepts, which in turn promotes performance both on summative exam items that also require application and on items that test retention and understanding (e.g., see Jensen, McDaniel, Woodard, & Kummer, 2014, for a biology classroom finding).

Another recommendation on which the cognitive science chapters converge is that review of core content will more effectively support learning and long-term retention if that review is spaced rather than massed (in single lectures or homework assignments). Much research supports this recommendation, so here we briefly touch on several pragmatic issues that pose challenges to implementing this recommendation in authentic classroom contexts. One challenge is modifying the common practice in college classes of focusing coverage on a particular topic, finishing the topic, and proceeding to the next topic. There is minimal revisiting of topics preceding the current topic until the students take the exam. Here are some techniques that some of our faculty colleagues are beginning to use to implement spacing in their courses. In chemistry, the weekly quizzes previously focused on the current week’s content; they are now being modified to include some questions revisiting topics, concepts, or problems from preceding weeks. In psychology, one instructor reserves a small amount of class time for review of material presented in prior weeks. In statistics, the standard practice for homework is to assign problems that reflect the type of problems being currently addressed in class. One instructor revised his homework sets so that they now include a problem or two from previous sections of the course.

A second question frequently posed by instructors concerns what the optimal interval is for spacing repetitions. So far, the research has provided is no precise answer to this question. However, it appears that the optimal spacing interval is some ratio of the anticipated retention interval. One study that sampled a number of spacing intervals and retention intervals (with retention intervals of up to a year), suggests as a rule of thumb that the intervals for spacing should be approximately 15 percent of the anticipated retention interval (Cepeda et al., 2008). For example, for the spring statistics course, which is a prerequisite for higher-level courses that students will be taking in the fall, material should be reviewed
about every 2 weeks (assuming a 15-week summer break). Innovative work with computer tutors is using complex models that include parameters like the difficulty of particular items and the student’s ability to provide an optimized schedule of spacing for each particular target item to be learned (for vocabulary learning in foreign language classes; Lindsey, Shroyer, Pashler, & Mozer, 2014). As experimental work with this technology matures, we expect that more precise guidelines for optimal spacing may emerge.

**Closing Thoughts for Instructors**

Returning to the Koedinger and McLaughlin chapter, their major proposal is that the effectiveness of domain-general instructional techniques revealed in basic cognitive science work (chapters by Bjork & Yan, Marsh & Cantor, Rawson & Dunlosky) may depend on the instructional domain. As one example, they suggest that for domains (or learning requirements) that are heavily fact laden and require remembering, that instruction that elicits responses is preferred (e.g., testing, as discussed above), whereas for domains that are heavily problem-solving oriented, telling (giving the answer) is preferred. This is a provocative suggestion that may prove fruitful in advancing the effectiveness of instruction. However, very recent work with computer tutors suggests that an even more fine-grained intermixing of instructional techniques may be optimal. These tutors base instructional policy on empirical analyses of the effectiveness of instructional choices at each micro-step of the instructional session (Chi, Vanlehn, Litman, & Jordan, 2011). The upshot is, at least in physics instruction, that for some of the micro-steps eliciting is better and for others telling is better. Thus, we believe that the techniques presented in this volume provide a valuable contribution toward optimizing science instruction, but this is not the endpoint. In the coming years even more sophisticated combinations of these techniques may provide significant gains in student learning relative to what we are accustomed to presently.

**Recommendations for Students**

Although we are dividing our commentary into advice for faculty and students, we admit that this division is somewhat arbitrary. After all, many points of advice for faculty are about how to induce good learning in students. It stands to reason that students should use these same strategies in their studies outside of class.

Bjork and Yan describe steps students should take to become a sophisticated learner. Let us first pause to discuss the unsophisticated learner, too often seen even in university settings: This is a student who does not read the assignments on time, does not come to class prepared, may not come to class frequently, waits until the last moment to complete assignments, and crams before exams. Advice representing low-hanging fruit is advising students to follow the opposite of all these practices, but, of course, they probably already know this, if asked. The trick is to get the students to follow better practices. The recommendations to instructors help to do this: If a quiz is given every day or every week, students must study and attend class and then the retrieval practice from the quiz will further help them learn critical concepts.
Many students would like to be sophisticated learners but do not know how. They read, reread, and review, but they can be fooled by illusions of knowing from the fluency of repeated reading. Bjork and Yan discuss bringing desirable difficulties into play so students force themselves to learn more and to be able to monitor their state of knowledge. They should space their practice in time, interleave learning of one subject with other subjects, test themselves at spaced intervals, and give feedback when they make errors. Students should not fear generating errors but welcome it if they are testing themselves under difficult conditions and striving to improve their knowledge above their current level. Feedback after errors is critical, however.

If there is one central element in Bjork and Yan’s advice to students it is that students should overcome their belief that the easiest method of learning is the surest path. Repeated reading or review is easy to perform, but countless research studies show that testing oneself with feedback provides better long-term retention and corrects metacognitive errors.

Rawson and Dunlosky’s chapter is also about how learners can self-regulate their learning to be more effective. They suggest the technique of successive relearning. If students are taking an introductory psychology course in which they cover 15 chapters of disparate material in 15 weeks, they will have to learn a huge number of conceptual terms: conditioned stimulus, opponent process color theory, Piaget’s stages of cognitive development, the fundamental attribution error, and on and on. The research by Rawson and Dunlosky suggests that students create flashcards (or the computer equivalent) and cycle through them, testing themselves at long intervals and then relearning the information as needed. Research with flashcards suggests that students often drop cards too soon (thinking that if they have gotten an item right once or twice, then must know it). However, much research shows that continued practice beyond the first correct recall greatly improves long-term retention (Karpicke, 2009; Karpicke & Roediger, 2008). Pyc and Rawson’s (2009) research shows that somewhere around 5-7 correct retrievals provide optimal long-term retention, at least in their experimental situation involving the learning of foreign language vocabulary. The method of successive relearning (especially if spread over wide amounts of time) would surely be effective in keeping knowledge accessible. However, successive relearning is hard work and perhaps only the most dedicated students will actually put it into practice. That is why cumulative quizzes in the classroom can be so helpful, as well as providing a cumulative final exam (Roediger, Putnam & Smith, 2011).

The chapter by Ross et al. provides advice for instructors in how to make students sophisticated in solving problems in physics. They should identify the underlying principle represented by the problem, justify why the principle is appropriate, plan on how to solve the problem, and then implement its solution (with equations, where appropriate). These steps represent excellent advice. As the authors comment, once students have learned to apply these steps in one domain, such as physics, they may well learn the general skill of using these steps to solve problems in math, chemistry, or engineering courses. Again, learning this approach would make the students more sophisticated learners.
**Summary**

The chapters we have synthesized here provide excellent advice for both instructors and students. Our summary comments hardly do justice to the full force of the ideas expressed in the chapters under consideration, but they highlight the potential positive impact that applications of cognitive science can have in STEM instruction.

**References**


doi:10.7936/K7KW5CX1
Stephanie V. Chasteen
http://orcid.org/0000-0002-7731-2243 and
Katherine K. Perkins

University of Colorado Boulder

Correspondence:
Stephanie V. Chasteen
Science Education Initiative and Department of Physics
University of Colorado, Boulder
Boulder CO 80309
http://www.colorado.edu/sei/

Abstract

The Science Education Initiative (SEI) is a university-funded project which aims to achieve highly effective, evidence based education for students. To achieve these goals, the SEI supports work at the departmental level to transform undergraduate STEM courses by establishing what students should learn, determining what students are actually learning, and improving student learning using evidence-based instruction. The main features of the program are a departmental focus, and the hiring of science education specialists embedded within those departments to support this work. The outcomes of SEI work are diverse and include the transformation of specific courses, addressing department and institution cultural issues, researching the impact of pedagogical techniques on learning, and dissemination of course and related materials. In this chapter an example of an SEI course transformation in physics is discussed in detail.

Introduction: Why the Push to Improve STEM Education Now?

In this chapter, we present one model of change—the Science Education Initiative (SEI) at University of Colorado Boulder (CU)—that has leveraged several key conditions to engage
departments and faculty and to transform how Science, Technology, Engineering and Math (STEM) is taught and learned at the University of Colorado Boulder (CU).

Responses to calls for improvements in STEM education are receiving unprecedented attention, support, and action—with many more organizations assuming some responsibility for addressing this need. Universities are investing resources and establishing major efforts (e.g. Association of American Universities, Association of Public and Land-Grant Universities, Florida International University, University of Utah, University of Colorado Boulder). Disciplinary societies are assuming new roles and spearheading new programs (e.g. the Physics Teachers Education Coalition, http://www.phystec.org/, or the New Faculty Workshop in Physics and Astronomy, http://www.aapt.org/Conferences/newfaculty/nfw.cfm, both of which aim to improve the educational preparation of college instructors). Business leaders are coming together to identify and support programs that work (e.g. Change the Equation, http://changetheequation.org/). New national efforts are creating networks of partners to increase the impact of the collective (e.g. 100Kin10, http://www.100kin10.org/). Federal agencies are creating new programs to target widespread change in STEM teaching (e.g. WIDER) (National Science Foundation, 2012).

Additionally, significant advances in our knowledge of how people learn in general (Bransford, Brown, & Cocking (Eds.), 2000) and, specifically, how students learn STEM topics are available to inform and guide these new efforts. The field of disciplined-based education research (DBER) has emerged with a focus on researching and transforming STEM education at the undergraduate level; its community of disciplinary researchers engage in developing new, effective teaching practices; conduct content-specific research on student ideas, thinking, and difficulties; and create better assessments of learning, among other research activities (Singer, Nielsen, & Schweingruber, 2012). While DBER is a growing field and there is much work yet to be done, the DBER community has already succeeded in creating a vast collection of new teaching practices and approaches that are aligned with the research on how people learn and that have been shown to improve student learning in science. Specific examples include Peer Instruction (Mazur, 1997), Just-In-Time-Teaching (http://serc.carleton.edu/introgeo/justintime/index.html), Process-Oriented Guided Inquiry Learning (http://pogil.org/), SCALE-UP (http://www.ncsu.edu/per/scaleup.html), or Tutorials in Physics (McDermott, Schaffer, & the Physics Education Group at the University of Washington, 2002). While the DBER community will continue to study teaching and learning, classroom teachers already have access to research-proven approaches that they can implement today to improve student learning and engagement in science.

Thus, we find ourselves at a unique time in the history of science education—a time at which the foundational knowledge, the practical approaches, and the resource investment needed to achieve substantive, widespread improvements in science education are present. The CU SEI
takes advantage of this powerful combination to create a department-centered, bottom-up model of educational change.

**A Vehicle for Reform: The SEI Program**

In 2005, the CU Science Education Initiative (SEI, [http://www.colorado.edu/sei/](http://www.colorado.edu/sei/)) was launched as a five million-dollar, university-funded project to support departments in improving science education. The project was spearheaded by Dr. Carl Wieman, who also started a sister program in 2007 at the University of British Columbia, [http://www.cwsei.ubc.ca/](http://www.cwsei.ubc.ca/). In this chapter, we describe the structure of the program, present the program’s approach to course transformation, provide an example of the program in action in the classroom, and summarize lessons learned. Finally, we provide a list of recommended readings for individuals interested in learning more about the SEI, about research-based instructional approaches for science education, or about institutional or faculty change in STEM.

The SEI program is structured with a small team of central staff (SEI Central) serving a role similar to that of a funding agency—albeit a funding agency that provides advice, training, knowledge, and oversight to its grantees. Departments initiate their involvement in the SEI through a proposal process, and seven departments have received funding to conduct their proposed work. The funded departments include Molecular, Cellular, Developmental Biology (2006-11, extension 2011-13); Integrative Physiology (2006-12); Geological Sciences (2006-11); Chemistry (2006-11); Physics (2007-11, extension 2011-13); Astrophysical and Planetary Sciences (2011-13); and Ecology and Evolutionary Biology (2011-2014).

A key element of the program is its departmental focus; at a college or university, the department is the natural unit where faculty are primarily affiliated, conducting their work and teaching, engaging in discussion with fellow faculty, and evaluating departmental policies and priorities. Within the SEI structure, departments must initiate participation, deciding as a unit whether to submit a proposal. In fact, the proposal must indicate the fraction of faculty who have voted in favor of participating in the SEI. This structure is designed to avoid the alternative top-down reform structure and, instead, create a scenario where departmental faculty have collectively discussed SEI participation, and the majority have expressed the desire and commitment to engage in improving science learning before proposing to participate.

During the proposal process, the SEI provides departments with a research-based framework for improvements in science education (see section 3). This framework guides departments in the types of activities that departments can support with SEI funding. While the framework encourages a range of productive activities, departments retain significant latitude in their
ability to structure their proposed work to reflect their faculty’s agreed upon priorities—e.g. introductory courses, laboratories, courses for majors or non-majors, etc. The most recent call-for-proposals, conducted in 2010, is available online: (http://www.colorado.edu/sei/about/funding.htm).

Finally, impactful change takes a significant investment of time and effort, as well as content and pedagogical expertise. For faculty, however, time is a precious resource. Recognizing that faculty work under significant time constraints, all departments receiving SEI funding have used the majority of that funding to hire Science Teaching Fellows (STFs) to partner with faculty on these efforts. STFs are hired as a member of the department and engage in a wide variety of activities (Table 1) to organize, facilitate, and enable the department faculty to achieve the changes set forth in their department’s proposal to the SEI.

To be effective in their roles, STFs need 1) significant content knowledge, 2) knowledge of research-based instructional practices and education research methodologies, and 3) excellent interpersonal and facilitation skills. STFs have typically earned a recent PhD in the discipline and are interested in a discipline-based, education-focused post-doctoral experience. With their PhD-level content-expertise, the STFs can meaningfully engage in content-specific work with the faculty, such as discussing learning goals, creating homework, and designing assessment items. Often, the STFs have also demonstrated significant interest in teaching and learning within their discipline. They have participated in their institution’s equivalent of teaching and learning programs or engaged in extra teaching opportunities. With only a limited number of PhD programs in discipline-based education research, most do not hold a PhD in education research in their discipline. Importantly, the department itself conducts the search for the STF and is responsible for all hiring decisions.

While the STFs are all situated and work within their respective departments, most new STFs are in need of additional training on education and cognitive science research, on implementation of research-based instructional practices, and on education research methodologies. Thus, the SEI Central project team has commonly provided training and advice to new STFs and coordinated regular meetings of the STFs, creating a community of STFs engaged in similar work within their respective departments.
Facilitating faculty communication and consensus building

Collecting, distilling, and communicating data to support and guide faculty efforts

Developing curricular materials and teaching approaches in collaboration with faculty

Serving as a local resource for department faculty

Facilitating sustainability by archiving and disseminating materials

These three elements of the SEI—the departmental focus, the guidelines on SEI-fundable activities, and the use of STFs as agents of change—are designed to create an environment where departments meaningfully and successfully engage in department-based, faculty-driven efforts to improve student learning in science. While departments have engaged in a variety of activities around improving science education for their students (see below), transforming individual courses—as described in the next section—has been a major activity in all departments.
Due to varying departmental needs, the outcomes of SEI work have been quite diverse, focusing on various aspects within the educational system:

**Departmental culture and community**
- Offering various department-based professional development opportunities
- Facilitating faculty brown bags and retreats to discuss undergraduate education within the department
- Creating and implementing new TA training programs

**Transformation of a specific course**
- Facilitating faculty discussions of learning goals
- Compiling literature on student thinking about course topics as well as existing instructional materials and resources
- Writing and collaborating on instructional materials: clicker questions, class activities, tutorials, homework, invention activities, labs, and recitation activities
- Supporting faculty in implementation and use of research-based teaching practices
- Providing instructors with feedback on classes, homework, and exams as well as any insights into student thinking from synergistic research activities

**Research on and assessment of learning**
- Investigating student thinking about disciplinary topics to gain a better understanding of student difficulties and misconceptions (e.g. conducting cognitive interviews, problem solving sessions, etc.)
- Researching the impact of pedagogical techniques on student learning and engagement (e.g., clicker use)
- Developing, administering, and analyzing pre-/post-conceptual assessments

**Dissemination**
- Compiling course and related materials into archives and making them freely available for re-use by the broader education community
- Creating written and video-based materials aimed at faculty on a variety of instructional techniques, including clickers, learning goals, and assessment

**The SEI Approach to Course Transformation**

Our model of course transformation is based on a model of “backwards design” (Wiggins & McTighe, 2005). In order to successfully align instruction and assessment it is critical to first define the goals of the course. Or, colloquially to answer the questions, “Where are you starting, where are you going to, and how are you going to get there?” Figure 1 below shows the three central questions directing all SEI course transformation efforts:
The STF collaborates heavily with faculty during each step of the course transformation, taking on many of the labor-intensive duties that teaching faculty do not have the available time or expertise to attend to. Course transformations take at least two semesters—one semester for a planning period and one for the course transformation itself. Ideally, a third or fourth semester is available for making adjustments and refinements based on the implementation and assessment results.

What should students learn? Identifying course learning goals

In the semester prior to a course transformation, the STF coordinates meetings of a volunteer working group of faculty to discuss the learning goals of the course in question. Learning goals are defined as what students should be able to do at the end of a course. For example, a goal in a journalism class might be that “students should be able to describe journalism's role in a democratic society.” These goals frame what the course is about and the big messages that an instructor wants her students to walk away with.
In addition to these “big messages,” or course-level learning goals, the faculty working groups also develop topic-level learning goals. These goals are more fine-grained and define what students should learn during a few lectures rather than the course as a whole. For example, a topic-level learning goal in a chemistry course might be “students should be able to predict relative melting points in a series of related compounds, and explain their reasoning.”

We typically suggest roughly 10 learning goals for the course as a whole, and two to three for any individual topic. Also, while it’s ideal if the faculty agree on the list as a whole, we are satisfied if they agree on 75 percent of the list, allowing for individual focus areas in a given course.

All learning goals should be sufficiently operationalized so that they are testable. If a goal is too general or vague, then it is unclear how students would demonstrate that they had successfully achieved that goal, and the goal is of limited utility and needs to be revised. For instance, “Students should understand ...” is a commonly-used phrase, but it is very imprecise. Two faculty members could have very different ideas of what “understand” means in the context of the course. The goal should be rewritten to operationalize the goals—that is, what will the students be able to do if they “understand” the topic or concept at the desired level.

Learning goals are valuable because they allow the faculty to more easily target instruction and assessment towards pre-defined goals, and they enable communication about course expectations to both students and fellow faculty.

**What are students learning? Research on learning as part of teaching.**

It is important to identify student prior knowledge on a topic, as well as common student difficulties, in order to properly target instruction. Thus, our course transformation model places a heavy emphasis on gathering data on student learning prior to, and during, a course transformation. This can take a variety of forms, including searching the research literature, observing students in class discussions, keeping field notes during classroom observations, reading homework and exams, interviewing students as they work through problems or answer questions about the content, and administering conceptual surveys. As with the development of learning goals, some of this work occurs in the semester prior to a course transformation.

In addition, it is important for faculty to know whether or not they ultimately achieve their goals—to know what aspects of the course are working well and what aspects still require work. Thus, for many courses, the STF and faculty engage in developing a new, research-based assessment to measure student learning for the course’s most central, typically
conceptual, learning goals. These assessments are designed with significant faculty input and
with insights on student thinking and difficulties from observations and interviews and then
validated using student interviews and faculty/expert responses.

What instructional approaches improve student learning?

Armed with information on what faculty think students should learn and what students are
actually learning, it is now time to address any gaps with research-based instructional
techniques. There are many ways to go about this work—the most important consideration is
that the course materials be aligned with the learning goals and the results of research into
student difficulties. STFs act as education experts in this work, suggesting and developing
instructional materials based on known techniques such as clickers with peer instruction
worksheet tutorials, Just in Time Teaching (http://serc.carleton.edu/introgeo/justintime/
index.html), case-based learning (http://serc.carleton.edu/introgeo/icbl/index.html),
preparation for future learning (Schwartz & Martin, 2004), and more.

Thus, to summarize, below we list the various outcomes from the typical course
transformation in the program:

- **Course learning goals:** Developed with faculty working group, for course as a whole
  and individual topics
- **Student difficulties:** Review literature, observe and interview students, create
diagnostic tests
- **Conceptual assessments:** Develop research-based conceptual surveys based on
learning goals to test student learning
- **Improved teaching methods:** Target student difficulties with instructional techniques
consistent with the research on how people learn
- **Archived materials:** Provide materials in organized online repository
- **Plan for sustainability:** Establish departmental structure and plan teaching
assignments to ensure ongoing use

We refer the interested reader to the Resources section of this paper, as well as our website
(http://colorado.edu/sei) for additional information about aspects of our course
transformation model.

In the rest of this paper, we will give a detailed example of SEI work in the physics
department.

**In Action: Transforming the Junior Level in Physics**

In order to provide a concrete example of how a department-based change initiative might
look in practice, we outline the results of an SEI course transformation in physics. The
physics department at the University of Colorado Boulder already had a history of
implementing interactive teaching methods and educational research in the introductory course sequence. However, investigation into student learning for upper-level physics majors had received little attention either in the department locally or in the physics education research field on a national level. There are many open questions as to how to best educate physics majors to achieve the skills and habits of mind required of a future physicist.

Thus, the physics faculty—in their departmental proposal—chose to focus their SEI efforts on the transformation of several core courses for majors: Sophomore Math Methods and Classical Mechanics, Junior Electricity and Magnetism (two-semester sequence) and Junior Quantum Mechanics. Here we describe the work in Junior Electricity and Magnetism I (E&M1: the first semester of the sequence). Readers interested in additional detail can reference our primary papers on this work (Chasteen, Pollock, Pepper, & Perkins, 2012a; Chasteen, Pollock, Pepper, & Perkins, 2012b; Chasteen, Pepper, Caballero, Pollock, & Perkins, 2012).

Why E&M1?

E&M1 is seen—by students and faculty—as one of the core courses that define the physics major. Typically taken in the junior year, this is one of the courses in which students learn to use more sophisticated calculational tools (e.g., vector calculus) and must solve problems that are increasingly lengthy and complex. Many students have reported that this course represented a milestone in establishing their identity as a physics major and helped them realize what the field of study was about. Faculty are also highly invested in this course and its outcomes and indicated an interest in modifying the course. Thus, E&M1 was a natural choice for the physics department to begin their efforts on course transformation.

As previously mentioned, the physics department had successfully implemented interactive techniques in the lower-division. Freshman physics courses used two research-tested learning strategies: (1) Peer Instruction (Mazur, 1997), where students are asked a challenging question, required to debate it with their neighbors, and then vote using an electronic response system (“clickers”), and (2) the replacement of traditional laboratory recitations with small group activities (Tutorials in Physics) (McDermott, Schaffer, & the Physics Education Group at the University of Washington, 2002) where students confront common misconceptions about the content.

Both of these techniques have been shown to increase student learning (Fagen, Crouch, & Mazur, 2002). Despite these new teaching approaches in the introductory courses, the upper-division courses continued to be taught, for the most part, in a traditional “chalk-and-talk” lecture mode, with the instructor deriving equations and solving example problems on the board. Some faculty thought that, by this point in their studies, interactive teaching methods
would overly “coddle” the undergraduate majors and that majors needed learn to learn effectively from lecture if they were to be successful in their undergraduate career. Part of the SEI project in physics was to investigate this as a research question: What instructional techniques can best help physics majors master the challenging cognitive tasks of expert physicists?

**Overview of Transformation**

Over the course of five years, the departmental STFs and physics faculty laid the groundwork for the transformations, generated instructional materials, implemented them in the classroom, and investigated the impact of the transformations on student learning. The STF who helped initiate this work is a co-author on this current paper (SVC). The outcomes of this work are listed below in Table 2:

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>Literature search, classroom observations, and student interviews</td>
</tr>
<tr>
<td>Learning goals</td>
<td>Developed explicit course goals in collaboration with faculty</td>
</tr>
<tr>
<td>Student difficulties</td>
<td>Based on ongoing research, identified common student difficulties with content</td>
</tr>
<tr>
<td>Clicker Questions</td>
<td>Developed an extensive repository of clicker questions for use with Peer Instruction</td>
</tr>
<tr>
<td>Interactive lecture elements</td>
<td>Lectures were mostly traditional, but included interactive techniques such as questioning, kinesthetic activities, whiteboards, simulations, and clicker questions (above)</td>
</tr>
<tr>
<td>Modified homework</td>
<td>Homework was redesigned to require students to make sense of their answer or use important physicists’ tools.</td>
</tr>
<tr>
<td>Tutorials</td>
<td>Weekly group tutorials were developed and refined over the course of the project. Originally optional and out-of-class, they are currently being reconfigured for modular, in-class use.</td>
</tr>
<tr>
<td>Homework help sessions</td>
<td>Instructor office hours (which are typically one-on-one, or with a few students) were redesigned as interactive group problem solving sessions</td>
</tr>
<tr>
<td>Conceptual assessment</td>
<td>In order to test whether students achieved the learning goals, a conceptual assessment was developed and validated (Fagen, Crouch, &amp; Mazur, 2002).</td>
</tr>
<tr>
<td>Online course archive</td>
<td>All materials were thoughtfully organized and made available on the web (The Science Education Initiative at the University of Colorado, n.d.a).</td>
</tr>
<tr>
<td>Publication s on this work</td>
<td>Throughout the project, data was collected on student demographics and outcomes as well as instructor use of the materials, resulting in a substantial body of scholarly work. (The Science Education Initiative at the University of Colorado, n.d.b).</td>
</tr>
</tbody>
</table>

**Table 2: Elements of the Course Transformation for E&M1 in Physics**
The transformed course has been taught seven times (at CU) to date. The first time it was taught by a faculty member whose research specialty is physics education research (PER faculty). This faculty member had experience with interactive teaching techniques and authored many of the new course materials required for the transformed course. The following semester it was co-taught by a different PER faculty member and a faculty member whose research was in a traditional area of physics and who was open to integrating in research-based teaching approaches. This same non-PER faculty member was purposefully scheduled to teach the course the following semester, solo, and given complete freedom over which aspects of the course approach to retain, if any. This sequencing of instructors was intended to enhance the sustainability of the reforms, as co-teaching has been shown to be an important element in instructional transformation (Henderson, Beach, & Famiano, 2009). After these three initial semesters, faculty rotated through the course as normal, without any intentional instructional selection by the department leadership.

✔ “What should students learn?”

To begin, faculty needed to articulate what they wanted students to get out of the course by creating operationalized learning goals. Thus, the semester prior to the transformation, the STF formed a working group of 8-10 faculty which met biweekly. The STF facilitated the meetings and summarized the results, framing the meeting as seeking faculty guidance rather than proselytizing about education. With some exceptions, these faculty meetings have been very successful in the physics department. In other departments, meetings have been less successful—either poorly attended or vetoed after the first such gathering. We note that, in physics, courses are taught by different faculty each term, creating a broader base of faculty vested in each course. The format of these meetings in physics, including pitfalls and elements of success, are summarized in another publication (Pepper, Chasteen, Pollock, & Perkins, 2011). In cases where the working group meeting format has failed, STFs have used alternative approaches, such as individual faculty interviews, to facilitate development of well-defined, course learning goals.

The working group meetings resulted in a set of 10 consensus course-level goals (and another more loosely-agreed-upon set of topic-level goals), such as “students should be able to sketch the physical parameters of a problem (e.g., electric field, charge distribution).” These learning goals then guided the course transformation efforts.
<table>
<thead>
<tr>
<th>Do...</th>
<th>Don’t...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encourage broad participation, inviting the entire faculty and targeting individual faculty members</td>
<td>Rely on mass emails alone</td>
</tr>
<tr>
<td>Distribute a clear agenda</td>
<td>Follow the agenda too closely</td>
</tr>
<tr>
<td>Choose a topic that will motivate faculty to attend</td>
<td>Call a general meeting without a topic of broad or urgent interest</td>
</tr>
<tr>
<td>Designate a knowledgeable facilitator who can guide and synthesize discussion</td>
<td>Hold an unfacilitated discussion, or one with a facilitator who is focused on expressing their own opinion</td>
</tr>
<tr>
<td>Approach discussions in the spirit of soliciting faculty guidance and input</td>
<td>Proselytize about education</td>
</tr>
<tr>
<td>Discuss course objectives and pedagogical issues</td>
<td>Create the impression you are telling faculty how to teach</td>
</tr>
<tr>
<td>Send out summaries of meeting accomplishments</td>
<td>Assume faculty will remember or recognize the progress made</td>
</tr>
<tr>
<td>Hold several meetings</td>
<td>Rely on one meeting</td>
</tr>
<tr>
<td>Synthesize meeting results and produce working documents for discussion in the next meeting</td>
<td>Expect faculty to do homework</td>
</tr>
<tr>
<td>Follow up with faculty about how their input has been used</td>
<td>Move ahead with the project, without letting faculty know the outcomes of their investment of time</td>
</tr>
</tbody>
</table>

Table 3: Elements of a more successful faculty meeting, from SEI experience (Pepper, Chasteen, Pollock & Perkins 2011):

✔ “What are students learning?”

The STFs used a great variety of data sources to gain insight into student thinking about the course content:

- Classroom observations
- Student work (homework, tutorials)
- Student interviews, typically in “think-aloud” format as they worked through problems or questions
- Instructor interviews on their use and experience with the course materials
- Attitude surveys
- Traditional exams
- End-of-course conceptual assessment

One important outcome of this aspect of the project was a body of work on student difficulties with the content, including the generation of published papers on common student difficulties with several problem-solving techniques (Wilcox, Caballero, Pepper, & Pollock, 2012) and mathematics (Wilcox, Caballero, Pepper, & Pollock, 2012; Pepper, Chasteen, Pollock, & Perkins, 2012; Wilcox, Caballero, Rehn, & Pollock, 2013; Caballero,
Wilcox, Pepper, & Pollock, 2012) in the course. This foundational work contributed substantially to an area that had not been previously well-researched and informed the subsequent instruction, which was designed to address these difficulties.

In collaboration with the faculty, the STFs also led the development of a conceptual assessment to serve as a post-course diagnostic of student achievement of the faculty’s learning goals. Traditional exams typically assess student skills in calculation and problem solving but do not address the more conceptual learning goals identified as important by the physics faculty members. Additionally, it would be problematic to give the same course exam questions each year, as these questions might be shared among students. Thus, a new conceptual assessment was developed to address this gap. Questions were created and refined with the assistance of the faculty working group and validated and improved through classroom use, student interviews, and faculty feedback. In order to test higher-order cognitive skills, the assessment was developed as an open-ended exam, allowing the STFs and faculty to examine student reasoning, sketches, and predictions, without the influence of pre-defined answer choices. Using a detailed scoring rubric, inter-rater reliability on the scoring of this assessment was high, allowing accurate determination of a student’s score on the exam as a whole to within five percent.

The final instrument is a 17-question test asking student to choose an appropriate problem-solving method for a given situation (and defend that choice), sketch and graph, and explain the physical underpinnings of steps in common calculations (see Figure 2). Currently, the department is transforming this assessment into multiple-choice and testing the effects of this change, with positive results (Chasteen et al., 2012).

DO NOT SOLVE the problem, we just want to know:

- The general strategy
- Why you chose that method

A grounded conducting plane with a point charge \( Q \) at a distance \( a \). Find \( E \) (or \( V \)) at point \( P \).

*Figure 2. Example CUE question*

This assessment has been extremely valuable in the course transformations to (a) determine how the transformed instructional approach compares to traditional instruction, (b) compare
different iterations of the transformed course to one another, and (c) further identify common student difficulties. More about this assessment can be found in another publication (Fagen, Crouch, & Mazur, 2002).

✔ “Which instructional approaches improve student learning?”

As outlined in Table 2, above, the class transformation included integration of a variety of instructional techniques, often borrowed from the department’s experience at the introductory level. In-class activities (clicker questions, interactive questioning, simulations, demonstrations, whiteboard activities and kinesthetic activities) and out-of-class activities (tutorials, homework, and problem-solving sessions) were geared to provide a variety of cognitive benefits: (a) formative assessment, (b) timely feedback to students and instructors, (c) multiple feedback channels, (d) soliciting student ideas, (e) connecting abstract physics to real-world phenomena, (f) providing instructors a window into student thinking, (g) practicing expert cognitive processes, (h) necessitating and modeling reflection, and (i) attending to student motivation. In other words, a unifying theme among course approaches was to provide students greater opportunities to reflect on their thinking and to communicate that thinking to the instructor and to each other. Here we provide some additional information about these techniques.

Clicker questions

Approximately 2-4 conceptual questions were asked in a 50-minute lecture. These questions, and the accompanying peer instruction (Mazur, 1997), allowed students to discuss and debate challenging, high-level ideas (Figure 3). The instructor then facilitated a whole-class discussion (Figure 4), focusing on articulation of reasoning and scaffolding productive argumentation. Clicker questions might include steps in derivations, application of an idea to a new context, or making sense of the physics.

Instructors and students alike were very positive about the use of clickers in this upper-level class. As one instructor noted, “Doing the concept test questions, it lets me... listen to the average student... and it lets me focus my attention much more on them... it’s let me have really more communication with the class, to have a better sense of what’s going on with them that I would have.” In a survey of 234 students across several upper-division courses using clickers, 82 percent indicated that they preferred lecture using clickers and 77 percent recommended using clickers in other upper-division courses. Additionally, most students (in an analyzed subset of that survey) indicated that clickers allowed them to be active in the class or helped them improve mastery of the material (Perkins & Turpen, 2010). On end-of-term surveys in E&M1, students consistently rated clickers and tutorials as the instructional techniques most useful for their learning (on end-of-term surveys). We note that the questions
themselves do not fully support student learning on their own, but the pedagogy of peer instruction provides several benefits (by leveraging the social nature of learning and making reasoning visible), which assist students in achieving the learning goals for the course. “[Clicker questions] give me a chance to talk over any questions I have with my classmates and work through the problem,” said one student.

Additional discussion about upper division clicker use within physics can be found in prior publications (Pollock, Chasteen, Dubson, & Perkins, 2010; Perkins & Turpen, 2010). The SEI has also produced several videos to demonstrate to other instructors how clickers and peer instruction may be used at the upper division: “Upper Division Clickers in Action,” “Types of Questions in Upper Division,” and “Writing Questions for Upper Division.” These are all available at http://STEMvideos.colorado.edu.

Figure 3. Students discuss a clicker question during Peer Instruction.
Figure 4. The instructor facilitates a class discussion around the question.

Tutorials

Tutorials consisted of a conceptually focused worksheet intended to be completed in groups of 3-5 in an hour-long weekly session (see Figure 5, below). Tutorial attendance was optional, with about 40 percent of the class attending on average. An instructor and undergraduate learning assistant (Teaching with Learning Assistants, http://serc.carleton.edu/sp/library/learning_assistants/index.html) attended to facilitate, asking Socratic questions to help students work through the material. Tutorials were intended to help students develop metacognitive strategies and to practice communicating their thinking, to enable the instructor to model such strategies in a more interactive environment, and to provide instructors with an in-depth view of student thinking and struggles. With respect to content, tutorials were designed to provide students with a deeper conceptual understanding of material and scaffold their development as expert problem solvers (by, for example, breaking problems into smaller pieces or showing multiple ways that a problem might be solved).

Students rated the tutorials highly in terms of their overall enjoyment, as well as their effectiveness in encouraging interest, intellectual challenge, and amount learned. Instructors also commented that the tutorials were a positive experience for them.

Additionally, we investigated the impact of tutorials on student learning (using the score on the conceptual assessment as a measure). Due to the optional nature of tutorials, we used multiple regression analysis to subtract the effects due to student background variables that
would tend to co-vary with student motivation. We found that even when GPA in prior courses was taken into account, attendance at tutorials was a significant predictor of student performance on the conceptual assessment.

Later work on the tutorials has included the addition of a pretest (to enable pre-/post-testing on specific content) and an in-progress experiment to reduce the length of the tutorials and incorporate them as an in-class activity for all students.

Figure 5. Students collaborate on a group tutorial with the aid of a large whiteboard.

Homework and homework help sessions

Traditional homework problems used in this course are calculation-based and drawn from a popular textbook, which, unfortunately, has many of the answers posted online. Homework questions were redesigned from external sources or revised from traditional sources to provide intellectual challenges that would engage students in active thinking processes better aligned with the course goals (see Figure 6 for an example). In particular, homework was
geared to require students to connect abstract problems to real-world contexts, articulate their expectations for an answer prior to calculating it, make sense of the final answer, and use common physicists' tools such as approximations and expansions.

Consider a field $E^r = c \frac{\hat{r}}{r^2}$

(a) **Sketch it.**
(b) Calculate the divergence and curl of this E-field. Test your answers by using the divergence theorem and Stoke’s theorem.
(c) Is there a delta function at the origin like there was for a point charge, or not?
(d) What are the units of $c$?
(e) What charge distribution would you need to produce an E field like this? Describe it in words as well as formulas. Is it physically realizable?

**Figure 6.** Redesigned homework problem. Non-traditional components are in bold.

Additionally, traditional office hours were replaced with twice-weekly homework help sessions, where instructors facilitated student work on homework in groups.

**Did it work? Impact and outcomes**

We have described the intention of the course materials but have not yet fully-substantiated their impact on students. For the E&M course transformation, the primary metric of success was the conceptual assessment described earlier. This assessment has been given to a total of 499 students at the end of nine lecture-based courses and nine courses using the transformed materials created with SEI funding at CU and elsewhere. Additionally, assessment results were collected from 16 graduate students in physics to represent an expert sample. The results are shown below in Figure 7. Courses using traditional lecture-based instruction achieved a lower average score on the assessment (45% ± 3.8%) compared to those using the transformed course materials (56%±4.1%). This difference is both statistically and pedagogically significant. Thus, the course transformations achieved at least some measure of success in improving student learning of the faculty-consensus learning goals.
Students were also very positive about the course transformations, reported that the majority of course elements were useful for learning, and felt that the course elements were well-connected. Students also reported spending more time on the transformed course than the traditionally taught courses, another positive indication of student engagement. Said one student, “I especially liked the tutorials, clicker questions, and help sessions—I wish every physics class was run like this one.”

Another measure of success was that of sustainability and transferability of the materials. The consistent scores in the assessment—across instructor and institution—suggest that the materials are able to consistently support student learning despite variability of instructional approach and context. Additionally, many subsequent instructors at CU, as well as instructors at external institutions, chose to use the materials. Indeed, multiple instructors from around the country request access to the materials every semester, suggesting that this project filled a national need. This transferability is facilitated by the course documentation and organized archive. “[These materials] allow the interested person to start teaching a transformed course without the huge time investment that it might otherwise have required,” said one instructor. On the other hand, use of the materials in more recent years at CU has been variable. Thus, our “if you build it, they will come” model of providing course materials may not be as effective as working personally with each instructor to encourage and assist them in using these techniques.
Of course, there is variability in how instructors implement these materials. There is some evidence (Chasteen, Pepper, Pollock, & Perkins, 2012c) that students score higher on the conceptual assessment when their instructor adheres more faithfully to the course and pedagogical approaches as envisioned (using the bulk of the materials in an exemplary way).

**Lessons Learned and Recommendations**

**Impact on faculty:**

Instructors of the E&M1 course, as well as those in the faculty learning group, had the opportunity to engage in conversations about teaching and learning, develop and test new instructional materials, learn about how their students were learning the material, and publish on teaching and learning. While there is some variation, these types of effects are seen across the departments funded by the SEI program. Thus, the SEI model can successfully impact faculty. Across the program as a whole, over 100 faculty have been impacted by the SEI, with over 90 having modified their instruction based on the SEI efforts. Over 50 faculty have taught a course with the support of an STF, and over 70 faculty have used or developed learning goals.

**Recommendations:**

- **Broadly involve faculty from the start; ask their opinions, and keep them updated.** This allows the faculty and STFs engaged most-closely in the course transformations to include their broader faculty’s feedback, use their expertise, and increase faculty buy-in.
- **Use an STF, or other facilitator, who is somewhat outside of the departmental structure, but still knowledgeable about the content of the discipline, to facilitate the transformation.** This enables conversations about teaching and learning that are less influenced by departmental politics, and decreases the overall burden on faculty.
- **Provide a la carte, modifiable course archives,** so that faculty can pick and choose materials that suit them, and modify them to their needs, rather than adopting a course approach wholesale.
- **Incentivize faculty involvement,** Explicit recognition and reward of faculty involvement in the department’s SEI efforts (at yearly events, or in publications for example) can facilitate faculty involvement. Also provide opportunities for implicit rewards such as observations of student engagement and learning in their courses or conversations about teaching and learning with colleagues.

**Impact on students:**

As indicated in the example of E&M1, above, the SEI model can successfully increase student learning. By supporting course transformations, faculty development, and other educational improvements, the SEI has the potential to positively affect the undergraduate education of thousands of students. After five years, the SEI project has impacted over 50 undergraduate
courses, with STFs working very closely with faculty on over 25 of these courses. Through dissemination of materials, the model also has potential to impact student learning at external institutions.

Recommendations:

- **Attend to transferability and fidelity of implementation of the course approaches —its goals, pedagogy, content materials, and structures.** Careful documentation of the course approach and rationale, videos and other methods to show instructors “what it looks like,” and in-person contact with future instructors can help ensure sustainability of the course approach.
- **Create an effective learning climate in the classroom.** Even if an instructor is using high-quality instructional materials, if students do not engage in the learning process, this can mitigate the impact of the course approach. For some ideas on how to introduce students to active learning, see our “First Day Framing” activities on our Faculty Resource page (http://www.colorado.edu/sei/fac-resources/index.html).

Impact on STFs:

The science teaching fellows themselves benefit from the program, enabling them to transition into a career in discipline-based education research or science education reform through the expertise they gain and the body of published work they generate (the project has generated over 40 peer-reviewed publications to date). Examples of future career paths have included assistant professor (mostly focused on discipline-based education research), lecturer, college instructor, associate director for teaching development, senior educational strategist, education consultant, learning assistant program coordinator, and academic advisor.

Recommendations:

- **Provide adequate training** in teaching and learning as well as in conducting education research, including reading and discussion of key papers and books.
- **Provide a community** where STFs can learn from one another, sharing experiences, challenges, and advice.
- **Urge and support publication of their work:** The daily tasks of supporting a class can easily overshadow the less-pressing but important job of documenting research and outcomes.

Impact on departments:

Due to the SEI efforts, several departments are now defining more clearly what students should learn in their courses, and measuring how well these goals are being met. Additionally, curriculum and pedagogy have been modified to improve student learning. The impact on individual faculty and courses, in this way, can translate to larger changes across the department itself. By the faculties’ own account, the project has changed
departmental culture. Discussions about teaching, learning, and departmental courses are much more frequent among the faculty and are now integrated into formal departmental structures such as faculty meetings. Faculty overwhelmingly report that the SEI has had a positive impact on their department. For example, the physics department at CU is now using interactive methods for many of the upper-level courses, including E&M1. This approach appears to be gaining popularity and becoming more normative within the department, especially as students come to expect the use of interactive techniques such as clickers in upper-level courses. In physics and several other SEI departments, the SEI work sparked the submission of several grants (at least seven NSF-CCLI/TUES proposals) to continue funding STF work (National Science Foundation, 2010, February 24).

Recommendations:

• Collaborate with the departmental chair and other respected figures. If course transformation efforts are seen as endorsed and valued by the departmental leaders, there can be greater engagement on the part of faculty.
• Attend to departmental culture. Different departments have different levels of overall acceptance of interactive teaching methods or different histories of instructional reform, for example, all of which can affect the faculty’s reaction to and involvement with its SEI-related efforts and course transformations.

Limitations of the SEI model

The successes of the SEI model, including direct impact on faculty and departments, suggest that our focus on the department as the agent of change is appropriate. However, there are some limitations to the model:

1. Course materials themselves do not constitute a successful transformation. While the project has generated many materials for instructor use, without a dedicated person (such as an STF) to work with faculty on their successful and sustained implementation, these new materials may have limited impact. Thus, continued faculty support is important, but not possible with a limited funding model.
2. Disseminating course materials proves a challenge. We have struggled to identify the best mechanisms and formats for providing materials to instructors. For example, how will another instructor best be able to find what she is looking for? Should the materials be organized by the type of material (e.g., clicker questions, tutorials), or by topic? How much is too much? Will instructors want the entire package or only certain focused sections? How will they navigate to what they are looking for? If organizing by topic, is there a “standard” list of topics for that particular course? And what is the best delivery mechanism? We have experimented with providing both online navigation to materials or a simple download of a zip file.
3. Creating STF community can pose challenges. Multiple start dates, busy schedules, decentralized locations around campus, and varying disciplinary backgrounds are all potential barriers to creating community. Attention of a focused facilitator can be key.
4. Supervising STFs and managing department SEI activities can be a challenge. While faculty serving as the SEI departmental directors are all dedicated to the goals of the program, they bring varying expertise in research-based teaching practices, education research, and education literature, and, thus, vary in their ability to provide direction and support to STFs in their department. Additionally, the influence of the SEI departmental director within the department as well as the influence and support of the chair impact broader faculty participation in and perception of the efforts.

5. Departmental culture affects the project success. For example, course teaching assignments vary across departments, such that a course may be taught by the same instructor every semester or by different instructors. Additionally, we have seen that departments differ in the faculty autonomy in teaching, departmental cohesiveness, how much student education is perceived as the responsibility of the department as a whole, and the prior use and acceptance of research-based teaching techniques. Each of these present potential challenges to the implementation of course reform.

**Getting Started: Helpful Resources and Readings**

You may find many helpful links to resources on various instructional techniques on our Faculty Resource page: [http://www.colorado.edu/sei/fac-resources/index.html](http://www.colorado.edu/sei/fac-resources/index.html).

Here we highlight some good starting resources related to the topics discussed above.

**The SEI Change Model**


Plus find course archives from all SEI departments on our Course Archive site at [http://www.colorado.edu/sei/fac-resources/course-archives.htm](http://www.colorado.edu/sei/fac-resources/course-archives.htm)

**Learning Goals**

‘At the end of my course, students should be able to ...’: The Benefits of Creating and Using Effective Learning Goals, M. K. Smith and K. K. Perkins, *Microbiology Australia*, Mar. 2010. This paper serves as a good starting point for faculty interested in developing their own learning goals.

See also our Learning Goals page: http://www.colorado.edu/sei/fac-resources/learn_goals.htm.

Research-based Instructional Practices

*Scientific Teaching* by Jo Handelsman, Sarah Miller, and Christine Pfund. A comprehensive guide targeted to faculty, covering the most important results of education research and common research-based instructional practices.


SEI’s collection of white papers on teaching and learning:

- Basic Instructor Habits to Keep Students Engaged
- What All Instructors Should Know
- What Not To Do: Practices that should be avoided when implementing active learning
- First Day of Class
- Group Work in Educational Settings
- Teaching Expert Thinking
- Succeeding at Learning 101 (aimed at students)
- Advice from Senior Undergraduates (aimed at students)

All available from our Faculty Resource Page: http://www.colorado.edu/sei/fac-resources/index.html.

The Science Education Resource Center (SERC) and the PER User’s Guide both introduce a wide variety of research-based instructional techniques: http://serc.carleton.edu/introgeo/instructionalmethod.html#teaching

http://perusersguide.org

Assessments

SEI-produced white papers on assessments,

- Assessments that Support Student Learning
- Promoting Course Alignment: Developing a systematic approach to question development
- Preclass-Reading Assignments: Why they may be the most important homework for your students

All available from our Faculty Resource Page: http://www.colorado.edu/sei/fac-resources/index.html.

This paper serves as a guide for faculty interested in the process of creating a valid conceptual assessment instrument similar to that used here to measure E&M learning.

**Research on How People Learn (general)**


**Institutional and Faculty Change:**


**Acknowledgements**

We would like to acknowledge the work of Dr. Steven Pollock, who helped develop many of the earlier course materials, and has been a valuable collaborator in research and instruction throughout the project. We also acknowledge Bethany Wilcox (CU) for her recent work on tutorials and the conceptual assessment, including the most recent data points in Figure 5, as well as Rachel Pepper (currently at University of California at Berkeley) for work in validating the conceptual assessment. We also acknowledge the generous contributions of several CU faculty members in the working group, as well as instructors of the transformed courses who created and modified materials (S. Pollock, M. Dubson, E. Kinney).

**References**


The Science Education Initiative at the University of Colorado. (n.d.a). Retrieved from [http://colorado.edu/sei](http://colorado.edu/sei)

The Science Education Initiative at the University of Colorado. (n.d.b). Resources. Retrieved from [http://www.colorado.edu/sei/fac-resources/](http://www.colorado.edu/sei/fac-resources/)


8. Process Oriented Guided Inquiry Learning

A recent article in the New York Times (Dreifus, 2013) included suggestions from a wide array of individuals—from Nobel Laureates to an elementary school student—on how to improve math and science education in the United States. Perhaps the most apt was provided by two high school students who said, “I’d like more hands-on projects where I would learn something about what I’m doing instead of just memorizing things” and “One of the problems I have during math class is not understanding the reasoning behind what we are doing. The teacher will put something on the board and say, ‘This is how you do it,’ and I’m thinking, ‘Why does that make sense?’ The teachers are sometimes reluctant to explain it because they think we won’t understand. But if something doesn’t make sense to me, I can’t do it. I’d rather understand than just memorize formulas.”

These students recognize that their learning would be improved by being engaged and thinking about the content rather than being told what to know and do. Over the past several decades, research in cognitive learning theory and classroom research have combined to confirm that students experience improved learning when they are actively engaged, interacting with their peers, and involved in the construction of their own knowledge and
understanding (see, for example, Bransford et al., 1999 and Lawson, 1999). Based on this research on how people learn, the “views of how effective learning proceeds have shifted from the benefits of diligent drill and practice to focus on students’ understanding and application of knowledge” (Bransford et al., 1999). Based on these ideas, and incorporating the well-developed principles and techniques of cooperative learning described by Johnson, Johnson, & Smith (1991), Process Oriented Guided Inquiry Learning (POGIL) is a student-centered pedagogic strategy in which students are guided to construct their own understanding through the use of small groups while also focusing on the development of important learning skills.

**What is POGIL?**

In a POGIL learning environment, students work cooperatively in self-managed small groups of three or four. The group work is focused on activities that are carefully designed and scaffolded to enable students to develop important concepts or to deepen and refine their understanding of those ideas or concepts. In addition, the learning environment is structured to support the development of process skills—the important learning skills and interpersonal skills that will promote life-long learning and be of great value in the workplace and in life. The instructor’s role is to facilitate the development of student concepts and process skills, not to simply deliver content to the students.

To get a sense of the difference between a typical, traditional classroom and one that is using POGIL pedagogy, we will contrast the way in which a fundamental topic in chemistry is approached. In many chemistry (and other science and mathematics) classes, at both the high school and college levels, the vast majority of time is spent with teacher-talk; the description that follows is representative of this type of pedagogy. The instructor enters the room, possibly makes some introductory remarks, and then addresses the class as a whole. The students prepare to take notes, based on some combination of what the instructor says and what she writes on the board. In some cases, PowerPoint slides are used, and the students may have printed copies of these, on which they may (or may not) make additional notations. The instructor begins by noting the topics that will be addressed during that class meeting: in this case, those topics include the components of the atom, the definition of some important terms, and the relationship of the numbers of subatomic particles to information on the Periodic Table. The lesson then begins. Every atom contains protons, electrons, and neutrons, with charges of +1, -1, and 0 respectively. The number of protons in an atom is known as the atomic number, and this is a characteristic of each element. This number can be found for each element on the Periodic Table; it is typically located just below the atomic symbol for each element. For example, for carbon (C) the atomic number is 6 and for hydrogen (H) the atomic number is 1. The nucleus of the atom contains both protons and neutrons, and these have
roughly the same mass, which is about 2000 times as large as the mass on an electron. The **mass number** is the sum of the number of protons and neutrons. In most cases, the atoms of a given element do not all have the same mass number; there are different types of atoms that vary in the number of neutrons that are present. Atoms that have the same atomic number (same number of protons) but different mass numbers (different numbers of neutrons) are referred to as isotopes. One or more examples of isotopes are given along with a determination of the numbers of protons and neutrons in each based on the designated mass number. If the number of protons and electrons present is not the same, an **ion** is formed. Ions with more protons than electrons are positively charged, and those with an excess of electrons are negatively charged, with the magnitude of the charge being equal to the number of excess protons or electrons. Examples of ions are then provided.

The POGIL approach to this content is very different. Students are seated in groups of four, and they have been assigned roles such as Manager, Recorder, Presenter, and Reflector. Each role brings with it specific responsibilities. The roles that are assigned may vary from day to day, depending on the nature of the activity, the instructor’s preferences, and the process-skill goals for that day.

In the author’s general chemistry classroom, the four roles mentioned above are typical. The Manager is responsible for keeping the group together as the students work through the activity. (Figure 1 is a reproduction of a POGIL activity for the content addressed above; the structure of this activity will be discussed below.) Even though there are a series of numbered questions to be addressed, the students are NOT to split up the questions among themselves (“you do #1, you do #2, I’ll do #3,” etc.). Rather the group considers each question together and reaches a consensus based on the information that is provided and the necessary analysis and reasoning. Each student then writes down that answer and explanation on their own activity sheet. The Manager keeps the group working cooperatively as they consider the questions. In addition, the Manager is the only member of the group who asks questions on behalf of the group. That is, if another group member has a question that he would like to ask the instructor, that question must be communicated to the Manager. If the group is unable to answer the question on their own, then the Manager raises his or her hand to get the instructor’s attention and asks the question on behalf of the entire group. This assures that the group considers each question on their own before asking for help and also reduces the number of people asking questions, making classroom management easier for the instructor. The Manager is also responsible for making sure that everyone in the group is engaged and contributing to the overall discussion.

The Recorder is responsible for recording the official responses for the group. Some instructors ask the Recorder to write the group consensus answers to the questions on the activity sheet onto a “Recorder’s Report” which is collected from each group at the end of the
class period. In the author’s classroom, the Recorder’s Report does NOT include the answers to the activity questions; rather, the Recorder is responsible for noting on the Recorder’s Report all of the important concepts that are developed during that class period, along with anything else that the instructor explicitly asks to be included. These reports are collected from each group, comments are written on them, and then a photocopy is made for each member of the group. These copies are then returned to the group at the beginning of the next class meeting. If the Recorder has written all of the key concepts on the Report, each group member then has a list of those important ideas, which provides an excellent resource for reviewing for exams.

The Presenter is the person who will provide reports to the rest of the class from the group. These reports may be oral, or the Presenter may be asked to put the group’s answer on a group whiteboard for sharing, on a blackboard or whiteboard in the front of the classroom, or on a transparency or sheet of paper for overhead projection. One of the important responsibilities of the Presenter is to be able to articulate not only what the group’s answer to a particular question is, but also how the group arrived at that answer. In a POGIL classroom, being able to generate correct answers is important, but it is not more important than being able to explain how to arrive at those answers.

The Reflector provides insight to the group about how it is functioning as a group—what the group is doing well, what needs improvement, and how that improvement might be achieved. The Reflector might be asked to provide feedback to the group privately at regular intervals (every 20 minutes, at the end of the class period), and this feedback might be general, or the instructor may indicate a specific aspect of group dynamics on which to focus. An important aspect of effective reports by the Reflector is to frame any comments—particularly improvements—in terms of the group rather than an individual. For example, if one student is dominating the discussion, the Reflector might be tempted to say, “An improvement would be for Rick to talk less,” but this is a comment about Rick and not about the group. A better reflection would be “An improvement would be if we could find a way for everyone to contribute equally to the conversations that we have. Perhaps the Manager could designate a different person to make the first comment about each question.”

One hallmark of the POGIL approach to instruction is that it is based on a set of guiding principles (primarily those of constructivism, inquiry, cooperative learning, and a focus on development of process skills) and is not substantially prescriptive. In fact, every POGIL classroom is different and is a reflection of the uniqueness of the particular context: the institution, department, physical space, student body, and instructor. However, there are four core characteristics that must be present in order for a classroom environment to be considered a POGIL implementation (The POGIL Project, 2013):
1. Students are expected to work collaboratively, generally in groups of three or four.
2. The activities that the students use are POGIL activities, specifically designed for POGIL implementation.
3. The students work on the activity during class time with a facilitator present.
4. The dominant mode of instruction is not lecture or instructor-centered; the instructor serves predominantly as a facilitator of student learning.

In addition, there are some common attributes of many POGIL classroom implementations and facilitation strategies that, when combined with the required characteristics above, provide a good starting point for any POGIL implementation:
5. Students have assigned roles within their groups.
6. The activity is designed to be the first introduction to the topic or specific content.
7. The students are not expected to have worked on any part of the activity prior to class meeting time.
8. Groups are expected to complete all of the Critical Thinking Questions (or equivalently designated questions) during class. There may be additional exercises or problems expected to be completed outside of class.

The description of the roles and the interactions given above are one example of a POGIL classroom and one that exemplifies the characteristics of the Basic POGIL Classroom Implementation. Shadle (2010) has recently provided an excellent description of the implementation of POGIL in her general chemistry classroom, serving as a nice complement to the one described in this chapter.

A POGIL activity that introduces much of the same content as in the traditional approach described earlier is shown in Figure 1. We now examine this activity in some detail to emphasize how it provides a very different learning experience from the more traditional presentation described earlier. Note that this activity begins with a Model, in this case a set of “cartoons” showing various atoms and ions, along with a few other pieces of information. The Model provides information that students will use, along with any necessary prerequisite knowledge, to develop the specific learning goals for that activity. In general, a Model could consist of any combination of text, graphs, tables, figures, diagrams, animations, or even a demonstration; the purpose of the Model is to provide sufficient information for the students, working together and guided by the series of questions that follow, to develop the concepts related to the learning goals for the activity. The first three questions can be answered directly from information provided in the Model without any analysis. Beginning an activity with a few directed questions such as these is important to make sure that the students can properly interpret the Model and to focus the students’ attention on aspects of the Model that will be important in the analysis to come. In Question 4, students are asked to draw conclusions (in this case, look for common characteristics) based on the first three questions. Question 5 extends this idea, without the scaffolding that was provided in Questions 1–3. In Question 6, students generalize their results from the previous two questions, make a connection with the
Periodic Table, and also, in essence, define the term *atomic number* based on this analysis—rather than being given the definition from an authority. Question 7 provides an opportunity to apply this newly developed idea to a different element. This ends the first section of the activity, one that has focused on developing the concept of atomic number.

Question 8 begins the examination of a new concept, providing an opportunity for students to explore and identify the characteristics of *ions*, albeit without the scaffolding that was provided for the development of the atomic number concept. In Question 9, students apply the knowledge that they have developed up to this point in the activity. A similar pattern is present in Questions 10, 11, and 12 with respect to the concept of *isotopes* and the definition of *mass number*. (Note that although examples of isotopes are given in the Model, no definition of the term isotope is provided, just as no definitions of atomic number and mass number are provided in the activity.) Question 13 is, in many ways, unconnected to the previous content in terms of the development of key concepts. This question provides an opportunity for students to use the information that is provided in the Model to answer a question while making their analysis and reasoning explicit.

In the author’s classroom, this activity is done during the first class meeting of General Chemistry I, and the students have *not* prepared in any way for this content: they have not been assigned any reading in the textbook, there is no preliminary lecture that provides any definitions or additional information, there are no additional handouts or class notes that they are provided, and they are NOT expected (or even allowed) to use their textbooks (or the Internet) to “find the answers” to the questions. While the students are working through the activity in their groups, the instructor circulates around the room, listening to the conversations that are taking place and observing the interactions. If a question is asked by one of the Managers, the instructor’s role is to avoid answering the question directly, but rather to find a way to help the students find the answer themselves. One way to do this is to first ascertain if the underlying difficulty is one of *process* rather than *content*. That is, many times the issue is one that is related to a need to improve some learning skill, communication skill, or interpersonal skill. The following responses by an instructor are examples that have been successful in just such situations:

- “All of the information that you need to answer that question is present in the Model. Think about what information you need, and then examine the Model carefully. I am sure that you will find what you need there.”
- “Have the Presenter read the question out loud to everyone and listen to what he is saying. If you still are having difficulty after that, let me know.”
- “Let’s have each person in the group say what he or she thinks is the best answer and why. After each person has given his or her answer and reasoning, try to reach consensus.”
Of course, there are times when a more direct response is needed, but the instructor’s first inclination (and one that is difficult for many instructors to develop because it is contrary to their previous experience and training as “the expert”) is to not answer questions directly, but rather respond with a suggestion or another question. For example, a Manager may ask, “Is this answer correct?” Sometimes the group is very frustrated or uncertain and truly needs affirmation that they have a correct answer. But another approach is to respond with a question such as “What is it about your answer/your analysis/your conclusion that you think may be incorrect?” This provides them an opportunity to reflect on what it is they are unsure of, and often they find that there is nothing! This approach can help them gain confidence in their own abilities and also help them develop as lifelong learners and problem solvers.

Nevertheless, students do want to know that they are on the right track and that they have developed the important concepts correctly before they leave class each day. Thus, it is important to have some closure to each class meeting that provides them with this feedback. There are numerous ways to do this. Presenters can be asked to provide the group’s answer (and reasoning) to some (or all) of the questions; discussion can take place when different groups have different answers. Another technique is to have each group take a couple of minutes to write down the “three most important things that you have learned in class today” and then share those with the class. Although it is also possible for the instructor to provide this type of wrap-up directly, having the students do it is of greater benefit to them as they are then forced to reflect on what they have learned, decide on what is important, and articulate it in their own words.

Learning in a POGIL classroom does not end when the class session ends. Students are expected to reinforce what they have learned after the concepts have been introduced in class. This reinforcement includes reading from relevant portions of an appropriate textbook. In addition, students should work through problems and exercises to reinforce the key ideas from class and as practice for solving questions for exams. In the author’s classroom, students also prepare for a short, one-question, five-minute quiz taken individually at the beginning of every class meeting. The quiz question addresses one of the fundamental concepts that was developed in the previous class and provides feedback to each student (and to the instructor) concerning the extent to which this important material has been mastered. The groups discuss the quiz immediately after it is completed so that each student gets immediate feedback. Not only does this provide assessment of the student’s understanding, but often it is also useful as preparation for that day’s activity, which may build upon this key concept developed in the previous class meeting.
What do we mean by “Process Oriented”? 

Although content mastery is important in all STEM disciplines, the vast majority of students in STEM classes at the high school and post-secondary levels (especially at the introductory level) are unlikely to be regularly applying most (if any) of the specific content knowledge they are expected to master outside of the course—with the possible exception of using some of the content in sequent courses. This is not to diminish the importance of this content, but rather to emphasize that the development of other skills—particularly thinking, learning, communication, and interpersonal skills—are likely to be much more useful and transferable over the course of an individual’s lifetime. This idea is emphasized in the Mission Statements of colleges and universities across the country. For example, the Mission Statement of Franklin & Marshall College is to inspire in its students “a genuine and enduring love for learning, to teach them to read, write, and think critically, [and] to instill in them the capacity for both independent and collaborative action.” These are the types of skills that are known as process skills, and the POGIL pedagogic philosophy emphasizes a commitment on the part of the instructor to design a learning environment that helps promote the development of these skills in a conscious way. That is, in the context of a particular major, program, course, and daily classroom experience, the instructor should have in mind specific process skills that are being developed and design the experience to promote the development of these skills. This is in contrast with some instructors who believe, for example, that the act of taking (and successfully completing) a chemistry course will help students develop their problem-solving and critical-thinking skills, simply because (to paraphrase comments that the author has heard from many faculty members) “learning chemistry requires students to solve problems and think critically.”

There are seven process skills that a POGIL classroom environment can develop in conjunction with effective facilitation of group learning and a well-designed POGIL activity:

• Communication
• Assessment, particularly self-assessment
• Teamwork
• Management
• Information Processing
• Critical Thinking
• Problem Solving

Still, every institution, department, program, student body, and course is unique, and the instructor is in the best position to determine what the important process skills are that her students should be developing. The key point is that in a POGIL learning environment, the instructor should have one or two specific process skill goals in mind for each classroom
experience, and the experience should be designed in such a way as to facilitate the development of those skills in the context of content mastery.

Two of the process skills on the list above deserve particular attention in terms of instructional practice. The first is “problem solving.” The key idea here is what a “problem” is. A problem exists when one does not know what to do or does not have an internalized algorithm for obtaining an answer. If one knows what to do, it’s not a problem—it is an exercise. Many STEM textbooks have numerous numbered items at the end of each chapter for students to solve; regardless of how these are labeled, in most cases the vast majority of these items are exercises, not problems. That is, students in general know what to do when they encounter these items: they either have an internalized algorithm for generating a correct answer, or they know that what they need to do to generate a correct answer is to look back through the chapter for a sample item that is phrased similarly and then mimic the solution that is presented there. If one wants students to get better at “problem solving,” then students need to encounter true problems to work on. Many students find that many of the questions in a POGIL activity are problems in this sense: they don’t know exactly what to do to generate a correct answer to the question. Question 13 in the Nuclear Atom activity is an example of a question that is specifically included as a “problem.” Although many students may already “know” that most of the mass of an atom is in the nucleus, this question requires them to use information from the Model (information processing!) to reach that conclusion. This question is included in this activity for specifically this reason—to develop the problem-solving ability of the students.

Another important instructional insight deals with assessment and differentiating between assessment and evaluation. To be clear, we define these terms in the following way:

- An evaluation is an activity that is designed to measure performance against a set of standards—frequently a pre-determined set of standards.
- An assessment is an activity that is designed to improve future performance.

These terms are similar to the terms summative assessment and formative assessment that are frequently found in the education literature; here the terms evaluation and assessment are used as defined above to make explicit what is meant by those terms and to avoid confusion based on differing interpretations of summative and formative assessment.

Examples of evaluations:

- tenure decisions
- assigning points for a student’s response to a question on an exam
- determining the grade for a written lab report
- selecting the members of the varsity field hockey team

Examples of assessments
• making comments on a draft of a paper
• providing effective feedback to pre-tenure colleagues
• having students practice homework problems

Once the members of the field hockey team have been selected, the most effective coach engages only in assessment activities from that point forward if the goal is to have the team improve.

One of the important insights about the impact of assessment and evaluation has to do with the result of providing both at the same time. For example, a common practice is to provide constructive comments on a submitted lab report and also provide a grade. The author’s experience is that many students look at the grade and, if they are satisfied, stick the report in their bookbag and (possibly) never look at it again. This impression is reinforced by the presence of identical errors and issues in the next report! An insight here is that—for many, many people—when assessment feedback and evaluative feedback are provided simultaneously, the recipient focuses on the evaluation and ignores (or mostly ignores) the assessment. The implication for instruction is that if we want our students to improve at something, we need to provide assessment feedback that is independent of evaluative feedback. This desire to separate assessment from evaluation is the reason that, in many POGIL classrooms, the POGIL activity is not graded. (Some instructors assign some “points” for completeness, but in general, students are not evaluated on the “correctness” of the responses that they provide.) The purpose of the activity is to enable students to develop concepts and deepen their understanding of those concepts. Interactions among students and between students and the instructor are intentionally assessment-oriented. Making this shift in thinking to the idea that an important role for the instructor is to provide assessment opportunities for students in the classroom can bring about a paradigm shift in what takes place in the classroom and for what purpose.

The types of process skills that are developed in the context of a POGIL learning environment are also the skills that employers are seeking in college graduates. In a recent study undertaken for the American Association of Colleges & Universities, Hart Research Associates (2013) found that “employers overall are most likely to believe there is a need to increase the focus on active skills such as critical thinking, complex problem solving, communication, and applying knowledge to real-world settings.” More than eighty percent of employers surveyed indicated that undergraduate education should place more emphasis on the following three areas, all of which are targeted process skills in a POGIL learning environment:

• Critical thinking and analytical reasoning
• Complex problem solving and analysis
• Written and oral communication
Guided Inquiry and The Learning Cycle

In the science-education literature, the term “inquiry” has multiple definitions. Different levels and types of inquiry (including “guided inquiry”) have been identified (see, for example, Trout et al. (2008). As Abraham (2005) has pointed out, “Inquiry can be seen as instructional activities that range all the way from the simple use of questions to the practice of open-ended research.” In POGIL, “guided inquiry” is defined in a particular way. First, the experience is “guided” rather than “open.” This implies two important ideas: the activity is appropriately scaffolded so that students are “guided” to develop the key concepts, and the experience is not open-ended research of a given question. The learning goals are clearly defined and the outcomes and answers are known (to the instructor). The activities are structured to follow a Learning Cycle of exploration, concept invention and term introduction, and application.

The Learning Cycle is an instructional strategy that is rooted in the ideas of constructivism and a Piagetian view of development. Cracolice (2009) has provided an excellent summary of the Learning Cycle as an instructional strategy and its relationship to guided inquiry. Karplus and co-workers developed the approach as part of the Science Curriculum Improvement Study in the early 1960s (Karplus & Thier, 1967). Much research has been done over the past several decades (see, for example: Lawson, 2003; Lawson, Abraham & Renner, 1989; Abraham & Renner, 1986) to demonstrate that including all three phases of the Learning Cycle in the designated order provides the strongest overall results in terms of student learning and retention. Learning Cycle instruction has also been shown to provide other, more general benefits. For example, ninth-grade science students receiving Learning Cycle instruction outperformed students receiving traditional instruction in terms of content knowledge and formal reasoning and showed increased improvements on an IQ test (Schneider and Renner, 1980).

In the exploration phase, the learner gathers data and processes it, looking for patterns or trends from which generalizations or conclusions can be reached. In the laboratory, a question might be posed, and the students would then collect data in an effort to address this question. In a POGIL classroom activity, rather than have students collect data (which in many cases is not practical), the data are instead presented to the students in the Model. The initial questions in the POGIL activity begin the process of exploring the Model, extracting relevant information, and eventually guiding the students to identify an important trend or pattern, or to reach a key insight from the data. In the Nuclear Atom activity (Figure 1), the first three questions serve the former purpose and focus the students’ attention on the numbers of protons, electrons, and neutrons. Questions 4 and 5 lead the student to notice the pattern that all atoms and ions of a given element have the same number of protons. These
questions represent the end of the initial *exploration phase* and the transition into the second phase of the Learning Cycle.

In the *concept invention/term introduction* phase, students develop the targeted concept; ideally, this is also where the associated vocabulary (or *term*) is introduced. Note that in the Nuclear Atom activity, the *term* “atomic number” is introduced to the students in Question 6, after the related concept has been initially developed in Questions 4 and 5. The advantage of this sequence is that the term is being introduced at a point after the student has developed the concept that is being named. In this case, the term gets attached to the concept in the student’s brain—both figuratively and (hopefully) literally. This sequence is different than the more traditional approach that is taken in many textbooks and by many instructors. In the more traditional approach, the names of the concepts are provided *prior* to any examples or evidence related to the concept. Frequently, after the term is given, the definition of the term is provided: “The *atomic number* is the number of protons in the nucleus of each atom of a particular element.” Then some examples are provided to show the students that the previously provided definition is “correct.” For many students, when this approach is used, the term is introduced on a purely verbal level: the students may be able to reproduce the definition and may be able to mimic the examples that are given, but there are no conceptual underpinnings upon which to build, and the ability to reproduce the definition is not necessarily related to being able to apply the concept in other, more meaningful, ways.

The third phase of the Learning Cycle is the *application* phase in which the newly developed concept is applied in new situations. This phase provides an opportunity to strengthen the concept, show how it can be used, and/or test its generalizability. In the Nuclear Atom activity, Question 7 constitutes the application phase for the concept of atomic number and completes the cycle. In some cases, the application phase of one cycle leads into the exploration phase of the next cycle. This is not the case for the Nuclear Atom activity, although a second Learning Cycle is present in Questions 8 and 9. In this case, the concept to be developed is that of an *ion*, and there is less scaffolding than was present in the first Learning Cycle. Question 8a, building again from the exploration provided by Questions 1–3, begins this cycle, and the term *ion* is introduced in Question 8b. The combination of Questions 8b and 8c can be considered the *concept invention/term introduction* phase, with Question 9 serving as the *application*. Note that Question 12 serves as further application for the concepts of *ion* and *mass number*. As mentioned previously, Question 13 is not part of any of these Learning Cycles; it is provided to provide an opportunity for students to work on several process skills—particularly problem solving and critical thinking.
Chi (2011) has recently proposed a framework for differentiating different levels of cognitive engagement in various types of “active learning” environments and has found that the relative achievement of students in many studies can be correlated with the level of engagement. The levels of engagement are defined in terms of the observable behaviors of the students, based on the assumption that these behaviors are a sufficient proxy for the underlying cognitive processes. The levels (in order from highest level of engagement to lowest) are:

- **Interactive**: two or more students engaging with each other; examples include explaining to each other, arguing (requesting and providing justification) with each other, and responding to a question from another student
- **Constructive**: students generate some information beyond that which is presented in the learning materials; examples include posing questions, providing justifications, forming hypotheses, and comparing and contrasting
- **Active**: students are doing something with their hands (or bodies) with the learning material; examples include taking notes, copying a solution from the board, underlining important sentences in a text, and rehearsing or repeating definitions
- **Passive**: students are oriented toward or receiving instruction (“paying attention”) but they are not doing anything else that is observable; examples include listening without taking notes, watching a demonstration, studying a worked example, and reading to themselves.

Chi has shown that evidence for the ICAP hypothesis is present in many reports in the literature (Chi, 2011). For example, in a study in a materials engineering context, different groups of students experienced each of the four levels: after reading the text on their own, all students took a pretest. Then, one group read again out loud (passive); a second group read and highlighted sentences (active); students in the third group were individually given four figures and asked to construct explanations for these figures (constructive); and the fourth group was given the figure-explanation task but did it in pairs (interactive). The pretest averages for all four groups were roughly the same; the gains achieved on the posttest strongly supported the ICAP hypothesis, with the observed learning gains (defined as [Post – Pre]/[Maximum—Pre]) being 0.07 (passive), 0.24 (active), 0.42 (constructive), and 0.62 (interactive). This means that, on average, the students in the “passive” group increased the average score on the posttest by only seven percent of the maximum possible gain they could have, while the students in the “interactive” group experienced an increase of 62 percent of the maximum possible gain. In this case, both of these groups started with essentially the same initial average score on the pre-test (52.2% vs. 52.3%). Here, the “interactive” group would be more properly described as “constructively interactive” as they were not simply interacting, but were also engaged in constructive activities such as providing justifications.
and (presumably) explaining their reasoning to their partners. This is precisely the type of activity that students engage in within a POGIL learning environment: the activities are designed for students to construct their understanding of the key concepts, and they work interactively within their groups.

Chi’s ICAP framework provides a basis for understanding why POGIL classrooms have been shown to be such powerful learning environments compared to more traditional approaches. Farrell, Moog, and Spencer (1999) described a study involving general chemistry students over an eight-year period. They report that students in courses employing a POGIL instructional strategy achieved a significantly higher success rate (defined as receiving an A, B, or C in the course, as compared to a D, F or withdrawal) than students who had been taught by the same instructors in previous years using a more traditional lecture-oriented approach. (See Figure 2.) In both cases, the students were in classes of about 24 students each, and similar exams were used for both groups of students. Lewis and Lewis (2005) also reported enhanced achievement for students who experienced POGIL learning environments. In their study, the control group experienced three lectures each week. The experimental group attended two lectures each week given by the same instructor as for the control group, with the third lecture replaced by a recitation section in which students worked in small groups of three or four using POGIL activities, with a graduate student or advanced undergraduate serving as the facilitator. They refer to this approach as Peer-Led Guided Inquiry (PLGI) because it combines many of the elements of Peer-Led Team Learning (PLTL) with the group-learning structure and guided inquiry activities of POGIL. Lewis and Lewis showed that the two groups were not statistically different in terms of the average SAT or ACT scores.
As shown in Figure 3, the PLGI group outperformed the control group on all four of the course exams (multiple choice exams written by the course instructors) and on the final exam (multiple choice exam constructed by the American Chemical Society’s Exams Institute). The effect size, as measured by Cohen’s $d$, increased for each successive exam during the semester, reaching 0.563 (above 0.5 is considered a “moderate” effect) for the 4th course exam. For the final exam, Cohen’s $d$ decreased slightly, to 0.367, which is still significantly above the lower limit of 0.2 for a “small” effect.

Figure 2.

Grade Distributions in General Chemistry

Data ($n = 905$) from small (~24 students) sections of three instructors using lecture approach (1990-94) prior to implementation of POGIL pedagogy (1994-98).

Another study demonstrated the impact that a POGIL learning environment can have on student retention of knowledge. In this study (Ruder & Hunnicutt, 2008), an unannounced multiple choice “quiz” was given to students on the first day of an Organic Chemistry II course at the beginning of the second semester at a large, public university in the southeastern United States. The quiz, written by an instructor of Organic Chemistry II who had not been involved in the course the previous semester, addressed content that had been covered in the first semester and was considered important for success in Organic Chemistry II. All of the students in the study had successfully completed the Organic Chemistry I course at the same university the previous semester, with the final examination in that course having been given about one month earlier. All sections of the course were large, ranging in size from 150–250 students. One group of students had experienced traditional instruction for the first semester; the other group experienced POGIL instruction for the vast majority of their “lecture” time, even though there were several dozen groups of students in the tiered lecture hall. The outcome, shown in Figure 4, is striking. A majority of the students (60%) in the lecture section scored below 50 percent on the quiz, and none of the students achieved a score above 90 percent. Less than five percent scored above 80 percent. In contrast, fewer than a
quarter of the students from the POGIL section scored below 50 percent on this quiz, and about 30 percent of the students scored above 80 percent, with over one-fifth of the POGIL students scoring above 90 percent.

Performance on Organic Chemistry 2
Unannounced First Day Pre-Quiz

• All students passed Organic Chemistry 1 at this institution during the previous semester
• All sections of Organic Chemistry 1 had more than 150 students.

A final example of the effectiveness of the POGIL approach is a two-year study of a POGIL implementation in an introductory anatomy and physiology course (Brown, 2010). A comparison was made between the performances of students in the final semester that the course was taught using a traditional, didactic approach and the following three semesters in which POGIL was implemented. Although there was no statistically significant difference in the overall student scores for the entire course between the first POGIL implementation and the last lecture implementation, in the two subsequent semesters the POGIL students had a stronger overall performance at the $p = 0.01$ level, with the mean final score raising from 76 percent in the lecture mode to 87 percent and 89 percent in the second and third semesters of POGIL instruction. In addition, the percentage of students earning a grade of D or F fell from 16 percent in the lecture implementation to less than 6 percent in the first POGIL implementation and to 0 percent by the third semester using POGIL.


Figure 4.
percentage of students earning an A in the course rose monotonically through all three POGIL semesters. These results are typical; many instructors find that the first time that a POGIL approach is implemented, the impact on student learning is not necessarily significant — and student complaints are not unusual. However, once the instructor gains some experience in facilitating in this new environment, the types of outcomes reported by Brown, and in the other examples given here, are generally experienced: increased student achievement and longer term retention of knowledge. In addition, students often recognize the advantages that this type of instruction has for them as learners. In a survey given to several hundred organic chemistry students at six different institutions, some of whom experienced POGIL instruction while others had traditional lecture classes, students were asked to respond to the following prompt: “I would recommend the method of teaching used in this course to a student taking this course next year.” Responses were on a five-point Likert scale from Strongly Agree to Strongly Disagree. Of the 381 POGIL students, 80 percent would recommend the approach, with 57 percent strongly agreeing with the statement, and only 6 percent either disagreeing or strongly disagreeing. This compares favorably to the responses from the 307 lecture students, of whom only 49 percent would recommend the lecture teaching method (with only 26 percent strongly agreeing) and 30 percent disagreeing with the statement (Moog et al., 2009).

Given the large amount of additional evidence for the effectiveness of POGIL and other student-centered, active-learning, inquiry-based approaches to science instruction, it is perhaps surprising that the adoption of these types of approaches is not more widespread. Numerous authors have addressed this issue over the past several years. In 2004, Handelsman et al. (2004) asked the question in this way: “Why do outstanding scientists who demand rigorous proof for scientific assertions in their research continue to use and, indeed, defend on the basis of the intuition alone, teaching methods that are not the most effective?” These authors suggest that the reasons include 1) an unawareness of the results demonstrating effectiveness; 2) a distrust of the data because many scientists (including themselves) flourish within the current dominant educational paradigm; 3) concern about the challenge of learning a new approach to instruction; 4) fear that being identified “as teachers will reduce their credibility as researchers.” Cracolice (2009) echoes many of these ideas; additional reasons that he proposes (based on his interactions with high school and college instructors in professional development contexts) are a concern that the pace will be too slow to “cover” all of the needed/desired content, alternative understandings of what is actually meant by “inquiry,” and a concern that the infrastructure and number of students in a course makes use of an inquiry approach impractical. All of these barriers, both real and perceived, have been overcome by individuals who implement POGIL and other student-centered inquiry approaches in their classrooms in both college and high school settings.
Summary

POGIL is a learner-centered approach to science instruction based on research on how students learn best. Students in a POGIL classroom are engaged in constructive and interactive work, consistent with Chi’s framework for the most effective learning environments. The key components of this environment are the use of cooperative learning groups with assigned roles, promoting the interactive aspect of the experience, and the use of specially designed activities that follow a Learning Cycle structure through which students construct important concepts. As of this writing, POGIL materials are available for most undergraduate courses in calculus, anatomy and physiology, and in most chemistry content areas (including biochemistry). Activities are also available for high school chemistry and biology courses. Additional information about how to obtain these activities, and about implementing POGIL in general, is available from The POGIL Project website at http://www.pogil.org. This website also provides information about professional development workshop for high school and post-secondary instructors interested in learning more about implementing POGIL and creating POGIL activities for their classrooms.

References


9. Professional Development of Faculty: How Do We Know It Is Effective?

doi:10.7936/K7G44N61
Diane Ebert-May¹, Terry Derting² and Janet Hodder³

¹Michigan State University, ²Murray State University, ³University of Oregon

Correspondence:
Diane Ebert-May
Department of Plant Biology
Michigan State University
612 Wilson Road
East Lansing MI 48824-1312
PHONE: (517) 432-7171
FAX: (517) 353-1926
ebertymay@msu.edu
http://ebertmaylab.plantbiology.msu.edu/

Abstract

Ongoing professional development (PD) of faculty is necessary for increasing faculty knowledge and developing materials and techniques that help engage students and foster learning. While workshops are a popular format of PD, little is known about the efficacy of workshops due to the reliance on self-reported data alone. PD programs should employ mixed approaches that include both assessments such as objective and self-reported data from faculty and students, direct observation of faculty teaching by external experts, and analysis of course materials, as well as surveys of student and faculty beliefs and approaches to teaching and learning. The Faculty Institutes for Reforming Science Teaching (FIRST II) and the National Academies Summer Institutes (SI) are examples of PD evaluation discussed in detail.

Introduction

In 2006, Connolly and Millar noted that workshops are one of the most popular ways for instructors in STEM fields to learn more about teaching and student learning. Yet research shows that we often do not know what instructors actually gain from participating in a
workshop. To date, the efficacy of workshops has been evaluated principally through self-reported surveys of faculty satisfaction, learning, and examples of applications in their classrooms (Connolly & Millar, 2006). What is often not measured is to what extent workshops result in effective changes in teaching practice and ultimately, the effect of changes in teaching practices on student learning. The goal of our chapter is to address the following assertion: Evaluations of the effectiveness of models of faculty professional development (PD) programs need to include more than self-reported data alone. In practice, if institutions are serious about improving teaching, student learning, and providing faculty opportunities to engage in substantive professional training, then it is necessary to employ mixed approaches to evaluating PD models. These approaches include both objective and self-reported data from faculty and students, such as direct observation of faculty teaching by external experts, analysis of course materials (i.e., goals, objectives, assessments), and surveys of student and faculty beliefs about approaches to teaching and learning.

**Background**

Since the turn of the 21st century, the national demand for improvements in undergraduate STEM education rings loud and clear across all scientific disciplines. In biology, for example, introductory courses were described as relying on “transmission-of information” lectures and “cookbook” laboratory experiences that did not help students build conceptual understanding and scientific reasoning skills (Handelsman et al., 2004). The Vision and Change in Undergraduate Biology Education report (Brewer & Smith, 2011) responded to this claim with input from over 500 biologists who called for reform, particularly in the introductory core courses of biology curricula. In response to this report, PD opportunities for faculty continue to grow. Yet, there are still few reported studies that rigorously evaluate the impact of these opportunities on producing measureable change in faculty teaching. Henderson et al. (2011) reported, in an analysis of 191 published studies on how to promote change in instructional practice in undergraduate STEM, that only 21 percent of the articles that reported the implementation of change strategies were categorized as presenting rigorous evidence to support claims of success or failure of the strategies.

**Case**

The basis for our claim about evaluation of PD programs is derived from our research on two national professional development projects, Faculty Institutes for Reforming Science Teaching (FIRST II, Hodder/Ebert-May PIs, NSF –DUE88847) and the National Academies Summer Institutes (SI) (Handelsman/Wood PIs, Howard Hughes Medical Institute). These two programs were selected because of their national prominence and our first-hand knowledge
of the development and implementation of both. The details of the studies can be found in Ebert-May et al. (2011).

The SI and FIRST II workshops were designed with the same overall goals, specifically to increase faculty knowledge about principles of active learning, to develop instructional materials that engage all students in the discipline-based practices of science, and to learn how to conduct learner-centered instruction. However, the two programs had somewhat different implementation plans. In the FIRST II program, faculty attended summer workshops that lasted from 2–4 days over a period of three years. Faculty teams learned to design, implement, and reformed courses or parts of courses. By reformed we mean courses that are learner-centered with higher cognitive goals for students rather than instructor-centered, information-transfer courses. In both of the programs, faculty learned how to use backward design (Wiggins & McTighe, 2005) to develop learning goals, assessments, and instruction that were aligned in terms of Bloom’s cognitive levels (Bloom 1956). Faculty learned and practiced tested pedagogies such as cooperative learning. The faculty in FIRST II were from all types of institutions, ranging from community colleges to research universities (Table 1). In contrast with the FIRST II approach, in the SI program, faculty attended a one-week institute and all participants were from research universities. The target courses for both projects were introductory biology and other core courses such as introductory genetics, ecology, evolution, and cell biology, as well as biological science for non-majors.

<table>
<thead>
<tr>
<th>Institution Type</th>
<th>Number of Faculty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Associates Colleges</td>
<td>40</td>
</tr>
<tr>
<td>Baccalaureate Colleges/Liberal Arts</td>
<td>21</td>
</tr>
<tr>
<td>Masters’ Colleges &amp; Universities</td>
<td>53</td>
</tr>
<tr>
<td>Doctoral/Research Universities - Extensive</td>
<td>31</td>
</tr>
<tr>
<td>Doctoral/Research Universities - Intensive</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 1: Number of faculty participants (N=185) in the FIRST II professional development program grouped by type of institution (Carnegie Classification).
Post-workshop evidence

Following the FIRST II workshops, we wanted to know if and to what degree faculty implemented what they learned into their subsequent teaching. Faculty were asked to complete several surveys so that we could determine improvements in faculty knowledge and experience with different aspects of active-learning pedagogy. As expected, faculty reported significant improvements in knowledge of each pedagogical area, including science-education reform, course and curriculum planning, use of technology in instruction, assessment, and cooperative learning, compared with their pre-workshop knowledge (Figure 1 from Bioscience 2011). Faculty also perceived (self-reported) improvements in their first-hand experience with each aforementioned pedagogical area, except for course and curriculum planning. Faculty already had substantial experience with this variable before the workshops and their initial high scores in planning remained high (Ebert-May et al., 2011).

Faculty provided similar data following the SI workshops (Pfund et al., 2009). There were significant increases in their self-reported knowledge of scientific teaching, active learning, assessment, and development of learning goals following the workshops. The consistencies in the self-reported data between the Pfund et al. (2009) and the FIRST II/SI (Ebert-May et al., 2011) were remarkable.

In both FIRST II and SI, faculty also provided data on their perceptions of their use of active-learning strategies after completing the workshops. A majority of faculty reported use of specific inquiry-based and learner-centered teaching practices following professional development. At least weekly or monthly and often in each class period, they used practices such as answering questions from individual students in class, cooperative learning, open dialogue and debate among students, communicating course goals and objectives to students, case studies, and problem-based learning. Overall, the majority of participants reported that they were using active-learning techniques, resulting in more learner-centered teaching one year after completion of professional development. We considered these data accurate because we believe that faculty perceptions of their teaching were genuine, and we had no reason to doubt their claims.

Based on the self-reported evidence alone, we would have concluded that both PD programs were effective, resulting in faculty changing their teaching practices following the workshops. Assessment of the FIRST II/SI project was based on a mixed-methods approach, however, that used more than self-reported survey data. Specifically, we also collected observational data on actual teaching practices in the classroom. To evaluate the utility of different evaluation methods we tested the null hypothesis that there is no difference in self-reported and observational data in terms of faculty teaching practice. We used the Reformed Teaching Observation Protocol (RTOP, Sawada et al., 2002) as a measure of the degree of active-learning
instruction and student involvement that occurred in a classroom (Sawada et al., 2002). A final RTOP score is obtained by summing five subscales that quantitatively rate an instructor on the following:

1. lesson design and implementation (i.e., the extent to which an instructor elicits students’ prior knowledge, to what extent the instructor engages students and allows them to influence the focus and direction of the lesson);
2. propositional knowledge (i.e., how a teacher implements discipline-specific content);
3. procedural knowledge (i.e., the inquiry process students use in the lesson);
4. communicative interactions (i.e., the proportion and variety of discussion that occurs among students and between the teacher and students); and
5. student-teacher relationship (i.e., attitudes of the teacher and students toward each other and the quality of their interactions).

RTOP scores range from 0 to 100, with higher scores representing more learner-centered classrooms (i.e., students actively participate, take primary responsibility for their own learning, interact with each other, and shape the direction of the discussion), whereas lower scores indicate lecture-based, teacher-centered classrooms (i.e., lecture is the primary mode of communication, with minimal roles for students beyond note taking or the use of personal response systems). The detailed method we used to rate videos using the RTOP is explained in Ebert-May et al. (2011).

Seventy-five faculty who participated in the FIRST II/SI projects provided both the self-reported data described previously and submitted two videos of their teaching over a two-year time period, post workshop. Of these faculty, 75 percent used lecture-based, teacher-centered pedagogy in the final videos of their teaching. Hence, faculty perceptions of their teaching that they reported in surveys did not match their actual teaching practice. Here we note that faculty were not giving false answers to survey questions; in fact, the majority of faculty were very excited about changing their teaching upon completion of the workshops. Rather, what faculty believed they did in classrooms post workshop is not what we saw. We assert that the majority of faculty did not understand all of the components of learner-centered teaching and needed much more practice and mentoring to be able to accomplish such teaching.

To help us understand what influences faculty teaching practice following PD, we identified a set of variables for development of a predictive model of actual teaching (i.e., RTOP score). We categorized the variables into three groups: experience, course, and appointment. “Experience” included four variables—number of years teaching, knowledge and experience with active learning, the type of PD program (i.e., FIRST II or SI), and instructor’s confidence. “Course” was defined based on number of students enrolled and instructor’s challenges to implementation of learner-centered classrooms. The biggest challenge reported by
participants was having time to design class meetings, grade and give feedback, and balance teaching with other responsibilities. “Appointment” was based on the percent of time spent teaching, tenure status, and self-reported data about departmental support. Faculty indicated that their departments were generally committed in various ways to improving undergraduate biology education such as improving retention of STEM majors, improving pedagogy (from traditional lecturing to active, learner-centered learning), and changing the culture and reward system for faculty.

**What variables predict teaching practice?**

The results of our model analysis indicated that four variables warranted consideration in relation to teaching reform and PD. Specifically, the most significant and consistent predictors of RTOP score were the number of years of faculty teaching experience and enrollment size in a course. Of secondary importance was the proportion of the faculty’s appointment devoted to teaching and their experience with curriculum reform and active-learning teaching practices. A negative relationship existed between years of teaching and RTOP score, indicating that more novice teachers implemented inquiry-based, learner-centered instruction to a greater extent than experienced teachers. One possible explanation for our finding is that faculty with more years of teaching may have greater difficulty implementing or are less inclined to implement learner-centered teaching. Collectively, the four variables explained approximately 20 percent of the teaching practice of the individual. Here we note that not all factors explaining the additional variance (80%) are likely to be measurable. We concluded from our study that traditional PD workshops alone were insufficient to enable faculty to implement learner-centered classes.

**The FIRST IV model for professional development**

In our subsequent approach to PD (FIRST IV, Ebert-May / Derting PIs, NSF-DUE 0817224), we modified the program design to include structured teaching experiences, reflection by participants on their own teaching, and mentoring by teaching experts, in addition to two, four-day annual workshops. The major changes in the workshops included who we recruited and what we did. In part, because of the negative correlation between RTOP and years of teaching, we focused on recruiting postdoctoral fellows to the PD program. Postdocs are our future biology faculty and are a population that has not yet established their teaching styles, may themselves have been educated in cooperative, active learning classrooms, and may more readily learn and adopt new teaching practices and pedagogy. The design of the FIRST IV workshops shifted from examining theories of pedagogy and putting techniques into practice (as in FIRST II) to having participants design an entire new course based totally on reformed instruction. Postdocs then taught that or a similar course at their home institution.
or a nearby institution with mentoring from expert faculty in FIRST IV, returned to a second workshop to reflect on and analyze the student assessments and videos of their teaching, revised their course, and taught for a second time the following academic year. Mentoring was emphasized as part of the PD process.

Two hundred-one postdoctoral fellows participated in the FIRST IV program, and nearly half are currently in early career faculty positions. A key research question we are evaluating is whether our revised approach to PD results in postdocs and faculty who implement learner-centered inquiry-based courses. Although data compilation is not complete, results from over half of the participants indicate that over 60 percent of the postdoctoral scholars implemented learn-centered courses, as determined by RTOP scoring of their teaching videos. This result is a dramatic improvement when compared with the 25 percent of FIRST II/SI faculty who demonstrated learner-centered classes following that version of our PD program. We will also analyze self-reported data to determine whether or not accomplishment of learner-centered teaching is associated with more accurate self-evaluation of one’s teaching practice.

Lastly, to evaluate the long-term effectiveness of our PD program, we are conducting a longitudinal study of the teaching practices of FIRST IV participants. With our multi-pronged approach to project evaluation, we are rigorously evaluating the effectiveness of the FIRST IV PD model to identify key variables associated with its impact on producing learner-centered, rather than teacher-centered, faculty. We assert that similar mixed-approaches to program evaluation are essential to the development of empirically-tested models of PD that can expedite teaching reform in STEM classrooms.

References


10. Some Cognitive Issues in Engineering Design

doi:10.7936/K72Z13FZ
Robert A. Linsenmeier\textsuperscript{123}, Jennifer Y. Cole\textsuperscript{4} and Matthew R. Glucksberg\textsuperscript{1}

Departments of \textsuperscript{1}Biomedical Engineering, \textsuperscript{4}Chemical and Biological Engineering, and \textsuperscript{2}Neurobiology, and the \textsuperscript{3}Northwestern Center for Engineering Education Research, Northwestern University

Correspondence:
Robert A. Linsenmeier
Department of Biomedical Engineering, Neurobiology, Northwestern Center for Engineering Education Research, Northwestern University
2145 Sheridan Road
Tech E368
Evanston IL 60208-3107
r-linsenmeier@northwestern.edu
http://www.nceer.northwestern.edu/

Abstract

Design is an important skill for engineers. Engineering students may encounter design courses first as freshman or not until they are seniors, depending on their universities. With the goals to learn the design process, enhance teamwork, and enhance communication within the team and between the team and the client, design is integrative, requiring students to use skills that are rarely needed in other courses. Students are familiar with the domain and the logical progression in textbooks from concept to concept, but design problems are often not well formulated, requiring students to formulate the problem more exactly before working toward a solution. Teaching design can be difficult as it requires moving students out of their comfort zone into the open-ended world of design, where there is not a single right answer. Mathematical modeling is an important aspect of design, and scaffolding appears to be helpful in improving students' abilities to generate and use mathematical models in biomedical engineering senior design. A study of student capabilities in modeling, followed by a classroom intervention in biomedical engineering design is discussed.
Introduction

Lumping engineering and technology with science and math under the acronym “STEM” recognizes that all these fields, particularly with respect to collegiate education, share commonalities in terms of content; indeed, more than a year of the engineering curriculum comprises courses in basic science and math. However, there are important features that distinguish the E in STEM from the S and M (we will not discuss T). A BS degree in engineering, in contrast to those in science and math, is a professional degree, and students receiving bachelor’s degrees often become practitioners immediately. Engineering is often defined as applied science, but the border between basic and applied science is fuzzy, so, educationally, the applied nature of engineering does not alone distinguish it from science. Many people who call themselves scientists apply their knowledge to practical problems. An education in engineering is distinct, though, because all engineering curricula and few if any science curricula include design courses. The teaching of engineering design is mandated in all engineering disciplines by ABET, which formerly stood for Accreditation Board for Engineering and Technology. One of the requirements for all programs is that students graduate with competency in 11 general outcomes, some of which all STEM students should possess, for instance, abilities in data analysis, communications, and life-long learning. Another outcome, outcome C, is one that is only expected of engineers—an “ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability” (ABET, 2006). Design was rated as being important by a large sample of engineering practitioners, with a mean rating of just below four (“quite important”) on a five-point Likert scale (Passow, 2012). It was slightly but significantly below several ABET outcomes, including teamwork, data analysis, problem solving, and communication, but, as noted below, design courses are important for teaching at least three of those outcomes. Therefore, the results of the survey reinforce the importance of design in the engineering profession and of design courses in the curriculum.

This chapter will be about the challenges of teaching design in the engineering curriculum: where it fits, the goals in teaching it, what students learn from it, why it is difficult to teach and to learn, and some aspects of how we might teach it better. This discussion will suggest some hypotheses about students learning design for which we do not yet have specific support or disproof or which may exist in parts of the rather large design education literature that we have not yet found. This chapter is just part of a discussion about what we might study in design education. However, we will also address a specific question, for which we do have data, about students’ use of mathematical modeling in design.
Where is design in the curriculum?

Essentially all engineering programs require design courses in the senior year, partly motivated by ABET requirements, and call it “capstone” design. Capstone design differs depending on the field of engineering. Here we will focus on design teaching that is more characteristic of mechanical engineering and biomedical engineering, the design of a device or product to meet a specific need. This lends itself to many different projects, so teams of a few students are typically assigned different projects. In chemical engineering and industrial engineering, design is usually about a process, rather than an artifact, and it is more common for the whole class to work on a single problem, also usually in teams. Within a field, such as biomedical engineering, there are many important common elements in design courses, but there is also some variation with respect to team size, time devoted to a single project, and emphasis on different aspects of the design process.

Across universities, there is also a lot of variation in whether design occurs earlier than senior year in the curriculum. A few universities have their students engage in a single design project over two years or more (e.g. Milwaukee School of Engineering); others have a sequence of project-based courses (e.g., Worcester Polytechnic and Johns Hopkins); and others have optional design certificates that include courses at different levels (e.g., Northwestern). For a long time, engineering faculty resisted the idea that students could design anything, or even learn about design, until they had taken math, science, and fundamental engineering science courses like circuits or fluid mechanics. Gradually, faculty realized that if a student was majoring in engineering, he or she might like to know what engineering was like earlier than the junior year. Most universities now require some type of freshman engineering course, and sometimes it is design-based (e.g., Northwestern, Rice, Johns Hopkins, and others reviewed by Sheppard and Jenison (Sheppard & Jenison, 1997). Freshmen can do some aspects of design, and it is generally believed that design education is good at that point (Hirsch et al., 2002; Sheppard & Jenison, 1997). Design experiences probably provide improved student engagement and identification as an engineer, and this may enhance retention in engineering (Knight, Carlson, & Sullivan, 2003; Ohland et al., 2008). There is insufficient data on all the benefits for college students, but it is notable that K-12 engineering outreach efforts, such as Project Lead the Way (http://www.pltw.org/our-programs/high-school-engineering-program) and the robotics competition US First (http://www.usfirst.org), very often engage students through design. Notably, in the context of what we will discuss below, these are short on mathematical aspects of engineering.

Goals of teaching engineering design

What does it mean to teach design? What are the learning goals of these courses? At the risk of great oversimplification, design courses often have three major goals. The first is to teach a
structured way to go from a real-world problem to the design of an artifact that can solve that problem—the “design process.” The various formulations of the design process involve several steps with several checkpoints requiring decisions about alternatives along the way. These decisions are often based on a team’s own research and analysis but are also shaped by feedback from stakeholders external to the design team itself. Figure 1 is one representation of a design process. Because of concerns about efficacy, safety, and reliability of a product, the design process may be embedded in another process to ensure design “controls.” For instance, FDA provides structured guidelines about validation and verification that students in biomedical engineering should learn. The second goal of design courses is to enhance teamwork. Most real-world projects involve teams with different kinds of expertise who need to work together efficiently. The third, related goal involves enhancing communication within the team, between the team and the client, and with those who will translate the design into practice. There is a strong requirement for documentation as part of the communication work.
Figure 1. One representation of the engineering design process indicating multiple requirements for validation and revision and checking that the design is aligned with constraints and specifications.

The educational methods for meeting these goals almost always involve grouping students into teams that do one or more iterations of the design process per quarter or semester, resulting in a design to solve a problem. Along the way students generate written and oral reports. Design courses usually involve some lectures, but they are never exclusively lecture courses; they always emphasize hands-on, active learning, which distinguishes them from many STEM courses.

What is difficult for students about design and design courses?

Interestingly, a substantial fraction of engineering students never become professional engineers. The Academic Pathways study from the Center for the Advancement of
Engineering Education surveyed 1,100 engineering seniors from a range of institutions and found that about 26 percent of engineering seniors were definitely or probably going to pursue careers outside engineering, with the percentage a little higher for women and a little lower for men (Sheppard et al., 2010). They go into business, medicine, scientific research, law, consulting, and other careers, presumably because, as stated by Jay Rogers, the vice president of recruiting for Randstand Engineering, a large engineering staffing company, “No matter what business you’re in, everybody has problems that they need to solve. Engineers are equipped to solve problems” (Boykin, 2013). While there are different perspectives about whether the US is producing enough engineers (Boykin, 2013), this diffusion into other fields is probably not much different than what happens to students majoring in other parts of STEM, who are valued for their transferable skills as much as for their content knowledge. Moreover, only a small number of those who take engineering jobs are really engaged in design. Engineers work in many areas, including quality assurance, regulatory affairs, marketing and sales, customer relations, and systems analysis. However, all students have to satisfy the design outcome, and even without the powerful stick provided by ABET guidelines, there is little question that everyone should take it, even though design courses are usually resource intensive. Part of the value of a design course is that it connects a student’s classroom engineering work to its social, economic and environmental implications, but we believe that much of the value relates directly to the cognitive challenges that students have to overcome to succeed in design courses. We propose several here but do not claim that this is an exhaustive list. We assert that the ability to deal with these challenges is almost certain to be of value to students going into any career in which the daily activities require flexibility, planning, time management, and problem solving.

First, design is the one activity where students absolutely have to use content from many earlier courses, as well as content that they never had before, so it is highly integrative. Design both provides a unity to a student’s education and stretches him or her to transfer earlier knowledge to a new situation. In most engineering courses, the domain of work is well-defined, and the instructor has worked hard, or there is a standard progression in the textbooks, all to make sure that the course moves logically from one concept or topic to another. Thus, a path is laid out, and the student rarely has to reach very far back to use material learned much earlier. In a design project, this guidance from the instructor, textbook, or syllabus is missing, and students have to figure out what material they need. This requires a more holistic kind of thinking and resourcefulness on the part of the student. A relevant way of describing this comes from the adaptive expertise literature (Schwartz, Bransford, & Sears, 2006; Walker, Cordray, King, & Brophy, 2006). Adaptive experts are innovative, but they are also efficient. Efficiency means being able to see the deeper features of a problem, possibly relating it to some prior experience, and calling to mind the most appropriate pieces of information from all those stored in memory to solve it.
Second, in many design courses involving real clients and real projects, it is not just the solution that is unknown. The design problem is often not well formulated either. In the scenario that we will discuss further below, the need is to find a way to treat jaundice in premature infants in resource-poor countries. Incubators and ICUs are not available, so where does one start? Students have very little experience dealing with complex situations that require formulation of the problem and consideration before working toward a solution.

There is some opportunity to give students practice in these two issues by appropriately structuring the engineering classes that precede the design course. The instructor can emphasize near transfer, perhaps within the course, by asking questions that require integrating current with previous material or by reminding students how earlier material is relevant. It is also possible to leave homework problem statements somewhat open-ended and requiring the students to identify their assumptions. To make another blanket statement without data, most students seem to dislike this. (One of the authors teaches a physiology course with a great deal of material and usually devotes some percentage of the questions on the final exam to “integrative” questions, rather than the systems covered most recently. This causes great consternation to students, and they need explicit guidance to see what threads have come up repeatedly (e.g. properties of epithelia, autonomic nervous system, regulation of metabolism, diabetes) and therefore make good subjects for integrative questions. This is somewhat helpful, but answers to these questions are still often disappointing.) In fact, it can seem artificial to do open-ended homework when the instructor knows the best assumptions or method or answer all along. That is not an issue in a design course where one may have only more or less complete or elegant designs, not right answers.

Third, once a team defines the problem and tackles one phase of the design process, how does the team know when their work is good enough, and they can move to the next step? Have all the important constraints and parameters been taken into account? Like point two, this concerns uncertainty but also involves confidence in one’s decisions. Maybe there is some highly useful piece of information lurking in the world if you could only find it. This can delay making a decision and taking the next step. Maybe another day of brainstorming or tinkering with a physical model will result in a flash of insight that will take the team down a different and more productive path. Students are used to completing a problem set or exam and moving on, using a tempo that is controlled by faculty. Of course the real world only works that way for employees, not the engineering managers or leaders we hope most of our graduates will become. Design requires working outside the boundaries of traditional schoolwork, although this point may not be made to them explicitly.

Fourth, once a team is committed to a path, there is a tendency to stick to it. This can lead to over-commitment to a flawed approach. If one reaches an unforeseen difficulty downstream, or predictions of physical or mathematical modeling do not work out, how far do designers
go to try to adjust along that path, and at what point do they move to a different path? Also, if students decide that the path is wrong, how far do they have to back up to get to a productive path? Design requires not only external review, but also constant meta-analysis of where designers are on the path, what took them there, and where the wrong turn might have been. It requires justifying decisions at each step as thoroughly as possible, again something that is not required very often in prior courses. Over-commitment is not a unique problem for students tackling design but applies frequently in research, where there can be a tendency to hold ideas longer than data justifies, so this is a point that is common to all STEM fields.

Fifth, it is not unusual for students to hold misconceptions very strongly, even about topics important to their majors. This is not surprising—even physicists have a parallel “folk physics.” Students may be able to hold them because the worst that happens in most contexts is the loss of a few points on an exam with no significant pressure to learn and use the right concept after the exam. We hypothesize, based on some of the work presented below, that misconceptions could be a bigger problem in design courses, because they could lead a design team down an unproductive path and could cause a flawed or failed design. They could go undetected until the prototype phase, which is very late. Even at that point we think that misconceptions about principles may be difficult to separate from small problems in making a prototype.

What about the innovation side of adaptive expertise? Is learning to solve design problems difficult because innovation is difficult? Possibly, but for many problems brainstorming can turn up many solutions worth exploring, especially in a group. There can, however, be a tension between trying to follow the design process faithfully and generating new ideas that leapfrog over certain steps or tear down previous progress, and this may be a challenge to student design teams.

**Is design hard for everyone?**

To be sure, design is not always the hardest course. Some students excel at design and gravitate toward it naturally. There are probably several reasons for this. It gives them an outlet for creativity and allows them to use a different aspect of engineering than their other courses. The connection to the social, economic, and environmental implications may be energizing. The extended length of a project means that there is more time for reflection, failure, and revision than in most college experiences. Creating a successful design is extremely rewarding, leading to positive self-image. These thoughts suggest that certain personality types may at least enjoy design more and could also be more successful. Success may also depend on whether we consider freshman or senior design. We hypothesize that early experiences in design may be especially beneficial for students who are not as strong in
the quantitative courses that characterize the rest of the engineering education. Design may keep them engaged and allow them to see the value of moving forward. This hypothesis does not suggest that they are necessarily better at design than students who are stronger by traditional measures, only that it may rescue students who would otherwise be lost to engineering. However, by senior design, we further hypothesize that these rescued students could be in trouble. The requirement to have done well in those other courses, and be able to use material from them adaptively, may hinder a student’s ability to succeed in designs that require more analysis.

**What is difficult about teaching design?**

Instructors face some difficulties in teaching design. We expect that most would say that finding appropriate clients and appropriate projects for the students is one of their biggest challenges. The projects must be ones where the students have enough background to realistically tackle the problem and make progress in a quarter or semester. The clients must be flexible enough to listen to a range of solutions. They need to be accessible. They need to provide feedback, but not a straitjacket, and need to understand that the students are learning a process as well as solving a problem.

However, the other challenge in teaching design, which some faculty may not even be completely aware of, is moving students out of their comfort zone of structured courses into the more self-directed, independent, and yet team-oriented, uncertain world of design that is outlined above. Faculty, who are used to uncertainty in their own lives based on research, design, or both, were probably good at taking this step themselves, and may not realize, or forget, the hurdle that it presents for some students. Consequently, we have worked on providing scaffolding to help students learn and instructors teach one of the difficult aspects of engineering design.

**The value of mathematical modeling in design**

Design consists of many stages, so if one wants to improve design education, what approach should one take? Our approach was to tackle one of the elements that we consider important and which our experience suggested was one of the most difficult for students. This is the use of mathematical modeling. Our work on this, a study of student capabilities followed by a classroom intervention in biomedical engineering senior capstone design, has been reported in three proceedings papers (Cole, Linsenmeier, Glucksberg, & McKenna, 2010; Cole, Linsenmeier, Miller, & Glucksberg, 2012; Cole, Linsenmeier, Molino, Glucksberg, & McKenna, 2011), and will be condensed and reviewed here.

Modeling can be used for several purposes in the design process. Modeling can
1. replace expensive and time-consuming tests of physical models, or can
2. predict the range of parameters that need to be tested in physical models.

Models can also help to

3. rule out seemingly reasonable designs that are destined to fail,
4. optimize design parameters and avoid overdesign of components,
5. explore the likely range of performance of a device, and
6. estimate failure probabilities given data on individual components.

When we use the term modeling, we mean mathematical modeling, implying that one identifies, solves, and interprets equations that describe a physical system. This always requires making simplifying assumptions and recognizing what those simplifications and assumptions do to the validity of the model’s output. In our work we treated modeling as a sequence of six steps, identified by Gainsburg (Gainsburg, 2006). Gainsburg reviewed several frameworks for modeling from the literature and created a synthesis that also depended on her detailed analysis of the work of a practicing junior structural engineer who found that he could not use existing models “off the shelf.” The steps, with a brief description of each one, are the following:

1. **Identify the real-world phenomenon**: Determine what needs to be modeled; sketch the system; identify the parameters and variables of the system.
2. **Simplify or idealize the phenomenon**: Revise the sketch; list assumptions; reduce the complexity.
3. **Express the idealized phenomenon mathematically** (i.e., “mathematize”): Use symbols to represent the parameters and variables; show relationships between parameters and variables using mathematical operators. Write equations that describe the system, either de novo or by modifying existing equations.
4. **Perform the mathematical manipulations** (i.e., “solve” the model): Generate solutions over a range of interest by varying the parameters; obtain numerical values or graphs representing the solution.
5. **Interpret the mathematical solution in real-world terms**: Use the results of the model to refine the design, recognizing the assumptions and simplifications that were made.
6. **Test the interpretation against reality**: Check the validity of the model against experimental data or physical model.

The authors consider mathematical modeling to be central to design. Do others see it this way? One data point was obtained by the Center for the Advancement of Engineering Education at the University of Washington, which surveyed graduating seniors and asked them to rate the most important elements of design (Atman et al., 2010). Each student was allowed to pick the top six of 23 elements. Figure 2 shows the percentage of students who chose each element as one of their top six. Modeling was ranked 12th, exactly in the middle of...
these elements, with about a quarter of the students naming it. Because other elements of design are also important, and students could only choose six, this can be interpreted to mean that they considered modeling reasonably important.

![Graph showing graduation students' evaluation of the importance of modeling in design](image)

**Figure 2.** Graduating students’ evaluation of the importance of modeling in design (Atman et al., 2010).

However, from our own work, presented below, we know that “modeling” means different things to students than it does to us. Despite all the quantitative courses they had been through, our seniors in biomedical engineering very rarely thought of mathematical approaches when we asked them to model. They thought mainly of physical models, as well as experiments with a physical system to generate descriptive equations, rather than starting from first principles and identifying and solving equations about the physical system (Cole et al., 2010). Physical models are certainly a valuable type of model, but our perspective, for reasons outlined above, is that mathematical modeling is at least as important.

**Scaffolding to help students learn about modeling**

We knew at the beginning of our study that students resisted modeling in biomedical engineering senior design and that encouragement by the faculty—“go model that situation
and see what you learn”—was not sufficient. When they did do modeling, with a high level of individualized instructor guidance, it was still not very satisfactory. We did not want to teach the use of a particular set of equations or the use of a modeling package, which is what many articles about design are concerned with. This may work for some engineering fields, but biomedical engineers may need many different kinds of models—transport, electrical, mechanical—depending on the design problem. Instead, we wanted to make sure that the students understood the whole concept of modeling, and factors that are common to the use of any model in design. These were captured well by Gainsburg’s steps. At the outset, we did not know exactly why students had difficulty in modeling or which of Gainsburg’s steps caused the most difficulty, but we thought that the earlier ones might be troublesome, because earlier courses had not provided much practice in these.

We worked in a biomedical-engineering senior-design course, taught by one of the authors in collaboration with another professor. Our strategy was to introduce mathematical modeling through a design scenario for the whole class that ran in parallel to the individual team projects. In a sequence of iterations, mostly done in class so that we would obtain individual responses, students were asked to do different parts of modeling aligned with Gainsburg’s steps. In the first year of the project, we simply collected data on student performance on these different steps. However, after each step we provided enough information that failure to do the previous step would not hinder performance on the next one. In the second and third years of the study, we incorporated some classroom lecture material after each stage of modeling. This was designed to discuss principles of modeling and student misconceptions on previous steps but not to provide guidance on the next step. We felt that allowing students to tackle each stage of the modeling themselves, rather than simply telling them about modeling, or telling them what we had learned from the previous classes, would make them more aware of their own limitations, and more receptive to the discussions in class. We analyzed student performance on different aspects of the scenario problem, and, to learn whether the scenario work impacted their design projects, we analyzed final reports and presentations from those projects as well.

Here is the scenario, which is about phototherapy for jaundice:

A design team has been asked to develop an inexpensive device to treat neonatal jaundice in developing countries. The device must be compatible with Kangaroo Mother Care (KMC), a blanket-like device that wraps a baby and holds it close to the mother’s chest, so that the skin-to-skin contact maintains the temperature of the baby. The inclusion of a method to treat jaundice in the KMC blanket is expected to replace the need for an incubator.

The design team’s initial research regarding this project uncovered that to some extent, every newborn is afflicted with jaundice. Jaundice is easily treatable with phototherapy, a process in which light is shone on the skin. Phototherapy used to treat neonatal jaundice
is typically conducted in an incubator, but this is too expensive for standard use in the developing world. However, the design team could not find a phototherapy device currently on the market that is affordable and compatible with KMC, leading them to design their own device. The phototherapy team became aware of blue LEDs and thought that they might be able to attach them to a flexible substrate that could be put next to the baby inside the KMC. These would be run by a battery power source. The team decided that the next step was to model this situation.

Students were also given some information about jaundice, properties of the LEDs, and recommended wavelength of the light, skin coverage area, and spectral irradiance (power per unit wavelength) required for treatment. Thus, the problem was constrained, and the exercises could focus on modeling rather than any of the design decisions that would have had to precede it. The scenario came from a project that an actual design team had done in a previous year, but no real team projects during the study were related to this one.

The sequence of activities, and their relationship to Gainsburg’s steps, is shown in Table 1.

<table>
<thead>
<tr>
<th>Stage 1 (week 1)</th>
<th>First year</th>
<th>Second and third years</th>
<th>Gainsburg steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1: Activity - Iteration 1</td>
<td>Class 1: Activity - Iteration 1</td>
<td>1, 2</td>
<td></td>
</tr>
<tr>
<td>Class 2: Lecture 1</td>
<td>Class 2: Activity - Iteration 2</td>
<td>1, 2</td>
<td></td>
</tr>
<tr>
<td>Class 3: Correction lecture</td>
<td>Class 3: Activity - Iteration 2</td>
<td>1, 2</td>
<td></td>
</tr>
<tr>
<td>Stage 2 (week 2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 4: Activity - Iteration 3</td>
<td>Class 4: Activity - Iteration 3</td>
<td>3, 4</td>
<td></td>
</tr>
<tr>
<td>Stage 3 (week 3)</td>
<td>Class 5: Activity - Iteration 3</td>
<td>3, 4</td>
<td></td>
</tr>
<tr>
<td>Homework: Activity - Iteration 3</td>
<td>Class 6: Lecture 3</td>
<td>3, 4</td>
<td></td>
</tr>
<tr>
<td>Class 7: Activity - Iteration 4</td>
<td>5, 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 4 (week 4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 1**: Steps in using the phototherapy scenario as a research tool, in the first year, and as a research and instructional tool, in the second year. Modified from (Cole et al., 2011).

The assignment in Iteration 1 was:

“As advisors to the design team, your job today is to tell the team:

a. What you think the design team should model.

b. What modeling approach you recommend for the phototherapy project.

c. How you expect the developed model will be helpful in the design.”
Note that this did not specify mathematical modeling, just modeling. Iteration 2 was similar, but more specifically highlighted mathematics. “To help design their device, the design team decided to create a mathematical model of the transmission of light from the device to the infant’s skin. The intended purpose of the mathematical model is to generate predictions of how different parameters will affect the performance of the device.” They were then asked to provide specific details relevant to the mathematical model, including sketches, relevant parameters and variables, and assumptions. We specifically asked as part of this that they should provide “A proposal for a possible mathematical approach to the problem.”

Iteration 3 was designed to determine whether students could generate and solve appropriate equations. “The design team knows that the light reaching the baby’s skin is a function of the distance from the LED to the surface of the baby and a function of the spacing between the LEDs. They want to use their mathematical model to determine how the different parameters under their control contribute to the amount of light on a patch of skin.

a. Help the design team by writing the equations to be used in the model.
   b. List any assumptions that the design team needs to make in their model.
   c. From your equations, calculate irradiance as a function of position on the skin.”

This iteration was done as homework, followed up by further questions in class. Before this iteration, students were given the equation for the light distribution falling on a surface from a point source, and we expected that they would try to generalize this to an array of LEDs.

For Iteration 4, students were given graphical results from model equations in which the parameters had been varied. They were asked to make decisions about the optimal spacing of the LEDs and the optimal distance from the skin.

In general, we found that students had great difficulty in the first three steps of Gainsburg’s modeling framework (Cole et al., 2010; Cole et al., 2011). There was not only difficulty in creating correct equations, but in identifying parameters that could be varied in a model, distinguishing parameters from the design constraints about wavelength and spectral irradiance (which they also wanted to vary), in drawing a sketch that would help them define relevant parameters and see relationships among them, and in identifying assumptions.

The dataset is very large and cannot be fully represented here, but a few key points can be made. Student performance on the question of the modeling approach that they would recommend is summarized in Figure 3. The key point here is that students could not think of mathematical approaches, and less than 20 percent did this even after our question specifically highlighted and prompted the use of mathematical models in Iteration 2. Responses to Iteration 2 showed that many of the students failed to give any sort of model, or
suggested literature research, or proposed a physical model without tying it to a way that it could help in the design. This graph aggregates data from Years 1 and 2, which gave similar results. In Year 2, Lecture 1 led to only a very slight improvement in Iteration 2.

![Graph showing students' responses to modeling questions](Image)

**Figure 3.** Responses to the question of how students would model the phototherapy scenario.

Table 1 shows selected results of the study on key elements of modeling: making assumptions, representing the physical system in equations, and evaluating whether the model represents reality. These data come from Iterations 1 through 4. Making and recognizing assumptions is crucial and is a skill that extends beyond modeling to almost any real-world decision. Ideally, in the present context, we were looking for assumptions that could be used to simplify equations, or, in the event that an existing model was available, the assumptions would be used to tailor those model equations to suit the student’s needs. In the first year, almost half the students were unable to articulate any assumptions that were needed. The good and bad assumption lines do not indicate fractions of students (as all other lines do), but rather the total fraction of good and bad assumptions made, as defined in the table legend. Students did considerably better in making assumptions in the second year, when modeling was discussed in the lectures, and everyone reported at least one assumption, although a few students only had bad ones. “Mathematizing,” or describing the
system with an equation, in this case is finding an expression for the light distribution on a surface when a line or array of LEDs is at some height above the surface (and one assumes point source characteristics). The solution for one LED was given in class and was provided as a starting point, but many students in both years took it as a final solution. Some students did not even get that far in the first year, however, and gave no equation. In the second year, more students recognized that this was at least a good starting point, all students worked to generate some equation, and some gave a completely correct equation. While the most appropriate equation involves only geometrical considerations and summation, which mathematically are well within the students’ capabilities, it is not a trivial problem. In evaluating whether the model was a good representation of reality, there is not a definitively correct answer. It is good in some respects and bad in others. Thus, in order to go further with the model, it is important to understand what its strengths and weaknesses are, and this requires some explanation of why it is considered good or bad. A much higher percentage of students justified their answer in the second year, although almost half still did not do this.

<table>
<thead>
<tr>
<th>Performance on the Phototherapy Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
</tr>
<tr>
<td><strong>ASSUMPTIONS</strong></td>
</tr>
<tr>
<td>Good assumptions</td>
</tr>
<tr>
<td>Bad assumptions</td>
</tr>
<tr>
<td>No assumptions</td>
</tr>
<tr>
<td><strong>MATHEMATIZATION</strong></td>
</tr>
<tr>
<td>Correct one LED equation</td>
</tr>
<tr>
<td>Correct multiple LED equation</td>
</tr>
<tr>
<td>Partly correct multiple LED equation</td>
</tr>
<tr>
<td>Incorrect equation</td>
</tr>
<tr>
<td>No final equation</td>
</tr>
<tr>
<td><strong>MODEL EVALUATION</strong></td>
</tr>
<tr>
<td>Answer with justification</td>
</tr>
<tr>
<td>Answer with no justification</td>
</tr>
<tr>
<td>No clear answer</td>
</tr>
</tbody>
</table>

**TABLE 2**: Selected results of student work on mathematical modeling related to the phototherapy scenario. All values represent the percentage of students whose work was judged to be in the categories shown, except in the good assumptions and bad assumptions rows. The percentage of good assumptions was computed as:

\[
\text{good percentage} = \frac{\text{total good assumptions made}}{\text{N students x M possible good assumptions}}
\]

A similar measure was used for bad assumptions. There were four possible good assumptions and six
possible bad assumptions in the work analyzed. The design course had 38 students in Year 1, and 74 in Year 2.

Of more importance than the students’ work on the scenario itself is whether they were able to carry the lessons to their work on their own projects. We analyzed this in the second and third years of the project, with the scenario in the third year being executed just as in the second year (Cole et al., 2012). Because the projects were all different, they were assessed with rubrics that again specifically looked for evidence of accomplishing the steps in Gainsburg’s framework (Gainsburg, 2006). Figure 4 shows a simplified presentation of part of this analysis. Students showed marked improvement in the mathematical modeling aspects of their project reports in Years 2 and 3, after the scenario work accompanied by discussion. Stating the problem to be modeled, at least partially listing parameters and assumptions, generating a mathematical representation of the physical situation, and generating outputs of the model all improved from threefold to fourfold. However, the mathematical representations were not always outstanding (Cole et al., 2012). Parameters were not always fully stated. Also, in some cases where an equation was given, it was simply an equation that students found and was not tailored well to the specific problem. In addition, while about half the teams reached the point of using the model to generate results, one of the key reasons for modeling—exploring the parameter space—was done adequately by only about 30 percent of the teams. As a result, the model could not really be used optimally to inform design decisions. We believe that some of these failures occurred simply because students ran out of time in the 10-week quarter. It seemed that most of the groups understood how models could be useful, which was not the case previously.

Figure 4. Aspects of mathematical modeling in senior design projects.
Issues for the future

The scaffolding provided by the scenario exercise appeared to be helpful in improving students’ abilities to generate and use mathematical models in biomedical engineering senior design. It allowed students to see modeling in a context where they had not used math previously and emphasized the critical early stages of creating a mathematical model of a physical situation, which is a frequent need in research and design, but had not been needed in their previous coursework. It was not a perfectly successful intervention judging from either scenario performance or project reports, but the scenario work appeared to give them a better conception of modeling than they had previously. One could argue that simply talking about modeling without having students work through the scenario might accomplish the same gains. However, we believe that the scenario work probably had several benefits in line with general education theories. Their failures in the scenario work probably made them recognize their limitations and prepared them to listen to instruction, as well as gave them practice, and made it clear that the instructors thought that modeling was important. It seems unlikely that a one-hour lecture covering the points of Gainsburg’s framework would accomplish the same thing. We are now testing whether a similar approach is useful in process-oriented, rather than artifact-oriented senior design course in chemical engineering.

As described earlier, modeling is only one of the elements of design. We believe that we have helped the students transfer earlier knowledge about math to design, but this still leaves many other elements of design that are different from their work in earlier courses and with it the general problem that in design students are doing something more open-ended and unstructured than in the past. Do we need to scaffold all the other elements of design as well? Possibly, but it might not take such extensive effort. Case studies of the ways in which previous design problems have been successfully handled and how decisions were made along the way might be sufficient.

We can also ask whether all of the scaffolding needed in design has to be left to the design courses. We think not, but this takes collaboration among faculty. Homework problems in earlier courses can require students 1) to help determine what problem needs to be solved, 2) to identify and justify assumptions, and 3) to explain whether their solutions are reasonable and are related to real world behavior. It would also be possible to explicitly clarify more often that the equations students are learning and solving in other courses are models and that there are assumptions behind them. This is common in some courses but not all.

There are a number of research questions remaining that could be explored by a collaboration of cognitive scientists and engineers. First, earlier we listed several hypotheses about students learning design that could be tested. For instance, we have emphasized that design requires transfer from earlier courses. Is the opposite also true? Does early experience in design
provide transferable benefits to students in other aspects of their curriculum? Second, how is proficiency in design, which could be defined in several ways, related to doing multiple design projects? Do students benefit from having done one design project in terms of making it easier to do another one? Do students gain metacognitive skills about pitfalls in sticking with an unproductive solution? Do they become more comfortable with uncertainty in other kinds of problem solving? One study suggests that students in senior design do not benefit greatly from having had freshman design (Kotys-Schwartz, Knight, & Pawlas, 2010), and if that is generally true then we need to ask how to correct that and carry the benefits forward.

While engineering design is mostly the province of engineers, it is reasonable to ask whether any of the discussion above or the lessons about design are relevant to teaching and learning in basic science. We believe that many of the features about design are valuable in any real-world context where the problem is not fully defined but must be solved anyway and within constraints. More specifically, students in basic science do face similar problems when they do independent research, so some of the same challenges and strategies are applicable there.

Acknowledgements

NSF Engineering Education Program grants 0648316 and 0891530, Ann McKenna (Arizona State University, and REU students Timothy Miller (Binghamton University) and Esteban Molina (Florida International University)

References

ASEE Proceedings papers are all available at http://www.asee.org/search/proceedings


11. Six Ideas That Shaped Physics: Rethinking Introductory Physics

doi:10.7936/K7JW8BSN

Thomas A. Moore
http://orcid.org/0000-0001-7797-5605

Department of Physics and Astronomy, Pomona College

Correspondence:
Thomas A. Moore
Department of Physics and Astronomy
Pomona College
610 N College Ave
Claremont CA 91711
tmoore@pomona.edu
http://pages.pomona.edu/~tmoore/

Abstract

Six Ideas That Shaped Physics is a comprehensive set of text materials, instructor resources, and web-based tools whose goal is to enable professors to pursue an innovative, activity-based approach to teaching introductory calculus-based physics course (even in situations where large classrooms are the norm). This chapter will describe what makes the Six Ideas approach distinctive, with special emphasis on why (in any STEM discipline) choosing appropriate instructional metaphors and constructing an interlocking, self-consistent course design is essential: all course elements (the textbook, class activities, homework, and exams) must work coherently together to produce genuine learning. I will also describe how this course design was adapted for use at Washington University in St. Louis and report on student outcomes there and elsewhere.

Introduction

The Six Ideas That Shaped Physics project represents an effort to completely rethink the teaching of introductory calculus-based physics for the 21st century. The overarching goal of this project is to make it practical for a professor to mount an introductory physics course that
not only presents a thoroughly 21st-century vision of physics, but also uses active learning
and other course design elements to help students learn the material more quickly and
effectively, even in situations where large classrooms are the norm. Instructor resources
include a detailed online instructor’s guide (which emphasizes the importance of creating a
self-consistent course design), computer software that supports both class activities and a
unique approach to homework, and a supporting textbook (Moore, 2003). In what follows, I
will describe what makes the Six Ideas approach distinctive, with special attention to how
important well-chosen instructional metaphors and a comprehensive, carefully designed, and
self-consistent course structure are to helping students learn effectively. I think these course
design issues have broadly applicable implications in STEM disciplines, where students
generally must master a huge amount of detailed technical information at the same time as
practicing critical-reasoning skills, applying and drawing logical conclusions from mental
models of the natural world, and correctly interpreting experimental data. The core question
is, what are the essential aspects of a course structure that enables students to develop expert-
like reasoning patterns as quickly and effectively as possible?

**Origins**

I have been pondering these questions and experimenting with solutions for more than 25
years. The Six Ideas project grew out of the Introductory University Physics Project (IUPP), a
national, NSF-funded project organized in 1987 to develop a fundamentally new approach to
teaching introductory calculus-based physics. The goal of this course is to provide science
and engineering students with enough facility in applying physics concepts to real-world
applications to enable them to succeed in their choice of science major. This is a very
important course, which is currently taught to roughly 200,000 physics, engineering, and life-
science students in the United States alone (Khoury, 2003; Mulvey & Nicholson, 2010). The
founders of the IUPP project, however, had noted a number of serious problems with the
course: Its content typically only went as far as 19th-century physics, which seemed
increasingly problematic as the 21st century approached. In many ways, its outlook was 19th-
century as well, in spite of the fundamentally new perspectives provided by 20th-century
physics. Since most of the 200,000 students would never take another physics course in their
lives, it seemed to the founders irresponsible to leave students with such an archaic picture of
the discipline. Moreover, standard physics textbooks had become massive reference tomes
that students were loath to lug around much less read). Their content was also frozen: they
were all so similar that opening each book to the middle would almost certainly land one in
the same chapter. The last great rethinking of physics content had occurred in the early 1960s
(in the immediate post-Sputnik era). Even in 1987, the student clientele and goals for the
course had evolved considerably since the early 1960s.
However, the most important concern was that educational researchers were just beginning to document just how poorly students in such courses were able to apply even basic physical principles to real-life situations. Before the IUPP project ended in 1995, research articles had abundantly documented that students in standard introductory physics were learning very little actual physics, and in some cases they were learning new misconceptions from the course to complement those they already had!

After pondering these problems, the IUPP steering committee (on which I served for eight years) issued a national call for alternative “model” curricula for this course consistent with the following explicitly stated goals:

• Recognize that less is more (a plea for greater depth and less breadth).
• Include 20th-century physics (in spite of the first goal!).
• Present a story line (to help with coherence, focus, and student motivation).

In response, I submitted a description of a model entitled *Six Ideas that Shaped Physics*, which had the following goals (in addition to the required three above):

• Support active learning.
• Build on what we know about student cognition.
• Explicitly teach expert-like scientific-modeling skills.
• Help students understand the hierarchy of physics concepts.
• Work within a standard classroom and class schedule.

The first two goals were especially important to me, partly because I felt that few other people were addressing these very important issues (indeed, none of the other submitted curricula addressed them at all).

The organizing concept of the *Six Ideas* curriculum was to sub-divide the introductory course into six units, each of which was focused on developing a single core idea that has fundamentally shaped physics during the last 350 years. Here is a list of these ideas in their preferred teaching order (though other orders are possible):

• Unit C: *Conservation Laws Constrain Interactions*
• Unit N: *The Laws of Physics are Universal* (Newtonian mechanics)
• Unit R: *The Laws of Physics are Frame-Independent* (Special relativity)
• Unit E: *Electricity and Magnetism are Unified*
• Unit Q: *Matter Behaves Like Waves* (Quantum mechanics)
• Unit T: *Some Processes are Irreversible* (Thermal physics)

This six-unit organization (1) by its very nature daily emphasizes to students that physics is organized hierarchically, with some ideas being more fundamental than others, (2) gives each
unit (and the whole course) a storyline that gives it purpose and direction, (3) builds contemporary physics into the structure, (4) provides motivation for cutting material (for example, whatever does not advance us toward understanding a unit’s core idea), (5) makes it reasonable to separate the textbook into small units that are less daunting and easier to carry, (6) makes it easier to present the material in different orders better suited to different calendars. Indeed the choice of *six* ideas (as opposed to five or seven) was mostly because 6 is divisible by both 2 and 3, making it possible to fit entire units into either a semester or quarter-based calendar. Also, though units C and N must be taught first and in that order, the other units are essentially independent and may be discussed in any order (or omitted entirely).

After a period of independent review, my model was selected (along with four others) for further development and testing. I was therefore mandated by the IUPP not only to create a development team, whose purpose was to flesh out the curriculum in more detail, but to formally test the resulting curriculum at Amherst and Smith Colleges in the 1992-1993 school year. These tests were evaluated by independent IUPP evaluators (*Di Stefano, 1996a, 1996b*).

It was this process that taught me most of the things that I will describe in this chapter. My original mandate was to rethink the teaching of introductory physics from the ground up, experiment with new approaches, disseminate the results to other institutions, and evaluate what happened. The evaluation tools available at the time were nothing compared to what is available now, but I still learned a lot (mostly by experimenting and failing).

One thing that I learned early on was the central importance of a coherent course structure. The course simply did not function well (leading to poor learning and student unrest) if all of the elements did not positively reinforce each other. This will be an important theme in what follows.

*Paths to Student Understanding*

Let us begin by considering what we are actually trying to do. Figure 1 illustrates perhaps the single most important goal of the course: building student understanding.
As the figure illustrates, building student understanding requires at least presenting the ideas clearly and coherently, as well as using active learning methods to ensure that the ideas actually engage students and become embedded in their minds. In what follows, I will explore the implications of each of these dimensions in turn.

**Cognitive Simplicity.** Physics professors generally pride themselves on being able to deliver clear lectures, but this is often as far as attention to this dimension goes. Since the mid-1980s, physics education research has increasingly documented the extent to which students bring active misconceptions about physics to the classroom. As these misconceptions are founded on successful (if partial and mutually inconsistent) intuitive models of the physical world arising on years of experience, they are very tenacious. Certain traditional textbook approaches to topics not only fail to rout these misconceptions (assuming that the student is a blank slate), but actually help students to fall into the very same conceptual traps that have also beguiled physicists in the historical past. Successfully presenting new and difficult ideas to students therefore not only involves developing clear and evocative metaphors, but also knowing where the misconception snares lie and constructing metaphors that guide students to take other cognitive paths.

As an example, consider the topic of special relativity. Presentations of special relativity in introductory physics courses (when it appears at all) often follow a historical path blazed by Einstein and others that was first designed for popular audiences. This path (perhaps not surprisingly) focuses on relativity’s most sensational elements (time dilation and length contraction), supplemented (in most textbooks) by a discussion of the (historically important but abstract) Lorentz transformation equations. However, this path has many buried traps that historically have snared not only students but also practicing physicists, as evidenced by...
lengthy discussions in the literature about various “paradoxes” (such as the “twin paradox” and the “barn and pole paradox,” to name just two of the most famous). These traps have generally given special relativity the unjust reputation of being too difficult for mere mortals to understand.

The problem, however, is entirely the result of choosing unclear and misleading metaphors. Let us examine the specific example of time dilation. This phenomenon is usually expressed by the pithy phrase “moving clocks run slow,” an idea that seems simple enough (even if somewhat unexpected). This simple statement is at best very misleading and in some cases actively wrong: *this is the wrong metaphor for describing this phenomenon.*

The classic “twin paradox” illustrates just one problem with this metaphor. Briefly stated, the “paradox” imagines twins Albert and Bethany, who synchronize their watches before Bethany climbs into a spaceship that travels at a relativistic speed to a distant star and then returns. How does the time displayed by their watches compare when they are reunited? Albert thinks that because “moving clocks run slow” and since Bethany has been moving at a high speed during her trip, her watch will register less time for the trip. In Bethany’s frame of reference, it is Albert who has moved away from her and returned at a high speed, so she expects *his* watch to register less time. Hence the “paradox.”

The above example exposes one problem with the idea that “moving clocks run slow,” which is the ambiguity of what “moving” means. This metaphor also (ironically) plays into, and so cognitively reinforces, the very deep misconception that relativity was created to correct: the misconception that motion is absolute (that is, you can tell with certainty who is “moving” and who is not). This is because almost anyone who encounters this phrase (even a physicist!) intuitively interprets this phrase as saying that motion (somehow) physically affects a clock’s workings so that it slows down, something like some magical form of friction. However, since a physical effect must either be present or not, one should be able to tell when something is moving and or not. The metaphor leads us inevitably in this direction and thus runs smack into the contradicting assertion that all inertial reference frames (that is reference frames that move relative to each other at constant velocities) are actually physically equivalent.

At this point, most students give up and say, “I am never going to understand relativity.” The problem is not with the student; it is with the metaphor. In the 1960s, J. A. Wheeler and E. F. Taylor presented an altogether different metaphor for time dilation that avoids these problems entirely (*Taylor & Wheeler, 1966*). This metaphor focuses on the deep analogy between measuring the time separation between two different “events” (something that marks both a point in space and an instant in time, like a firecracker explosion) and measuring the spatial separation of two points on a map. As Figure 2a illustrates, there are
three different ways to quantify the separation of two points on a map: (1) we can measure their north-south separation on some arbitrary coordinate axis (map grid), (2) we can measure the path-length between them using a tape measure laid along a curved path, or (3) we can measure the straight-line distance between them. Analogously, in the physical world of space and time, as Figure 2b illustrates, we can quantify the time separation between events (which look like points in the “spacetime map” shown in the figure) in three fundamentally and physically distinct ways: (1) we can measure their separation using a pair of fixed and synchronized clocks in a certain inertial reference frame, one present at each event (which on the spacetime map is equivalent to projecting their separation on the $t$ axis), (2) we can measure their time separation using a clock that is present at both events but travels from one event to the other along an accelerating path (which looks like a curved path on the spacetime map), or (3) we can measure their separation using a clock that is present at each event but that travels from one event to the other along a constant-velocity path (which looks straight on the spacetime map). We call the results of these physically distinct procedures for measuring the time between these events the “coordinate time,” “proper time,” and “spacetime interval” between these events, respectively.

Figure 2.

Now, we know that the north-south separation of two points is smaller than the distance between them, which is smaller (or equal to) any path length between them. We are not surprised by these differences, and we know that this is not the result of any magical stretching or shrinking of the rulers used to measure those separations: we understand that
we are measuring these separations in physically distinct ways. So, by analogy, we really should not be surprised that the three physically distinct ways of measuring time between the two events should yield different results. It turns out that the relative magnitudes are reversed (coordinate time is longer than spacetime interval which is longer than proper time) from what you would expect from the analogy, and though one has to work very carefully with students to make sure that this aspect of the metaphor is not confusing, the validity of the basic metaphor is not affected. Indeed, the very act of exposing this problem with the metaphor underlines the important metacognitive point that *this is a metaphor* (something that is just as true of the statement that “moving clocks run slow,” but not as clear).

This metaphor instantly resolves the twin “paradox” (which really does not even arise in this conceptual framework). Albert and Bethany are both present at the departure and reunion events but measure different *kinds* of time. Bethany’s spaceship accelerated during takeoff, slowed down when arriving at the destination, accelerated when leaving, and slowed down when arriving home. Albert, sitting at home, experienced none of these accelerations. Therefore, Albert measures a spacetime interval, while Bethany measures a proper time, which is always smaller than a spacetime interval, so Bethany’s watch registered less time. There is no ambiguity and no paradox. Moreover, this metaphor correctly focuses one’s attention on the relevant physical distinction between the twins: Bethany is accelerating and Albert is not, and that physical distinction between their reference frames is *not* “relative.”

My point with this lengthy discussion is that *metaphors matter*. One’s choice of metaphors makes a *big difference* in what difficulties students will encounter and how successfully they can overcome them. As this example illustrates, following the historical trajectory for teaching these concepts can lead your students to fall into the same traps into which physicists before them have fallen. At least in physics, a 21st-century perspective often provides metaphors for introductory topics that are broader, more powerful, and yet simpler than those used historically. Moreover, most people who have done the work of creating modern metaphors have done so precisely to *avoid* the traps that the historical approaches have exposed.

Providing metaphors is the job of the textbook, and in developing the *Six Ideas* project, I have consciously searched for new metaphors for various physical concepts (usually taking them from the literature, but sometimes devising new ones myself) to avoid standard misconceptions. As a result, the *Six Ideas* text uses non-traditional metaphors in almost every topic area in introductory physics. I believe that these metaphors really do help students to learn the subject more quickly and successfully, as I hope to convince you later.

*Active Learning.* However, we know that even a crystal-clear presentation does not ensure that students will learn. If physics education research has taught us anything in the past three
decades, it is that students learn best from activities that help them confront their misconceptions, practice new thinking patterns, and develop their critical-reasoning skills. They can no more pick up the ability to do physics from a lecture than they could learn to play piano from going to a concert. The key is guided practice with feedback.

The importance of active learning has been well established and is a feature of other efforts at reforming the teaching of introductory physics. I don’t expect that I need to justify this assertion to those reading this article, but some readers might like to know a bit about how I make this case to students and colleagues (both of whom are quite comfortable with the lecture format). With students, I begin with the simile mentioned in the previous paragraph: learning to think like a physicist is like learning to play a musical instrument. I say to them, ‘You would be outraged by a piano teacher who spent your entire lesson playing beautiful music and then said, ‘Go and do likewise.’ You know that such an approach is not effective. Then why would you put up with that from me in this class? You should be demanding that I make you practice, practice, practice on suitably chosen pieces and then offer you extensive and personalized feedback. This is how you will learn best and you know it.’ (I usually supplement this with analogies to learning to play a sport to make sure that I am reaching almost everyone.)

With physics colleagues, a version of these analogies can help, but it is also helpful to show them an article by Richard Hake (1998), which I will discuss in more depth later.

The Six Ideas project seeks to provide professors with resources to make it easier for them to change from a lecture-based classroom to a student-centered classroom, even in large-enrollment formats. Some of these are embedded in the textbook, but many are not and cannot be. My point is that the Six Ideas project is more than just a textbook; it is also about providing resources for transforming the classroom.

**A Course Ecosystem**

Even providing random resources is not enough. As I mentioned before, one of the most important things that I have learned from my experience is that a course is like an ecosystem, in which every part fills an essential niche and works together to create a life-sustaining whole. If any part is out of balance, the whole suffers. (Figure 3 conveys the same idea using a different metaphor.)
Let’s consider some of the implications of this interrelatedness. Let’s start by considering the basic differences between a traditional lecture-based classroom and a student-centered or “flipped” classroom (a feature of a few physics reform efforts in addition to the *Six Ideas* model). Figure 4 illustrates some of the differences (at least for the particular format of the “flipped” classroom associated with the *Six Ideas* model.)
In the traditional model, content is delivered through lectures, and students practice by doing homework outside of class. In this ecosystem, there really is no niche for the textbook, so it is not surprising that students in traditional classes report that they rarely refer to the textbook unless they need to look up a formula. The ecosystem of a typical large classroom also means that homework problems and test problems need to be easy to grade in bulk, which has traditionally meant problems that have simple numerical answers (whether the grading is done by computers or people). Since the only practice students get is from doing the homework, where they are also graded for “getting the right answer,” the only kinds of problems that students won’t fail to do reasonably well on are problems that either slavishly follow examples in the book or simple problems that involve finding the right formula, plugging numbers into it, and generating a numerical answer. The ecosystem demands this.

Therefore, it should be no surprise that such an ecosystem evolves students that are fittest to thrive in it; students who are good at finding formulas, grinding out answers, or carefully following examples from the lecture or textbook. As careful reasoning, deep understanding, and flexible modeling skills are not selected for in this ecosystem, these traits will not arise except in a few mutants (who go on to become physics professors).

I remember attending a talk at a national meeting about 20 years ago that detailed the paths of two students through an introductory physics course (determined through a sequence of interviews). One student was very enthusiastic about physics and very much wanted to understand the concepts, feeling that this was essential. The other student, from the beginning, was interested in doing only what was required to get a good grade. By the
middle of the course, the first student was working much harder than the second student, but was only getting a B, while the second was earning an A. By the end of the course, the first student, greatly discouraged, had given up trying to understand anything and had adopted the strategies of the second, and both were earning As. The first student (tragically) had evolved to succeed in the ecosystem.

The first step in constructing an ecosystem where students evolve to actually learn something is to refocus the classroom on student activity where students are guided to practice using the concepts being taught and get immediate feedback. This implies that students must receive the basic course content from somewhere outside of class. In the Six Ideas version of the “flipped” structure, the textbook fills this niche, but it takes a different kind of textbook to fill this niche well. Traditional textbooks are adapted to the traditional ecosystem, where they mainly serve as reference (and must contain lots of information to fit whatever the instructor might lecture about) and as a source of formulas and examples. These textbooks are neither meant for nor designed to take the place of lecture as the primary source of information. In the flipped structure, the text must be streamlined, conversational, and engaging, so that students can actually read it without going crazy.

This is not the only kind of change that has to take place. If the goal is to construct an ecosystem that really selects for students who deeply understand physics concepts instead of being able to simply crank out formulas, all aspects of the course must work together to support that goal, as Figure 5 illustrates.

**Figure 5.**
If our goal is to foster active learning of physics concepts, then, of course, class time must be allocated to do this. This means that the text must serve the role that lecture did in the traditional ecosystem, which has implications for the design of the text. Even a well-designed text will not serve its role if students don’t *read* the text, something that they do not desire to do freely. Therefore some kind of homework system must be devised that rewards students for reading the text.

Creating a student-centered class structure also requires time and experience, and many professors will not be able to do this without support. This can come partly from the text (which can provide “clicker” questions and other problems suitable to serve as classroom activities), but it may also need to come from online resources such as sample class plans. Finally, students will not learn the concepts unless the ecosystem makes it necessary to do so. Therefore, the text needs to emphasize concepts as well as providing “clicker” questions and homework problems where conceptual understanding is the focus, and exams must evaluate student mastery of concepts. Students quickly adapt themselves to do well on what really counts for their final grade, so the reward system must be carefully crafted so that the desired student traits reproduce themselves.

Suppose that one of our goals is also that students learn how to apply physics to the kinds of realistic situations they will encounter in their actual lives (see Figure 6).

---

**Goal: Students Learn to Apply Physics to Realistic Situations**

**Requirements:**
- Text must provide such problems
- Grading must reward *effort*
- Students must get feedback
- Exams must reflect this goal

**Text**

**Homework System**

**Exams**

*Figure 6.*
This will not happen if the text does not provide such problems, and many traditional texts do not, because (as we’ve seen) traditional homework problems are adapted to the large-classroom environment, where it is difficult to grade and, more importantly, to provide detailed feedback on large numbers of student problem solutions. So reaching our goal presents an even deeper challenge: how can we devise a practical homework system for large classrooms that enables students to get appropriate feedback on complicated homework problems? Moreover, even bright and able students are likely to fail the first time in solving realistic problems: if students are not to be discouraged, they must be rewarded for practice, (that is, for making a thoughtful effort), even if it is not ultimately successful. After all, this is precisely what good coaches of sports teams do: in practices, they set challenges for their players that are appropriately just beyond their current abilities, and then they reward and encourage the players during the process of trying and failing until they eventually succeed.

There are probably a number of ways to design a practical homework-grading system to provide this kind of reward structure, but let me describe one such system that works reasonably well for large-enrollment courses such as the Six Ideas course taught at Washington University in St. Louis. At the beginning of class, students hand in daily assignments consisting of three homework problems, two of which are simpler problems based on the assigned reading for that class, and the third of which is a richer and more realistic problem on material discussed in the previous class session. (This structure rewards students for carefully reading the assigned chapter in the text before class but also ensures that students work on challenging problems after they have received some guided practice in class.) Graders then use a rubber stamp to mark and grade each problem using the rubric shown in Figure 7.

![Rubric Image](image.png)

*Figure 7.*
Graders then assign points concerning whether the problem solution is complete (up to 3 points), clear (up to 2 points), and correct (up to 3 points), as well as one point if the final result is plausible and one bonus point if the initially submitted solution is entirely correct. The grader simply records the numbers in the left slots in the stamped rubric: they do not write any detailed feedback on what the student did wrong (if anything). The process typically takes literally seconds per problem, and graders do not need a great deal of expertise to make a reasonable assessment (at Pomona, we use only lightly trained undergraduates).

When students receive their graded solutions, they can view complete and carefully written solutions online (using an online system devised especially for users of Six Ideas) and use these solutions to correct a certain number of the homework problems they did incorrectly. They then can resubmit these corrected solutions, and the graders will update the scores for the complete, clear, and correct categories (though not the other two), recording the updated scores in the right slots. The student’s final grade on the problem solution is based on the updated scores.

This system has several important advantages over traditional homework grading. Most importantly, it provides detailed feedback (by motivating students to compare their solution to the online solution, and fix whatever was wrong with their solution) without requiring impractical time or unrealistic expertise on the part of the grader. Second, a student who makes a serious initial effort on a problem and then corrects it fully can earn 8 or 9 points out of 10 on that problem, even if the initial effort was entirely wrong. Therefore, this makes it possible to assign challenging problems without discouraging students if they try and fail. In other words, the system rewards practice and provides feedback, while still remaining practical for large-enrollment classrooms.

This is simply one example of a homework system whose design can support a more fruitful course ecosystem. The online instructor’s manual for the Six Ideas project guides professors through the process of adapting or inventing a homework system that can work well at their particular institution.

In summary, the goal of the Six Ideas project is to provide resources and guidance for professors seeking to construct a course ecosystem that evolves genuine student learning of introductory physics. These resources include not only the Six Ideas textbook, but also a lengthy online instructor’s manual (which discusses the issues I have presented in this chapter in much more depth), online (password-protected) serving of problem solutions (managed according to a list submitted by each professor), other computer software to support classroom and homework activities, sample lesson plans, and so on.
Measures of Success

Does the Six Ideas ecosystem do its job? To address this question, I’d like to discuss some data concerning implementations of Six Ideas at various colleges and universities, with special emphasis on the specific implementation of a Six Ideas course at Washington University (WU), a prestigious private university in St. Louis, Missouri. Roughly 750 students take introductory calculus-based physics at WU every year. WU started teaching one Six Ideas-based section in 2004. At present, it offers five sections of the course (Physics 197/198) with about 120 students in each section, making it one of the largest single implementations of Six Ideas in the nation. The student clientele includes pre-med and pre-engineering students as well as potential physics majors. It continues to offer one section of traditional physics (Physics 117/118), making comparisons between the two ecosystems possible.

The reason that the course grew from one to five sections at WU was that students voted with their feet. This was only partly because the original professors were pretty cool (see Figure 8). Mainly, it was because the course early on obtained the reputation of being the better course. (It got to the point, I am told, that an anxious parent called the dean to demand that his or her child be allowed to enroll in one of the Six Ideas sections, because the parent claimed this was essential to the child’s future success.)
However, we know that reputations do not always reflect realities. What might provide more concrete evidence that the course ecosystem does its job?

Several good tools for assessing student learning of physics concepts currently exist. Perhaps the best-known is the Force Concept Inventory (FCI) test, developed by David Hestenes and his collaborators in the early 1990s (Hestenes, Wells, & Swackhamer, 1992). This test consists of a set of multiple-choice conceptual questions carefully designed to examine whether a
student has mastered the world-view of Newtonian mechanics sufficiently to avoid certain beguiling misconceptions. These questions are deceptively simple, so much so that most physics professors looking at the test would confidently predict that their students would score quite well. In fact, even quite able students often do poorly on this test.

In 1998, Richard Hake published an article that used this instrument to evaluate student learning of Newtonian mechanics in a large number of physics courses nationwide (Hake, 1998). The instructors in these courses offered the FCI both before and after instruction. To control for the wide range of average class pretest scores, which ranged from 20 to 70 percent depending on the institution, Hake focused on a measure he called the “normalized gain $g$,” which was the fraction of the possible increase in the average class score that was actually achieved by that class. (For example, if the class’s average pretest score was 40%, and the post-instruction average was 70%, then their gain was halfway between their pretest average and perfection, so the normalized gain in this case would be $g = 0.5$). Hake found that there was a very sharp distinction in the normalized gains experienced in courses with traditional class structures (where $g = 0.23 \pm 0.04$) and those he described as having “interactive-engagement” structures (where $g = 0.48 \pm 0.14$), almost two standard deviations larger.

At Washington University, the FCI has been administered as a pretest and a posttest in both the traditional course (Physics 117) and the Six Ideas course since 2009. The normalized gain for Physics 117 for the 2009, 2010, and 2011 academic years has been completely consistent with Hake’s results for traditional courses, and the normalized gain for the Six Ideas sections ($g = 0.4$) was about twice as high, and well within the range seen in interactive-engagement courses (Hynes, private communication). Class size still has an important effect: at Pomona College, where the course is offered to classes ranging in size from 20-35 students, much more interaction is possible, and we have found that during the period between 2001 to 2006, the normalized gains at Pomona averaged about $g = 0.63$. A professor at The Ohio State University, where the Six Ideas course is offered as a small honors course, reported a gain of $g = 0.72$ in 2001. Both of these scores are at the high end of the “interactive-engagement” range.

At both Pomona College and Washington University, achieving large gains are complicated by the fact that average pre-instruction scores are quite high (in the 60% to 70% range) at these very selective institutions, which means that the difference between a normalized gain of 0.4 and one of 0.7 involves answering just two or three more FCI questions correctly. Since even graduate students have difficulty getting all of the FCI questions right, it is pretty difficult to get good gains with such high pretest scores.

The FCI examines only student understanding of Newtonian mechanics. Study of electricity and magnetism can constitute up to half of a standard introductory physics course. There is no E&M assessment tool that is as well-known or as carefully vetted as the FCI, but at
Pomona, I have used the Brief Electricity and Magnetism Assessment (BEMA) tool developed by Ruth Chabay and Bruce Sherwood (Ding, Chabay, Sherwood & Beichner, 2006). Since students typically enter an introductory physics course with little experience with electricity and magnetism at all, pre-instruction scores tend to be little better than what one might expect from random guessing. Therefore, the measure of interest here is simply the post-instruction score. Chabay and Sherwood report that average post-instruction scores for traditional courses are close to 40 percent. When I offered the BEMA test at Pomona for the Six Ideas course a few years back, my students scored 58 percent. This was true even though the electromagnetism section in the Pomona constitutes only about 40 percent of one semester (to make room for some contemporary physics), whereas traditional courses typically spend as at least twice that much time.

Another important element of class success involves bringing student attitudes about physics more in line with those held by experts. The Colorado Learning Attitudes about Science Survey (CLASS) provides an assessment of success in this area (versions of the test are available for physics, biology, and chemistry classes: see http://www.colorado.edu/sei/class/). This test is administered as a pretest and as a posttest, and results typically indicate a sharp decrease in expert-like attitudes as a consequence of traditional instruction (Adams, Perkins, Podolefsky, Dubson, Finkelstein, & Wieman, 2006). This test has been administered at Washington University since 2009, and results for the traditional section for the period 2009-2011 were consistent with the typical decrease, but for the Six Ideas section, expert-like attitudes (modestly) increased (Hynes, private communication), which is quite unusual.

I would like to close this section with some anonymous student comments from Washington University students, collected from their 2011 evaluations. These comments provide some (albeit anecdotal) evidence that students can perceive for themselves the benefits of a thoroughly redesigned course ecosystem. (Note: “Two-minute problems” are one of the active-learning devices presented in the Six Ideas text.)

•“This course totally changed my perception of the field of physics. In high school, I struggled with physics and found it incredibly dry. After taking this course, not only have I developed a deep understanding of many basic principles in physics, I have actually taken a real interest in the field! One thing in particular that I enjoyed about this course was the use of two minute problems. These were a sometimes welcome break from lecture and brought extra engagement into a large course.”
•“It is the best designed introductory science course I have taken. Everything is well thought out with all details covered.”
•“I feel that the most important part of physics 197/198 is that it imparts real knowledge... I feel as though I have gained a great deal of knowledge of not only equations or problems, but of the actual techniques and thought patterns of a physicist. After speaking with Dr. [...] about research, I realized that our books actually taught physics from a physicist’s perspective, which impressed me... It seems to me that [in
physics 117/118], students go to lecture, and then work on a weekly problem set in order to complete the required task, and then cram for the exam. This process is rather unsuccessful in engaging students and encouraging them to seek out knowledge to gain from the material. In physics 197/198, the daily assignments and involved in-class atmosphere keep students thinking about physics, and they become absorbed in scientific methods. For a course intended for scientists, physics 197/198 is fantastic, and I would recommend it fully.”

Conclusions

My goal in this chapter has been to present arguments on behalf of two fundamental features of the Six Ideas project: that both choosing the appropriate instructional metaphors and designing a directed and coherent course ecology can make an important difference in helping students effectively learn physics.

Doing the latter is not always easy. Flipped classrooms are still unusual, and students may find a redesigned course structure perplexing or even scary. I have found that a crucial part of the process is selling the structure to students at the beginning of the course, describing in some detail why the course is designed the way that it is.

The flipped classroom can also be daunting for the inexperienced professor. Student-centered classrooms feel chaotic and incomplete, while delivering a masterful lecture feels so good and can earn one positive comments from students. Surely a crystal-clear lecture that holds students enraptured is doing its job, right? Research consistently tells us that the answer is (sadly) no; students do not learn much physics from even the most scintillating lecture. Remember the piano teacher analogy! It doesn’t matter how beautifully the piano teacher can play if the students do not practice. Guided practice is what enables learning.

Finally good course ecologies are tricky to design. Just as the health of a real ecology can be surprisingly dependent on seemingly unimportant organisms, I have found that student outcomes and attitudes can be strongly dependent on what seem to be the most inconsequential details. The course designer (whether he or she uses Six Ideas or not) must be attentive to what is happening with the students, do research and assessment to support any observations, and above all, be willing to change the ecosystem on the basis of those evaluations to make it more successful.

My goal with the Six Ideas project has been to make this difficult and daunting task as easy as possible for professors. For much more information see the project website at www.physics.pomona.edu/sixideas/.
References


Di Stefano, R. (1996a). The IUPP evaluation: What we were trying to learn and how we were trying to learn it. American Journal of Physics, 64(1), 49-57. Retrieved January 12, 2013, from http://aip.aapt.org/resource/1/ajpias/v64/i1/p49_s1


12. The Benefits of Cross-Talk: Cognitive Psychologists and Stem Educators from Multiple Disciplines

doi:10.7936/K7BG2KWM
Regina F. Frey and Susan M. Fitzpatrick
http://orcid.org/0000-0003-4034-5778

Washington University in St. Louis

Correspondence:
Regina F. Frey
Washington University in St. Louis
St. Louis MO
gfrey@wuchem.wustl.edu

Introduction

The inaugural 2012 CIRCLE conference and this accompanying book come at an important time when there are many national calls to transforming STEM education. The 2012 report from President’s Council of Advisors on Science and Technology (PCAST, 2012) and numerous other reports (e.g., National Research Council, 2012; Brewer & Smith, 2011; National Science Board, 1996) have issued a variety of recommendations that emphasize several key proposals including the following: implementing empirically-validated teaching methods, engaging students actively in their own learning, exposing students to research thinking and problem solving, and providing students with opportunities to engage in research and hands-on activities and to study real-world problems. These various types of active exposure lead to the development of key process (or professional) skills such as information processing, problem solving, and critical and analytical thinking (Michaelsen et al., 2002; Prince, 2004) that are needed for success in the workplace. These experiences also affect persistence of students in STEM majors; some studies show that such exposure can reduce or eliminate the achievement gap between majority and minority students (Haak et al., 2011; Rath et al., 2007). Hence, using evidence-based active-learning teaching practices in multiple STEM courses could result in diversifying STEM majors, in increasing retention of students in STEM majors, and in increasing the likelihood more students will consider STEM careers after graduation.
As a community, we are still learning and studying the most effective ways to implement the evidence-based pedagogies that are known and evaluate these implementations in order to examine how well these pedagogies improve learning and retention over a diverse set of students, over many STEM disciplines, and over many course levels and course sizes. The 2012 CIRCLE conference provided two unique opportunities for faculty. The first was bringing together cognitive scientists with STEM-education discipline-based faculty focusing on teaching and coursework designed for higher education. Previous collaborations between cognitive scientists and science-and-math educators have typically focused on how novice learners gain knowledge, the knowledge structure of experts, or on experiments carried out with teachers and students in K-12 schools (e.g., Brown & Cocking, 2000). In contrast, the 2012 CIRCLE conference was the first conference that intentionally gathered together for discussion and collaboration cognitive-science researchers focusing on higher-education learning and STEM-education discipline-based faculty focusing on teaching in higher-education.

As a STEM community, we are still learning how to integrate the ideas from cognitive science and the discipline-based education research to best improve our students’ learning and experiences. Cognitive-science researchers study learning and memory under experimental-laboratory conditions. Their studies show that active participation in one’s learning (such as through retrieval practice) results in increased retention and application (Benassi, Overson, & Hakala (Eds.), 2014); however, it is not always clear how to best transfer these results to the classroom environment. Discipline-based education researchers study student learning informed by pedagogical knowledge of their respective disciplines, but their pedagogical methods are often developed and evaluated within their own discipline. Hence, even though the organizers recognized that the shared language needed for fruitful collaborations might not yet exist, the focus for the 2012 conference (and the continuing conversations it was intended to ignite) was to bring together a diverse group of scientists and researchers committed to solving common problems existing both within and across disciplines.

From our experiences at Washington University in St. Louis, we have found continued discussion among cognitive-science researchers and multi-disciplinary STEM faculty to be essential in expanding collaborations among these two groups, spreading pedagogical ideas across STEM fields, and improving student-learning outcomes by engaging students more actively in their own learning. Six years ago, we started a STEM education research group (ERG), which meets weekly in a format that resembles a laboratory-group meeting and brings together a multidisciplinary group representing five areas: 1) faculty who teach undergraduate STEM courses, 2) researchers who conduct research in cognitive science, 3) researchers who conduct research in the learning sciences, 4) researchers who conduct discipline-based education research, and 5) faculty-development experts. For example, our
STEM ERG consists of faculty and staff members interested in research on teaching and learning in STEM, from fields such as Biology, Chemistry, Education, Physics, Psychology, and Biomedical Engineering, as well as The Teaching Center, Cornerstone (our student learning center), and CIRCLE (Fisher & Frey, 2011). Collaborations among these groups have resulted in innovative research projects and in the refinement and evaluation of effective pedagogical practices. Projects include development and assessment of supplemental group-study programs in General Chemistry (Shields et al., 2012), assessment of interactive-engagement methods in general physics (Cahill et al., 2014), and implementation and assessment of a multiple-strategies approach for active learning in introductory STEM courses.

In addition to strengthening the discussion between cognitive scientists and STEM faculty, the 2012 CIRCLE conference also encouraged discussion among faculty from multiple STEM disciplines who study teaching and learning in their disciplines. At this conference, the mix of participants were two-thirds STEM discipline-based education-research faculty or STEM faculty teaching in their discipline and one-third cognitive scientists whose research focuses on applying cognitive-psychology principles to STEM instruction. The STEM discipline-based participants were widely varied and included biology (30%), chemistry (26%), physics (19%), and engineering (10%). Hence, this was a unique opportunity for discussion about teaching and learning among STEM faculty. For example, at most conferences, it is more common for biology faculty engaged in teaching to meet and share information with other biology faculty. Because discipline faculty give presentations at their own professional society meetings and discuss pedagogical developments with faculty from within their own discipline, each discipline has naturally developed pedagogies within their field and are often unaware that similar issues are being addressed and potential solutions have been identified across the disciplinary landscape. Gathering this knowledge and sharing experiences across departments can enrich the curriculum-reform efforts within all disciplines. As Chapter 8 by R. Moog suggests, many of these pedagogical methods can transfer to multiple STEM disciplines.

Encouraging faculty and researchers with these multiple perspectives to meet and discuss their research and experiences in the classroom will enrich cognitive science laboratory research, STEM education discipline-based research, and the practical experiences in the classroom. This conference is expanding the idea that teaching and the study of teaching and learning should be a shared endeavor of the broader interdisciplinary community.

The chapters in the first part of this volume summarize information and strategies for enhancing student knowledge based on years of cognitive science and cognitive psychology research (McDaniel and Roediger). In the second part of this volume, chapters from the discipline-based education researchers focus on curricular pedagogies (Linsenmeier et al.,
Integrating across the chapters from the discipline-based education researchers yields five key ideas that cut across instructional practices in STEM disciplines. The first three key points we present below concern student learning, and the last two key points discuss faculty development and implementation.

**Pedagogical Methods (Curriculum)**

The key curricular ideas include a focus on active learning, which has been promoted for many years to enhance student learning. As a result, a wide variety of methods has been developed and tested. A second focus is to provide opportunities for STEM students to experience inquiry and real-world problem solving. STEM courses and curricula at colleges and universities must fulfill a variety of objectives, including acquiring and retaining foundational knowledge content (knowing “what”), developing process skills or professional skills (knowing “how”), and improving the ability to apply knowledge and skills to authentic, real-world activities. Faculty design courses typically with an emphasis on combining these objectives.

**Key Point 1: Use real-world experiences to develop an understanding of the discipline**

To attract and retain students into STEM and to keep them excited about their anticipated field of study, it is advisable to have first-year students experience activities similar to those encountered in the real-life discipline. Too often, such experiences do not come until junior year. The benefits of authentic STEM experiences include increased student engagement with the material, increased identification with their chosen field, and improved retention. Individual opportunities to carry out supervised research are where students commonly receive such experiences; however, inquiry-based design (or laboratory) courses can also provide an opportunity to engage in real-life experiences. Authentic activities that combine acquiring and applying knowledge to solve problems give students experiences similar to those they will likely encounter in their future careers. These experiences help to develop and sustain a student’s passion or curiosity. Such activities also reinforce how to learn, how to problem solve, and how to work collaboratively.

**Chapter 10** authored by Linsenmeier et al. describes how design courses in engineering can achieve these real-world experiences for students. “Design... is highly integrative. Design both provides a unity to a student’s education and stretches him or her to transfer earlier knowledge to a new situation.” (Linsenmeier et al. Chapter 10) The opportunity to engage in the creative processes required for research or engineering design might also remind students why they were initially interested in these fields and spark their enthusiasm to continue their studies.
However, just as faculty need support and training in the design and management of courses incorporating active learning, students engaging in these real-world experiences in design (or laboratory) courses must also be supported via scaffolding. Students are typically uncomfortable with the uncertainty that is an essential part of being a scientist, engineer, or mathematician. For many students, their prior experiences with academic success meant getting the right answer; therefore, working on problems that have multiple “right” solutions requires new skills and a different mindset. In addition, for many students, learning has been a solitary activity. Effective group work requires a very different set of skills. Groups can be inefficient and the workload can be perceived to be unevenly shared. It is not reasonable to assume that students naturally know how to form and work effectively in groups. Students need to be taught how to work collaboratively by including the development of these skills as an explicitly designed component of the course curriculum. As Linsenmeier writes, “the other challenge in teaching design, which some faculty may not even be completely aware of, is moving students out of their comfort zone of structured courses into the more self-directed, independent and yet team-oriented, uncertain world of design ... Faculty, who are used to uncertainty in their own lives based on research, design, or both, were probably good at taking this step themselves, and may not realize, or forget, the hurdle that it presents for some students.” (Linsenmeier et al. Chapter 10)

Key Point 2: Actively engage students in structured collaboration with peers

Although studies have shown that active learning encourages deeper learning and increased engagement with the material, active-learning methods need to be implemented correctly for the benefit to be achieved (Prince, 2004). Effective active-learning activities need to be well-structured with clear, appropriate learning objectives. Many types of active-learning pedagogies include group work, both in-class and out-of-class (Eberlein et al., 2008). For these pedagogies to be effective, the materials must be complex enough such that members need each other to solve the problems, and faculty must facilitate the group work such that the students are responsible for the construction of each other’s knowledge. For group construction of knowledge to occur, the facilitator must ensure that meaningful discussion occurs and the group interactions need to be structured such that all members have equal opportunity to participate. There are several methods to ensure that discussion and equal participation occurs, such as the use of group roles or the use of collaborative-learning strategies. Such facilitation is complex and requires training. In addition to group work increasing content learning, group work is also an excellent approach to teaching process (or professional) skills such as teamwork and communication. Last, it is essential that faculty explicitly acknowledge and address the issues students may have with active-learning environments (see H. I. Modell, 1996, R. Felder, 2007 and R. Felder, 2011).
One nationally known collaborative-learning pedagogy is Process-oriented Guided inquiry learning (POGIL). Chapter 8 by Moog describes the POGIL philosophy and gives practical advice about how to transform a classroom into a collaborative learning environment. Although POGIL was started in chemistry, the method Moog describes is now used in many STEM disciplines. In addition, this method has been implemented in a variety of ways in a classroom; for example: 1) having no lecture and introducing all concepts through POGIL group work; 2) having a combination of mini-lectures and POGIL group work to introduce concepts; and 3) having each week some sessions predominately lecture and one or two class sessions entirely POGIL group work. The flexibility of implementation allows for instructors to include the POGIL method in a manner that best fits their course’s learning objectives and their own approach to teaching.

One of the main strengths of collaborative learning is teaching students problem-solving skills. As Moog discusses in his chapter, “If one wants students to get better at ‘problem solving,’ then students need to encounter true problems to work on. Many students find that many of the questions in a POGIL activity are problems in this sense: they don’t know exactly what to do to generate a correct answer to the question.” (Moog Chapter 8)

However, in working together and discussing the problems as a group, students begin to think more deeply about the questions and how to solve them. When students start to move from routine solving to developing complex problem-solving skills, many students are uncomfortable, because for the first time students may not be able to quickly arrive at an answer—the more typical classroom goal. However, scaffolding the problems as well as scaffolding the group dynamics allows students to discuss their ideas and bring out their questions in a safe, yet challenging, environment. This ability to communicate one’s ideas in a group is an essential skill that our students need to develop as they progress through their undergraduate careers.

Key Point 3: Think of your course as an “ecosystem”—a whole comprised of integrated parts

Especially in introductory STEM courses, there are many components to making a course successful. These components include the content material, classroom activities, problem sets, grading policies, quizzes or exams, supplementary-support programs, and feedback mechanisms. All of these components work together to improve student learning, and hence it is essential to think about how and in what combination of these components are integrated in your course. However, although instructors know that students do not learn only in class, these extra class activities are rarely integrated into the “course ecosystem” (as Moore calls the linking of these components). If you begin with the view that all of these components work together to improve student learning, it becomes easier to see value in thinking and
planning the integration of these components in your course. Introductory STEM courses contain students who have a wide range of backgrounds, experiences, and interests. To a large extent, the students are taking introductory courses because these courses are required and not because the students see these courses as essential parts of their field of interest. Having multiple ways to meet the students where they are and to encourage them to engage with the material results in improved learning.

In Chapter 11 by Moore, he describes his course and the philosophy behind his course structure. As Moore says, “A Course Ecosystem—Even providing random resources is not enough...every part fills an essential niche and works together to create a life-sustaining whole. If any part is out of balance, the whole suffers.”

**Faculty development and implementation**

The last two key points concern faculty development and implementation of innovations. Active learning has not become the standard for STEM courses, primarily because it often requires instructors to radically change their teaching methods—often in the absence of positive student feedback. One of the clear lessons from the existing literature is that active-learning techniques must be well executed to be effective. Recent meta-analyses of active-learning studies have shown that the pedagogical training of the faculty member in active learning—and hence his or her implementation of active-learning techniques—greatly affects student-learning outcomes (Andrews et al., 2011). Chapter 9 by Ebert-May et al. discusses a study comparing instructors’ self-report of their use of active learning in their classrooms with classroom observation. Chapter 7 by Chasteen and Perkins discusses the need for faculty to be given assistance if sustained changes want to be made throughout multiple courses.

**Key Point 4: Combine faculty’s self-perception with direct observation of their implementation**

Most faculty-development programs focus primarily on introducing the faculty to best practices and current literature on evidence-based pedagogies using the workshop model. However, in the study by Ebert-May (Ebert-May et al., 2011), it was discovered that faculty were not implementing the pedagogies taught them to the extent that they self-reported they were. In fact, the faculty did not modify their pedagogy to any great degree after participating in workshops. As Ebert-May says, “Here we note that faculty were not giving false answers to survey questions; in fact, the majority of faculty were very excited about changing their teaching upon completion of the workshops. Rather, what faculty believed they did in classrooms post workshop is not what we saw...We concluded from our study that
traditional PD workshops alone were insufficient to enable faculty to implement learner-centered classes.”

Therefore, faculty developers need to re-design faculty professional-development programs to not only teach and reinforce faculty knowledge of active learning and successful implementation strategies, but to scaffold and support faculty during the design, implementation, and evaluation of active-learning innovations in their courses. We need to create a feedback loop to our faculty about their implementations to allow them to improve their implementations. We also need to create a support structure and collaborative teaching culture in which faculty can develop, try out, evaluate, and refine new teaching ideas. Faculty need to be implementing these changes in an atmosphere in which they receive support from teaching centers and from peers across STEM departments as well as from colleagues within their departments. Our goal should be to establish a culture where teaching is a shared and collaborative endeavor (not a solitary one).

**Key Point 5: Create structure to support faculty in implementation, evaluation, and refinement**

Two examples of different programs that have attempted to put structures in place to accomplish institutional and sustained changes are: 1) CU Science Education Initiative (SEI) at University of Colorado – Boulder, and 2) WU CIRCLE Fellows Program at Washington University in St. Louis.

(Chasteen and Perkins) in their chapter describe their SEI program whose mission is to support departments to improve STEM education. Perkins describes in the chapter a number of lessons learned, recommendation, and successes. Below are listed a few key points (among many) for institutions to think about when setting up their own initiatives:

1. “Provide a la carte, modifiable course archives, so that faculty can pick and choose materials that suit them, and modify them to their needs, rather than adopting a course approach wholesale.”
2. “Use an STF, or other facilitator, who is somewhat outside of the departmental structure, but still knowledgeable about the content of the discipline, to facilitate the transformation. This enables conversations about teaching and learning that are less influenced by departmental politics, and decreases the overall burden on faculty.”
3. “Department culture changes. Discussions about teaching, learning, and departmental courses are much more frequent among the faculty and are now integrated into formal departmental structures such as faculty meetings.”
4. “Faculty overwhelmingly report that the SEI has had a positive impact on their department...students come to expect the use of interactive techniques such as clickers in upper-level courses.”
At Washington University in St. Louis, we started a CIRCLE Fellows program as a collaboration between CIRCLE (Center for Integrative Research on Cognition, Learning, and Education) and The Teaching Center. A CIRCLE fellow engages long term (two-year minimum) in a learning community where information from all the relevant domains of study is discussed. As they implement and assess their teaching adaptations, they become reflective teachers. These learning communities bring together the CIRCLE fellows, education researchers, cognitive scientists, assessment experts, and faculty-development experts. Each fellow works with a graduate-student intern, The Teaching Center, and CIRCLE to design, implement, and evaluate a teaching innovation. The project group meets bi-weekly during the academic year in a laboratory-research-meeting style format to discuss their projects. They also attend the summer STEM Faculty Institute on Teaching in their first year and then present in their second year. Each CIRCLE faculty fellow receives a small stipend to motivate and reward the faculty member for focusing on these teaching innovations and participating in the learning community over the two years of the project.

**Conclusion (Developing a unified research agenda)**

The conference organizing team and editors of this volume came together on this project as the result of our own trans-disciplinary “cross-talk.” The challenges faced by one of us (SF) as a university neuroscience instructor interested in adapting the more student-centered, active-learning pedagogical approaches discussed throughout this volume is an example of the challenges faculty can face when altering methods of instruction at the postsecondary level. To implement such changes, it is essential to have the support of researchers and faculty developers who have the expert knowledge of learning and the pedagogical approaches as well as the practical experience.

In addition, the willingness of faculty to adopt new methods is, to some extent, influenced by the unspoken norms and expectations students hold about what a college classroom “looks like.” For many students, the traditional college lecture has an almost iconic status. The professor in front of the class with students seated in an auditorium-like classroom is a common portrayal in popular media. Students moving from high school into colleges and universities have an implicit expectation that professors talk, students listen, and an exam is given based on the lecture notes. There is an unspoken “pact” about roles and what is required to succeed. Altering the pact, even when it results in a more dynamic classroom and more engaged learning, requires providing students with explicit explanations (including evidence) for changing the norms, roles, and expectations. Professors, particularly those whose own education was primarily in predominately lecture-based courses and who have not been engaged in research on teaching and learning, will succeed only if there are
opportunities for explicit and ongoing professional development on navigating the transition away from predominately lecture-based courses.

Faculty success at navigating the change to newer instructional methods requires the institutional leadership at universities, colleges, and departments to support the infrastructure for pedagogical research, for translating this research broadly into practice, and for fostering a feedback cycle of continuous improvement. As discussed by Ebert-May et al., faculty are not able to objectively monitor how much or how well they implement the available pedagogical knowledge into their classrooms. Learning communities like ERG and structured support from teaching centers or STEM education centers (as described by Chasteen and Perkins) are essential if we want to broaden participation of faculty to incorporate evidence-based methods into their courses.

Hence, although the first part of this volume provides suggestions for how disciplinary experts might weave the findings of cognitive science into their pedagogical repertoire, Moore’s observations provide an important and interesting set of challenges for cognitive scientists. Discipline-specific cognitive research, such as teaching algebra or the concept of mass, is often built around specific “units” or “modules” (see Koedinger and McLaughlin) and rarely provides guidance for how to build an entire semester or multiple semesters of course work. Similarly, the cognitive psychologists who focus how best to implement decades of experimental work on learning and memory, exemplified by the ideas put forward in the chapters by McDaniel and Roediger (retrieval practice) or Bjork (spacing and interleaving), rarely provide detailed insights into how different approaches can be used together for effective classroom practice. Hence, there are gaps in our knowledge that collaborative work between cognitive scientists and discipline-based STEM education researchers could fill.

For example, one area of partnership between the disciplinary experts and cognitive scientists that could be crucial for success and could help faculty determine effectiveness is the design of dynamic assessments and outcome measurements that provide data about what information students are and are not learning and retaining, where conceptual and methodological errors are occurring in problem solving, and how successfully students are able to generalize and transfer knowledge to novel settings. Other collaborations could derive from how one reconceptualizes the “problem.”

Last, for sustained changes to STEM education, faculty efforts to adopt active-learning pedagogy require institutional leadership at all levels of the organization: departments, schools, and the university. To assist in this institutional change, STEM educators might consider expanding their collaborations to include experts knowledgeable in social and organizational change, particularly when it comes to structuring the supportive infrastructure necessary for initiating, propagating, and sustaining best practices across the
different STEM disciplines. Although there has been relatively little study of change strategies in STEM higher education, enough evidence is available to suggest that many studies of change in other disciplines and environments are applicable. Recent reviews provide some insight (Henderson et al., 2011; Henderson et al., 2010; Borrego, Froyd & Hall, 2010). The literature suggests that presenting best practices to faculty, then expecting those practices to be adopted by a broad community of educators has not resulted in significant and widespread change. Similarly, top-down strategies that seek to modify classroom practice by implementing policies to mandate change are ineffective in achieving sustained change. Third, the lack of cross-talk and the use of different languages of researchers and practitioners in three relevant domains—discipline education, faculty development, and higher-education research—also contribute to the lack of broad-scale change. Henderson and colleagues conclude that working across these domains using strategies that fall into multiple categories (such as disseminating curricula and pedagogy and developing reflective teachers, a shared vision, and institutional policies) will be most effective both in catalyzing change and in developing a better understanding of effective change strategies. They also conclude that lasting change requires long-term intervention. In addition, teaching and research on teaching and learning are most effective when performed within a community. (Shulman, 1993; Fisher & Frey, 2011). We hope this book will help to create this broader multi-discipline community and impact teaching and learning beyond the conference.

References


This volume collects the ideas and insights discussed at a novel conference, the Integrating Cognitive Science with Innovative Teaching in STEM Disciplines Conference, which was held September 27-28, 2012 at Washington University in St. Louis. With funding from the James S. McDonnell Foundation, the conference was hosted by Washington University’s Center for Integrative Research on Cognition, Learning, and Education (CIRCLE), a center established in 2011.