2014

Applying science of learning in education: Infusing psychological science into the curriculum

Victor A. Benassi
University of New Hampshire, victor.benassi@unh.edu

Catherine Overson
University of New Hampshire, catherine.overson@unh.edu

Christopher M. Hakala
Western New England University

Feedback

Feedback regarding the editorial content of this book or any of its essays should be directed toward the individual authors or the book's editors. They (authors and editors) are solely responsible for the substance of the text.

Feedback regarding technical matters of formatting or accessibility of this text via the online environment of the Internet should be directed to the **Internet Editor**. If you have any complaints or difficulties in accessing these materials, be sure to provide as detailed a description of your problem(s) as you can; you should include information about the browser you are using (e.g., Firefox, Safari) and its version number well as the type of computer you are using and its operating system.

Copyright and Other Legal Notices

Suggested Reference Format

We suggest that the overall text be referenced in this fashion:

Benassi, V. A., Overson, C. E., & Hakala, C. M. (2014). *Applying science of learning in education: Infusing psychological science into the curriculum.* Retrieved from the Society for the Teaching of Psychology web site: **http://teachpsych.org/ebooks/asle2014/index.php**

Individual chapters may be referenced in this fashion:
Ambrose, S. A., & Lovett, M. C. (2014). Prior knowledge is more important than content: Skills and beliefs also impact learning. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.). *Applying science of learning in education: Infusing psychological science into the curriculum.* Retrieved from the Society for the Teaching of Psychology web site: **http://teachpsych.org/ebooks/asle2014/index.php**

Applying Science of Learning in Education

Infusing Psychological Science into the Curriculum

Editors
Victor A. Benassi
Catherine E. Overson
Christopher M. Hakala

Division 2, American Psychological Association 2014

Acknowledgments and Dedication

We thank William Buskist, Editor-in-Chief of the STP e-Books series, for suggesting that we prepare this book, and for his support and encouragement throughout the process. We also thank Jeffrey R. Stowell, STP Internet Editor, for his production of the final copy of the book.

Preparation of this book and the work presented in several chapters in the book were supported in part by a grant from the Davis Educational Foundation. The Foundation was established by Stanton and Elisabeth Davis after Mr. Davis's retirement as chairman of Shaw's Supermarkets, Inc. We are appreciative of the confidence the Foundation has shown toward our work. Our special thanks go to Leanne Greeley Bond, our grant program officer.

We also acknowledge the support and resources that have been provided by the Office of the Provost and Vice President for Academic Affairs, University of New Hampshire.

We dedicate this book to the many teachers who participated in the University of New Hampshire's Cognition Toolbox project (2009 – 2013).

e-book cover: Catherine E. Overson.

About the Editors

Victor A Benassi

Victor Benassi is professor of psychology at the University of New Hampshire (UNH) and currently Faculty Director of the Center for Excellence in Teaching and Learning. He is a fellow of the Society for the Teaching of Psychology (APA, Division 2), and was its 2013 President. He was principal investigator of a Davis Educational Foundation (DEF) grant--The Cognition Toolbox: Implementing Cognitive Principles and Assessing Student Learning in College Courses. He is currently principal investigator of a DEF grant: Teaching and Learning with Multimedia (2012 – 2015). Victor is co-editor (with William Buskist) of *Effective College and University Teaching: Strategies and Tactics for the New Professoriate* (2012, Sage Publications). In 2003, he received the American Psychological Foundation's Charles L. Brewer Distinguished Teaching of Psychology award. Earlier in his career, his research interests centered on judgment and psychology of belief. His current research focuses teaching and learning, including the application of science of learning principles to teaching in colleges and universities.

Catherine E. Overson

Catherine Overson, University of New Hampshire, is Project Director of the Center for Excellence in Teaching and Learning's Teaching and Learning with Multimedia grant (funded by the Davis Educational Foundation). She was previously a research associate on the Center's Cognition Toolbox grant (also funded by the Davis foundation). She is an experienced teacher of university courses in psychology, statistics, and the science of learning. She is a member of Division 2 of the Society for the Teaching of Psychology (American Psychological Association) and has presented her research on teaching and learning at annual APA and other national conventions. She has published in the same area (Overson & Goldstein), in *Your graduate training in psychology: Effective strategies for success*, Sage Publications, 2011; Stiegler-Balfour & Overson, in *Effective University and College Teaching*, Sage Publications, 2012). Her research has focused applications of the science of learning in college and university courses.

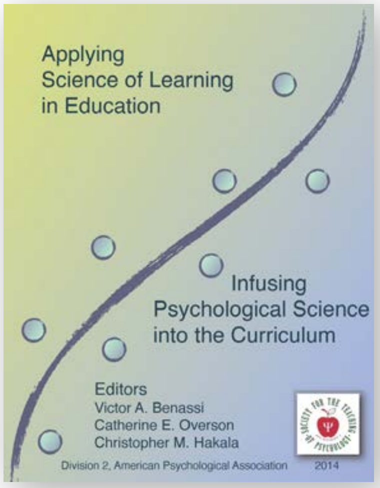
Christopher M. Hakala

Christopher Hakala is Professor of Psychology at Western New England University. He received his Ph.D. at the University of New Hampshire in 1995 and had previously taught at Gettysburg College and Lycoming College, both in Pennsylvania. His areas of expertise are reading comprehension and psycholinguistics. Over the past three years, Chris has split his time between teaching, research and developing the Western New England University Center for Teaching and Learning, and serving as its director. Chris has also served as Associate Editor of Teaching of Psychology and Associate Director of the Office of Teaching Resources in Psychology for the Society for Teaching of Psychology (Society for the Teaching of Psychology, APA Division 2).

Table of Contents

Part 2: Preparing Faculty to Apply Science of Learning

Part 3: Putting the Science of Learning into Practice



Introduction

Victor A. Benassi, Catherine E. Overson
University of New Hampshire

Christopher M. Hakala
Western New England University

What is the Science of Learning?

The field of specialization known as the science of learning is not, in fact, one field. Science of learning is a term that serves as an umbrella for many lines of research, theory, and application. A term with an even wider reach is Learning Sciences (Sawyer, 2006). The present book represents a sliver, albeit a substantial one, of the scholarship on the science of learning and its application in educational settings (Science of Instruction, Mayer 2011). Although much, but not all, of what is presented in this book is focused on learning in college and university settings, teachers of all academic levels may find the recommendations made by chapter authors of service.

The overarching theme of this book is on the interplay between the science of learning, the science of instruction, and the science of assessment (Mayer, 2011). The science of learning is a systematic and empirical approach to understanding how people learn. More formally, Mayer (2011) defined the science of learning as the "scientific study of how people learn" (p. 3). The science of instruction (Mayer 2011), informed in part by the science of learning, is also on display throughout the book. Mayer defined the science of instruction as the "scientific study of how to help people learn" (p. 3). Finally, the assessment of student learning (e.g., learning, remembering, transferring knowledge) during and after instruction helps us determine the effectiveness of our instructional methods. Mayer defined the science of assessment as the "scientific study of how to determine what people know" (p.3).

Most of the research and applications presented in this book are completed within a science of learning framework. Researchers first conducted research to understand how people learn in certain controlled contexts (i.e., in the laboratory) and then they, or others, began to consider how these understandings could be applied in educational settings. Work on the cognitive load theory of learning, which is discussed in depth in several chapters of this book (e.g., Chew; Lee and Kalyuga; Mayer; Renkl), provides an excellent example that documents how science of learning has led to valuable work on the science of instruction.

Most of the work described in this book is based on theory and research in cognitive psychology. We might have selected other topics (and, thus, other authors) that have their research base in behavior analysis, computational modeling and computer science, neuroscience, etc. We made the selections we did because the work of our authors ties together nicely and seemed to us to have direct applicability in academic settings.

Organization of Book

Part 1: Science of Learning—Principles and Approaches

The 14 chapters in this section of the book address important concepts, principles, theories, and research findings related to the science of learning. The reader will notice a similar overall organization in each chapter. We asked authors to provide:

1. an overview of their chapter topic, including definitions of terms and concepts as well as examples of relevant principles and instructional techniques.
2. a description and discussion of the relevant research on the science of learning principle. We asked them to comment on main effects and, if appropriate to their topic, interactions found in research studies. We also asked them to comment on any caveats and boundary conditions. Where possible, we asked them to address studies done with students in real academic settings.
3. advice on when, how, and under what conditions teachers can use the principles they have discussed (for example, experts vs. novices, kind of learning tasks, kind of learning outcomes [facts, concepts, problem solving]).

Several of the chapters in Part 1 present overviews of important general concepts and principles in the sciences of learning and instruction (e.g., desirable difficulties, prior knowledge, metacognition and self-regulation of learning, feedback and its role in the learning process), while several other chapters address a prominent theory of learning (e.g., cognitive load theory). Each of the chapters reviews research on instructional methods and their application in academic settings (e.g., the testing effect, worked examples, spacing and interleaving of practice, self-explanation, the generation effect, techniques to improve learning from textbook assignments). There has been a lot of interest recently in the use in academic settings of *intelligent tutors*, which are based on science of learning. For that reason, we have included a chapter on Project ARA (see also the work being done at Carnegie Mellon University on Cognitive Tutors—www.http://ctat.pact.cs.cmu.edu).

Part 2: Preparing Faculty to Apply the Science of Learning

In Part 2 of the book, we open with a chapter on "assessing the impact of instructional methods" because our intended audience is not restricted to readers with strong backgrounds in research and assessment of learning methods. Next, we present two chapters on approaches to working with faculty on applying the science of learning in their college/university courses—The University of New Hampshire's Cognition Toolbox project and the City University of New York's Bridging the Gap Faculty Development Program. We end Part 2 with a chapter on "Helping Students to get the most out of studying." This chapter could have been included in Part 1 of the book, but we included it in Part 2 because of Stephen Chew's focus on offering teachers a variety of techniques—grounded in cognitive psychology—that they can help students learn and apply during their study of course material.

Part 3: Putting the Science of Learning into Practice

Part 3, the final section of the book, provides six examples of research that has been done in real academic settings and that applied one or more science of learning principles. These chapters provide excellent examples of the science of instruction (Mayer 2011) in practice. The first chapter is written in a different format than the others (Kornell and Metcalfe). This chapter is in the form of a research article that might appear in a scientific journal. Thus, it is more detailed technically than the other chapters in this section. We wanted to include one example in the book of what original scholarly research looks like and we wanted to illustrate how much of the research on the science of learning and science of instruction is programmatic and systematic. In a series of seven experiments, Kornell and Metcalfe examined "The Effects of Memory Retrieval, Errors and Feedback on Learning" in samples of middle-school and university students.

Note that the learning issues and instructional interventions described in the chapters are related to material presented in chapters in Part 1. For example, the research and applications discussed in "Applying Multimedia Principles to Slide Shows" draws in Mayer's "Principles for Designing Multimedia Instruction." Similarly, the chapters describing work done in general chemistry and introductory psychology were informed by research on the testing effect (Pyc, Agarwal, & Roediger).

Suggestions for Reading the Book

As you prepare to read this book, please consider the following guidance.

1. Each chapter was prepared to stand alone, so you do not need to start at the beginning and move through the book. At the same time, many of the chapters are related to one another either directly or indirectly, and you will notice a lot of cross-referencing of other chapters.
2. We instructed authors to write their chapter so that teachers from any field or discipline could read and understand its content. You do not need to be an expert in science of learning to understand and make use of the recommendations presented in this book.
3. We asked authors to keep jargon to a minimum, to avoid technical details on research methods and statistical analyses, etc. Still, readers will notice that some of the chapters include descriptive statistics (Means [*M*s] and Standard Deviations [*SD*s]) and the results of statistical significance testing). Do not be concerned if you do not know the meaning of any of this technical detail. You should be able to understand authors' main points.
4. Our focus is on research that can, and has been, applied by teachers in formal educational settings—face-to-face, online, and hybrid courses (which combine face-to-face and online instruction).

Is the Research Described in This Book Ready for Wide Dissemination and Implementation?

We will not go in to detail here on this question. The December 2012 issue of the *Journal of Applied Research in Memory and Cognition* includes a target article on "Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice" (Roediger and Pyc) as well as several commentaries (Mayer; Daniel; Dunlosky & Rawson; Kornell, Rabelo, Jacobs Klein; Pellegrino). A subsequent article by Dunlosky, Rawson, Marsh, Nathan, & Willingham (2013) in *Psychological Science in the Public Interest* provides an excellent summary of which instructional principles are ready for 'primetime' and which still require further study in real academic settings.

We agree that much more work needs to be done in a wide variety of educational contexts that examine the principles and instructional methods presented in this book (Daniel, 2012; Dunlosky & Rawson, 2013). At the same time, we also agree with Roediger and Pyc (2012) that we know a lot about cognitive principles and that there is every reason to believe that they can be applied to good purpose in the academic courses:

> Do the principles governing learning stop when we switch from a lab to a classroom? All the evidence we know leads us to suspect that generalizations can be made, even though, yes, complexities will arise in the process and some pieces of advice will need to be revised as we learn more. Of course, the data base of research in classroom experiments is not zero, after all, and so far the returns seem promising. What is the downside of applying what we know now, even if the knowledge is not perfect? [p. 263]

Also, consider for a minute the alternative to applying the science of learning in academic setting. On what basis are instructional interventions posited and used? Are they, at least sometimes, based on "opinions, slogans, or quotations from experts" (Mayer, 2011) instead of evidence? We need to learn more about under what conditions science of learning principles have positive effects on students' learning, but we agree with Roediger and Pyc that the "returns seem promising." For example, in our work, we have done many examinations of the testing effect in university courses. With one exception (due to a miscommunication with the instructor about how to implement the intervention), we have found positive results (see chapter on Cognition Toolbox in Part 2).

Other Resources That May be Useful to Teachers

The present volume is similar to other edited science of learning books in that we invited top scholars to prepare a chapter on their specialty (e.g., Sawyer, 2006; Mayer & Alexander, 2001; Zheng, 2009). It differs from those books in that we asked authors to mindfully and explicitly write their chapters with nonexperts as the target audience—teachers who may have little or no background in science of learning, research-based approaches to teaching and learning, or even general principles of psychological science. The books cited above are outstanding, but their contents would be generally inaccessible to a typical teacher of history, political science, English literature, foreign language, etc. Our e-book will not only be accessible to teachers conceptually, it will be freely available through the internet.

There are several books that are written with nonexperts in mind (e.g., Mayer, 2011; Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010). Our book will complement these sources in that we have brought together many of the leading science of learning experts to write about their work and its application to real academic settings.

For readers who wish to delve more deeply into current work being done on science of learning and science of instruction, we recommend the following:

Pittsburgh LearnLab:

http://www.learnlab.org/

http://www.learnlab.org/research/wiki/index.php/Instructional_Principles_and_Hypotheses

Johns Hopkins Science of Learning Institute:

http://scienceoflearning.jhu.edu/

Institute for Educational Sciences:

http://ies.ed.gov/

http://ies.ed.gov/funding/ncer_rfas/casl.asp

The International Society of the Learning Sciences:

http://www.isls.org/index.html?CFID=75710752&CFTOKEN=69962141

Center for Integrative Research on Cognition, Learning, and Education (CIRCLE), Washington University:

http://circle.wustl.edu/Pages/Home.aspx

Human-Computer Interaction Institute Carnegie Mellon University:

http://www.hcii.cmu.edu/

References

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: 7 research-based principles for smart teaching* (1st ed.). San Francisco, CA: Jossey-Bass.

Daniel, D. B. (2012). Promising principles: Translating the science of learning to educational practice. *Journal of Applied Research in Memory and Cognition, 1*, 251–253.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*, 4-58.

Mayer, R. E., & Alexander, P. A. (Eds.). (2011). *Handbook of research on learning and instruction*. New York: Routledge.

Mayer, R. E. (2011). *Applying the science of learning*. Upper Saddle River, NJ: Pearson.

Roediger, H. L., & Pyc, M.A. (2012). Applying cognitive psychology to education: Complexities and prospects. *Journal of Applied Research in Memory and Cognition, 1*, 263-265.

Roediger, H. L. (2013). Applying cognitive psychology to education: Translational educational science. *Psychological Science in the Public Interest, 14*(1),1-3

Roediger, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying Cognitive Psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition, 1*, 242-248.

Sawyer, R. K. (Editor). (2006). *Cambridge handbook of the learning sciences*. New York: Cambridge University Press.

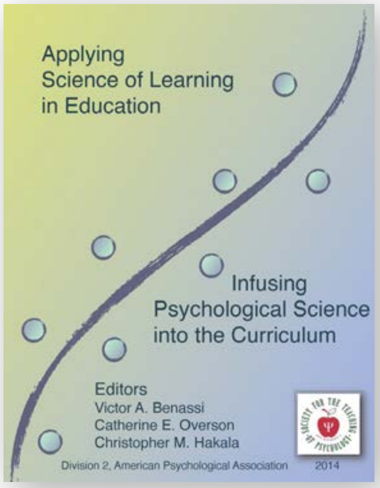
Zheng, R. (2009). *Cognitive effects of multimedia learning*. Hershey, PA: IGI Global. doi:10.4018/978-1-60566-158-2

Part 1

Science of Learning: Principles and Approaches

Division 2,
American Psychological Association

2014



Prior Knowledge is More Than Content: Skills and Beliefs Also Impact Learning

Susan A. Ambrose
Northeastern University

Marsha C. Lovett
Carnegie Mellon University

Introduction

Anyone who has ever taught a course knows that students are not a homogeneous group. They come into our courses with differing levels of knowledge about subject matter content, broad ranges of intellectual and metacognitive skills, and a variety of beliefs and attitudes toward the topic and toward learning. In fact, prior knowledge is one of the most influential factors in student learning because new information is processed through the lens of what one already knows, believes, and can do. In this chapter we define prior knowledge broadly, to include content, skills, and beliefs, because all are "knowledge" in that they result from past experiences and impact subsequent learning and performance. When prior content knowledge is accurate, sufficient, active and appropriate, students can build on that foundation, connecting new content knowledge to already established content knowledge in a framework that will enable them to learn, retrieve, and use new knowledge when they need it (Ambrose, Bridges, DiPietro, Lovett & Norman, 2010). When prior skills – both domain-specific and more general, intellectual skills – are honed, accessed appropriately, and used fluently, they help students to learn more complex skills. And when prior beliefs support behaviors that lead to learning, students' performance is enhanced (Aronson, Fried, & Good, 2002; Henderson & Dweck, 1990). However, students will not have a stable base on which to build new knowledge if their existing content knowledge is distorted, if their skills are inadequate, and/or if their beliefs lead to behavior that impedes learning.

This expanded definition of prior knowledge raises two questions: how can faculty members determine what students know, can do, and believe, and then how can faculty members adapt their teaching to address prior knowledge issues and promote student learning? In this chapter, we focus on helping faculty to better understand (1) different kinds of prior knowledge with which students enter courses; (2) how prior knowledge, skills and beliefs can help or hinder learning; and (3) what strategies faculty members might use to address prior knowledge, skills, and beliefs that are harmful and leverage those that are potentially helpful.

Specifically, we focus on four of the most potent kinds of prior knowledge:

Content-specific knowledge involves knowing what, when, how, and why in a particular domain,

Intellectual skills are the vehicle by which students express, apply, and demonstrate their content knowledge,

Epistemological beliefs focus on the nature of knowledge and learning, and

Metacognition encompasses a student's ability to reflect on and direct his or her own thinking and learning.

In the sections below, we describe key features of these types of prior knowledge, share some research results, illustrate prior knowledge effects in action via hypothetical student cases, and provide corresponding strategies to address prior knowledge differences.

Content Knowledge

Prior knowledge of "content" is the kind of prior knowledge that instructors most often notice – either because students lack critical content necessary to subsequent learning and performance or because students have inaccurate content knowledge that hampers learning and performance. Either way, prior content knowledge relates to specific domain knowledge that involves knowing what (facts), how (skills and procedures), when (conditions under which facts and skills can be applied), or why (connections and relationships between various facts and skills in the domain).

Knowing **WHAT**	Knowing **HOW**	Knowing **WHEN**	Knowing **WHY**
Newton's Second Law	Applying Newton's Second Law when told	Recognizing that Newton's Second Law will help solve a particular problem	Explaining the relationship between Newton's First and Second Laws
Principles of visual hierarchy	Analyzing a poster's visual hierarchy	Applying visual hierarchy appropriately to convey meaning	Relating visual hierarchy principles to how humans naturally perceive

Figure 1. Prior content knowledge includes multiple aspects of knowing: What, How, When, & Why

The accuracy of students' prior content knowledge is critical to teaching and learning as it is the foundation on which new knowledge is built (e.g., Bransford & Johnson, 1972; Resnick, 1983). If students' prior knowledge is faulty (e.g., inaccurate facts, ideas, models, or theories), subsequent learning tends to be hindered because they ignore, discount, or resist important new evidence that conflicts with existing knowledge (Dunbar, Fugelsang, & Stein, 2007; Chinn & Malhotra, 2002). For example, if first-year Physics student Toby mistakenly believes "heavier objects fall faster," he is likely to see what he expects when shown a demonstration of Newton's Second Law. In general, students can significantly benefit from lessons that directly challenge their misconceptions or that leverage accurate conceptions as a bridge to dislodge misconceptions (e.g., Clement, 1993; Hunt & Minstrell, 1994). So, instructors who are aware of their students' inaccurate prior content knowledge can design instruction to target and correct misconceptions.

Even accurate prior content knowledge has important effects on students' capacity to learn new related material. For example, when college students were taught new facts about familiar individuals (i.e., high prior knowledge), they retained twice as much as students who were taught the same number of facts about unfamiliar individuals (i.e., low prior knowledge). In addition, research has shown that students can maximize the potential benefits of accurate prior knowledge if they are prompted to "activate" that information (e.g., Peeck, VanDenBosch, & Kruepeling, 1982; Garfield, delMas, & Chance, 2007). For example, imagine a student, Donna, starting her second semester of statistics, and her instructor asks the class to generate daily-life examples of variability before studying the concept more deeply. By calling these examples to mind, Donna is better able to connect her existing knowledge with new knowledge and hence learn the new material better. In other words, instructors who are aware of their students' accurate prior knowledge can design instruction to connect new information more effectively with what students already know.

Because of the distinct types of content knowledge (Figure 1), instructors will benefit if they are aware that students may have one type of prior knowledge but not others. For example, students may know how to perform various statistical tests, but not know when to employ them. Further complicating matters, students, especially low-performing ones, often overestimate the knowledge they have (see Ehrlinger, this volume, for review; Dunning, 2007). So, instructors might:

> Design and administer a diagnostic performance assessment (e.g., a short quiz or mini-assignment) to reveal students' prior knowledge of material.
>
> Ask students to rate their knowledge by using a descriptive scale that distinguishes cursory familiarity ("I have heard of the term"), factual knowledge ("I could define it"), application ("I could use it to solve problems"), and conceptual understanding ("I could teach it to someone else"). This strategy also reinforces for students the different levels of knowledge.
>
> Look for patterns of error in student work to identify possible gaps or misconceptions in students' prior knowledge.
>
> Activate students' prior knowledge and promote connections by prompting students to think about what they already know about the current lesson's topic. This can occur as a pre-class exercise or an in-class discussion or activity.
>
> Use analogies and examples that connect new material to students' everyday knowledge.
>
> Highlight conditions of applicability (and other heuristics you use) to help students who are missing knowledge of "when."
>
> Address misconceptions by helping students recognize the contradictions in their thinking by creating experiences where they need to apply the new concept/approach/model (rather than just asking them to recite the "correct" answer). Recognize that the more deeply held the misconception, the more time it will take to dislodge, so build in multiple opportunities for students to practice the new way of thinking.

Adjust the level of support in assignments so that more support (e.g., worked examples, annotated cases, step-by-step breakdown of problems) is available for low prior knowledge students and less support (e.g., open-ended, more complex problems) for high prior knowledge students (see Lee & Kalyuga, this volume).

Intellectual Skills

In this chapter, we use the phrase intellectual skills to delineate skills that can be applied across a wide range of different content areas. Pascarella and Terrizini (2005) described these skills as those that help students "process and utilize new information, communicate effectively, reason objectively and draw objective conclusions from various types of data, evaluate new ideas and techniques efficiently, become more objective about beliefs, attitudes and values; evaluate arguments and claims critically; and make reasonable decisions in the face of imperfect information (p. 155)."

These skills are important because they are inextricably tied to the content we teach. To absorb and digest the information they are given in classrooms, students must be able to express, apply, demonstrate, and use their content knowledge through these intellectual skills. Moreover, assessment of students' knowledge is often a blend of domain-specific and domain-general knowledge and skills. For example, students need to be able to write well in order for teachers to clearly assess their knowledge of the reasons behind an historical event. Similarly, students need strong skills in analyzing and interpreting data to define and support a public policy decision.

Intellectual skills form the foundation of the ability to learn throughout life, particularly in today's work environment where professionals will likely change jobs and perhaps professions several times over the course of their careers. Employers across a variety of industries already recognize the importance of these intellectual skills for their companies' continued/future health and success (*It takes more than a major*, 2013; *Job outlook 2013*, 2012; *The role of higher education in career development*, 2012).

In our experience, faculty members often neglect or minimize the development of intellectual skills in students. Some faculty do not believe it is their responsibility to teach these skills, others worry that they do not have the expertise to do this kind of teaching, and still others make an attempt, but do it poorly. We argue here that it is critical for instructors to take on the responsibility of developing intellectual skills in their students, as disciplinary variations embedded within intellectual skills are critical and core to learning and developing expertise.

Faculty attempting to teach intellectual skills will be able to do so more effectively if they teach about explicit components of intellectual capacity to help students become aware of these separate aspects of intellectual skills. This is often a challenging task for faculty because research has shown that experts tend to skip and/or combine steps in a process once it has become second nature to them (Anderson, 1992; Koedinger & Anderson, 1990). But being able to grasp the component parts of intellectual skills is akin to being able to grasp and manipulate the trajectory of each ball when performing a complex juggling act. If students do not have the "microskill" of fluently grasping each ball, the fluid activity of juggling will fail. Similarly, creating an argument requires a set of specific steps that we engage in but often do not explicitly acknowledge: stating a position, examining assumptions, using evidence, refuting other arguments, drawing conclusions, and many other steps. And each of these steps has sub-steps. Learning to quickly and fluidly maneuver among these intellectual moves comes only after years of practice building understanding of individual steps.

Although some teachers believe that intellectual skills cannot be taught, we believe not only that they can be taught but that such skills should be taught explicitly within respective domains. We look, for example, to Harskamp and Suhre (2007) who found that they could improve students' problem-solving skills by intervening through formal and informal hints in a computer program, particularly the aspects of problem-solving that focus on the solution approach (planning) and feedback on the correct solution (verifying). This research was based on Schoenfeld's (1992) work in which he helped students improve their problem-solving skills by asking them both to recall the steps they had taken in solving a problem and to reflect on the next steps, reinforced by class discussion. Research by Garfield, delMas and Zieffler (2012) showed an improvement in statistical thinking – particularly around the modeling and simulation processes – when teachers "scaffolded" instructions (i.e., provided detailed directives at the beginning and then gradually reduced them by the end of the course), provided an illustrated summary and synthesis of both processes, and discussed the ideas in class. Research by Nadolski, Kirschner and van Merriënbuer (2005) also validated how worksheets helped students learn the steps necessary to prepare and complete a law plea. Again, explicit explanations of individual steps in complex tasks is what made instruction valuable for learning in the above studies.

Information about students' intellectual skill set should impact the activities and assessments we design for them. As students move through our respective courses and curriculum, they should apply these skills in progressively more complex and challenging assignments, projects, problems, etc., eventually transferring them to novel/new contexts. Because we need to build on the skills students enter our courses with, and address the gaps when they have not acquired those skills at the level they need to use them, we should be ready with strategies to promote learning and expansion of those skills, such as:

- Give students a performance task related to the skill(s) they will need for the course to gauge their actual level and identify where you need to provide support.
- Deconstruct expert skills for students by creating and teaching the key subcomponents of the skills you want students to learn and use.
- Model the skills you want students to learn by explaining out loud your process for understanding, solving, and reflecting on an activity or assignment.
- Provide guided practice on the skills that are missing or weak.
- Use rubrics in the course to reinforce the intellectual skills and their component parts.
- Ask probing questions to get students to reflect on their learning and the intellectual process they engaged in during their work.

Epistemological Beliefs

According to Billett (2009), students' beliefs about learning arise from their "capacities, earlier experiences, and ongoing negotiations" (p.211) with the world, which certainly can either promote or hinder learning. Many believe that learners' beliefs about their capabilities – often preconscious and often inaccurate– can be more important drivers of behavior than their actual capability. However, research indicates that these beliefs evolve over time, meaning that faculty can influence students' personal epistemologies as they develop (Schommer 1994; Perry, 1970; Belenky, Clinchy, Goldberger & Tarule, 1986; King & Kitchener, 1994; Baxter Magolda, 1992). To have influence, first faculty must recognize students' personal epistemologies as part of their prior knowledge, and then they need to decide when and how to confront beliefs that negatively impact behavior and which alternative perspectives will lead students to behaviors that will result in learning.

There are several epistemological beliefs (Perry, 1970; Schoenfeld, 1983, 1985; Schommer, 1990) that reside on a continuum and are independent of each other (Schommer & Walker, 1995). As we describe these beliefs below, we provide examples of students at the respective ends of the spectrum to illustrate how beliefs influence the behavior that impacts learning and performance.

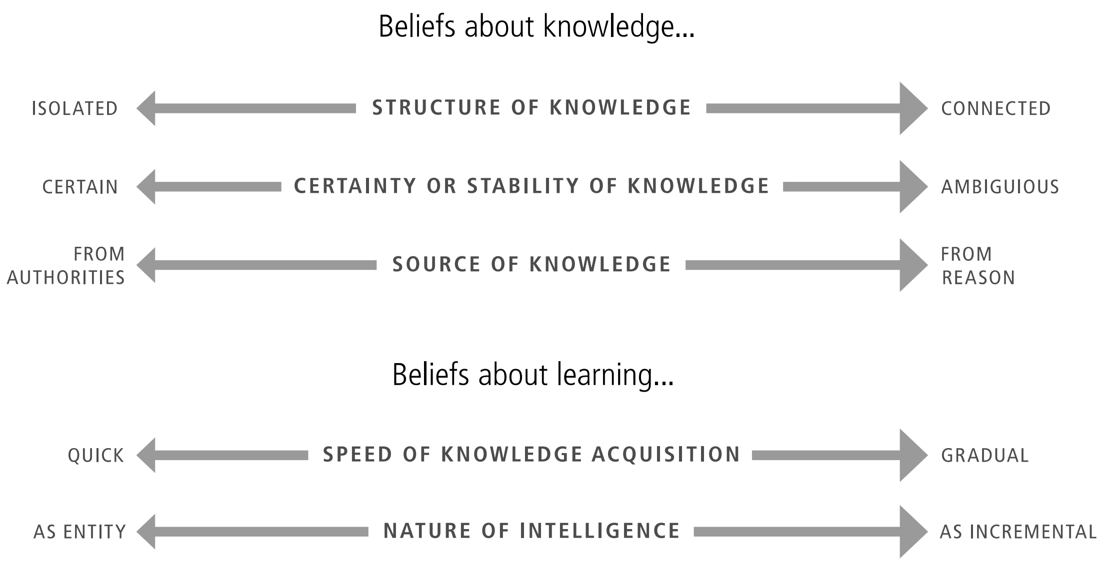


Figure 2. Students' beliefs about knowledge and learning vary along multiple dimensions.

Structure of knowledge: Students can view knowledge as isolated bits of information or integrated concepts and connected structures (Perry, 1970; Schommer, 1990d, 1994a). For example, contrast Toby, who reads assigned articles with the goal of memorizing names, facts, and concepts, with Claudia, who reads for connections and the "big picture."

Certainty or stability of knowledge: Students can also view knowledge either as unchanging and highly certain, or as fallible and highly uncertain (Perry, 1970; Kuhn, 1991). For example, Donna believes all problems have a right or wrong answer, and therefore searches for "the" definitive solution. Josh, on the other hand, welcomes competing theories, accepts the legitimacy of different viewpoints and evolving knowledge and recognizes that context and frame of reference matter.

Source of knowledge: Some students believe that knowledge emanates from the expert authority while others believe it originates from the self, based on reason and empirical evidence (Belenky et al., 1986; Baxter Magolda, 1992). One student, Will, may believe that his teacher and the authors of his textbooks are experts, and that their role is to impart their knowledge to him. He may skip labs because he does not believe he needs to actively engage in learning and he takes the expert's words as "truth." On the other hand, his lab partner Abbie attends labs regularly, enjoys tweaking experimental conditions to see if she can alter outcomes in interesting ways, and believes that empirical study may eventually yield results that challenge current knowledge.

Speed of knowledge acquisition: The speed at which one learns can also impact learning behavior (Hofer & Pintrich 1997; Schoenfeld 1983c, 1985b). If students believe that learning is quick and easy, then, as they read a passage or try to work a problem, they will not persist when they do not "get it" the first time. In high school, Leo quickly and easily finished his assignments, but in college, when he struggles with a problem that he cannot solve in five or ten minutes, he abandons it. Conversely, Kate, because she believes learning takes time, perseveres by rereading the text, reviewing her notes, and seeking out more information and new examples.

Nature of intelligence: Research suggests that students who view intelligence as innate focus on their ability and its adequacy/inadequacy, whereas students who view intelligence as malleable use strategy and effort as they work toward mastery (Schoenfeld, 1983c, 1985b; Bandura & Dweck 1985; Leggett, 1985). Imagine Charlie, who has never achieved high grades in math and therefore believes he will never be good in math. As a result, he does not attempt extra problems, seek out alternative explanations, go to recitation or review sessions, or join a study group. He views those actions as futile given his self-view of "low ability" in math. Nancy, on the other hand, believes that the more she engages with a subject, the more she will learn, regardless of her innate ability. As a result, she asks a lot of questions during class, goes to review sessions, joins a study group, and seeks out her TA and professor when uncertain about something. She believes her efforts will pay off. Each of their beliefs impacts behavior, one in a positive way and the other in a negative way.

These examples illustrate how beliefs that learners hold might meaningfully influence their behavior and, in turn, their learning and performance. Some research (e.g., Schommer 1990, 1993; Schommer, Crouse & Rhodes, 1992) shows that, after controlling for general intelligence, epistemological beliefs predict grade point average and can influence comprehension. In other words, belief can be a more powerful predictor of performance than ability.

Researchers have documented interventions that move learners from beliefs that are simplistic to ones that are more productive. Some of these interventions provided information that conflicted with subjects' own naive beliefs or fostered dissatisfaction with subjects' current beliefs through rebuttals of those beliefs using scientific evidence, with both interventions prompting changes in students' beliefs (Gill, Ashton & Algina, 2004; Kienhues, Bromme & Stahl, 2008). Other interventions that provided students with lessons that focused on the "brain as a muscle" (i.e., intelligence as incremental and improvable) showed that those who received this intervention endorsed incremental beliefs afterwards (Blackwell, Trzesniewski & Dweck, 2007). In another study (Dweck, Dinces & Tenney, 1982), learners who originally believed that intelligence was fixed, adopted goals that reflected a view of intelligence as acquirable after reading passages that described intelligence as a series of incremental changes.

These and other studies indicate that we can create conditions to help students develop beliefs that will enhance their learning. In order to do so, instructors can:

- Administer an epistemological beliefs survey in advance of the first day of class that will provide information to students and the instructor.
- Have a discussion with the class about the nature of knowledge and learning, both in general and specific to your discipline (either based on results from the survey or just in general if you do not administer a survey).
- Model for students how to confront multiple perspectives, theories, solutions, and designs and use evidence to make a decision/commitment to one theory or perspective. Then, give students an exercise asking them to do the same.
- Continually make connections among concepts, principles, and theories, as well as connections among lecture, lab and recitation, or among major sections of the course to help students who believe knowledge is a set of isolated pieces.

- Design assignments that force students to make connections among concepts, principles, and theories through visual representations like mental models or concept maps.
- Build opportunities into your course for students to reflect on the behavior they engage in and how this behavior arises naturally from their beliefs.

Metacognition

Metacognition is the set of processes involved in monitoring and directing one's own thinking (Flavell, 1976; NRC, 2001). Students benefit from metacognition because it enables them to reflect on their own approaches to learning, accurately assess what they do and do not know, and make better choices as a result (e.g., Butler & Winne, 1995; Zimmerman, 2001). Metacognition is critical to learning and performance in any domain, and it becomes increasingly important as students advance through their studies – through the school years, into college, and beyond – and take more and more responsibility for their own learning. For example, compared to high school, students in college are often required to complete more tasks on their own, to learn more independently, and to manage their time and approaches with less support. This kind of work demands that students recognize what they already know that is relevant, identify what they still need to learn, plan an approach to learn that material independently, and monitor and adjust their approach along the way. Faculty may naturally assume that students come into their courses with these and other metacognitive skills. However, research shows that most students lack or are rather weak in metacognition (see below). Given what the research tells us, it is not surprising that one of the major intellectual challenges students face upon entering college is managing their own learning (Pascarella & Terenzini, 2005).

Metacognition is key to being an effective self-directed learner (sometimes referred to as a self-regulated or life-long learner), and yet it is not a monolithic capability. Drawing on the literature in metacognition and self-directed learning, we can identify several steps in the cycle of monitoring and controlling one's own learning (Butler, 1997; Pintrich, 2000; Winne & Hadwin, 1998; Zimmerman, 2001). Figure 3 presents five distinct components, each of which relies on distinct knowledge and skills (adapted from Ambrose et al., 2010):

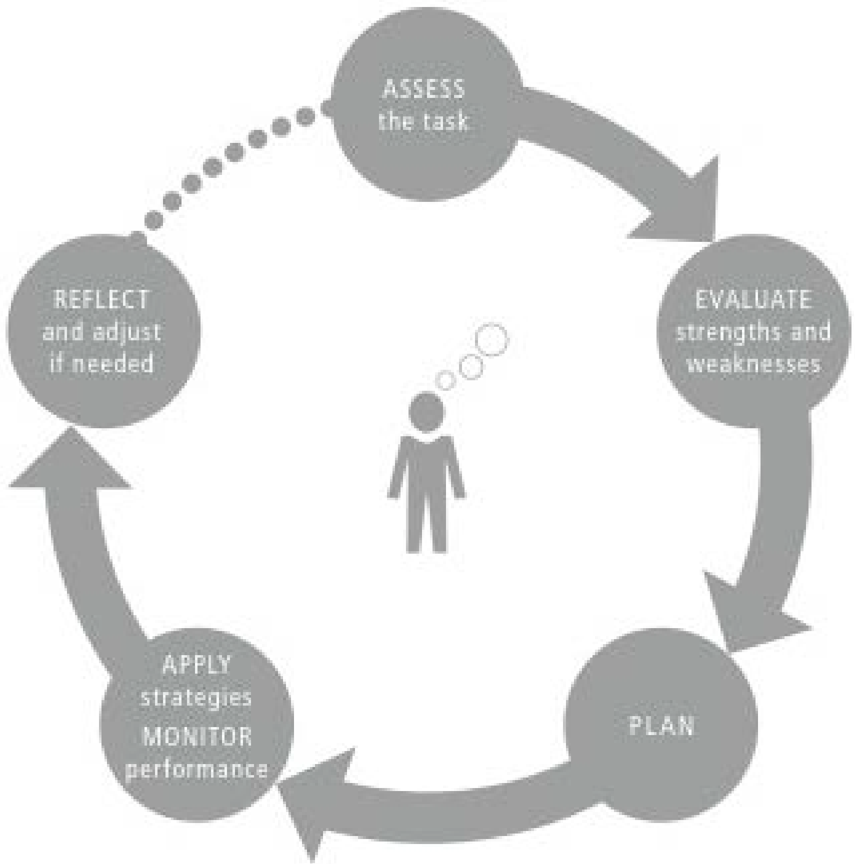


Figure 3. Five components of metacognition that enable effective self-directed learning

Because these are all distinct skill areas, instructors may see a variety of profiles for students' metacognitive prior knowledge. Most important, students may be stronger in some areas compared to others (e.g., Ertmer & Newby, 1996; Winne & Hadwin, 1998). In addition, different learning situations may make greater demands on some of these skills than others. For example, imagine Joey, a first-year physics student, who scored poorly on his first exam only to double-down on his strategy of memorizing formulas by spending twice as much time doing the same kind of studying when it came time for the second exam. When Joey does poorly on the second exam, he may be left confused by the result unless a teacher helps him realize that it is his approach to study that is problematic and that there are other approaches that might yield better results.

Given that students will likely vary in their metacognitive prior knowledge, is there a baseline level that instructors can rely on? Unfortunately, the most likely answer is "no." In a study looking at students' proficiency at assessing academic tasks (step 1 in Figure 3), Carey et al. (1989) found that half of the college students they observed ignored the instructor's instructions for the writing assignment and instead used a generic writing-as-knowledge-telling strategy that they had used in high school. Dunning (2007) has also found that people generally have great difficulty recognizing their own strengths and weaknesses (step 2 in Figure 3), with novices (i.e., students) as the worst judges of their own knowledge and skills. As yet another example, research shows that students – especially novices – are also generally quite poor at planning (step 3 in Figure 3); they often skip the planning process altogether or simply do too little planning for it to be useful (Schoenfeld, 1985; VanLehn, 1998).

The fact that students typically have low levels of prior metacognitive knowledge should not be surprising because metacognition tends to fall outside the content area of most courses. However, teaching metacognition in an abstract or generic way (e.g., in a mini-course entitled “College 101” or “How to be a good student”) is not necessarily the solution, as students in such courses often show difficulty implementing metacognitive skills later in specific situations (Carr, Kurtz, Schneider, Turner, & Borkowski, 1989; Lizarraga, Baquedano, Mangado, & Cardelle-Elawar, 2009; Schraw, Crippen, & Hartley, 2006). In other words, just as with many kinds of knowledge and learning, students do not automatically apply what they learn outside the context in which they learned it, unless a concerted effort is made for students to practice these skills in multiple contexts and reflect on the deep features that will help students recognize when to apply these skills appropriately broadly.

The following strategies thus provide instructors with avenues for addressing students’ prior metacognitive knowledge and skills within the context of typical courses:

- Identify and articulate the specific metacognitive skills you want students to learn to inform them about where to focus their efforts.
- Create (multiple) opportunities for students to practice the targeted metacognitive skills, and provide feedback so students can refine their skills.
- Expose students to metacognitive skills across a diverse range of situations to assure greater ability to generalize the skills.
- Contextualize the metacognitive processes you want students to learn by giving examples and modeling your processes in a way that is tied to the content of your course.

Conclusion

There are three main “take-aways” from this chapter that can be summed up in one sentence:

(1) Prior knowledge plays a critical role in learning, which means that (2) faculty members need to assess the content, beliefs and skills students bring with them into courses and (3) use that information as both a foundation for new learning as well as an opportunity to intervene when content knowledge is inaccurate or insufficient; skills are not adequately developed; and beliefs are interfering with productive learning behaviors.

Understanding students’ prior knowledge can help instructors to create effective learning environments. After all, the only way faculty members can influence learning is to influence the behavior in which students engage. Understanding where that behavior comes from is the first step in effective teaching.

References

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: 7 research-based principles for smart teaching* (1st ed.). San Francisco, CA: Jossey-Bass.

Anderson, J. R. (1992). Automaticity and the ACT theory. *American Journal of Psychology*, *105*, 165–180.

Aronson, J., Fried, C. B., Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology,38*(2), 113-125.

Baxter Magolda, M. B. (1992). *Knowing and reasoning in college: Gender-related patterns in students’ intellectual development*. San Francisco: Jossey Bass.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's ways of knowing: The development of self, voice and mind.* New York: Basic Books.

Billett, S. (2009). Personal epistemologies, work and learning. *Educational Research Review*, *4*(3), 210–219.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*(1), 246–263.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717-726.

Butler, D. (1997). *The roles of goal setting and self-monitoring in students' self-regulated engagement of tasks*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research, 65*(3), 245 -281.

Carey, L. J., Flower, L., Hayes, J., Shriver, K.A., & Haas, C. (1989). *Differences in writers' initial task representations* (Technical Report No. 34). Center for the Study of Writing at University of California at Berkeley and Carnegie Mellon University.

Carr, M., Kurtz, B. E., Schneider, W., Turner, L. A., & Borkowski, J. G. (1989). Strategy acquisition and transfer among U.S. and German children: Environmental influences on metacognitive development. *Developmental Psychology*, *25*, 765-771.

Chinn, C. A., & Malhotra, B. A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology, 94*, 327–343.

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' misconceptions in physics. *Journal of Research in Science Teaching, 30,* 1241–1257.

Dunbar, K. N., Fugelsang, J. A., & Stein, C. (2007). Do naïve theories ever go away? Using brain and behavior to understand changes in concepts. In Lovett, M.C. & Shah, P. (Eds.), *Thinking with data.* New York: Lawrence Erlbaum.

Dweck, C. S., Dinces, N., & Tenney, Y. (1982). Implicit theories of intelligence as determinants of achievement goal choice. *Unpublished raw data.*

Dunning, D. (2007). *Self-insight: Roadblocks and detours on the path to knowing thyself.* New York: Taylor & Francis.

Ertmer, P.A., & Newby, T.J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science, 24*(1), 1-24.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp.231-236). Hillsdale, NJ: Erlbaum.

Garfield, J., delMas, R. C., & Chance, B. (2007). Using students' informal notions of variability to develop an understanding of formal measures of variability. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 117-147). New York: Erlbaum.

Gill, M. G., Ashton, P. T., & Algina, J. (2004). Changing preservice teachers' epistemological beliefs about teaching and learning in mathematics: An intervention study. *Contemporary Educational Psychology*, *29*(2), 164–185.

Harskamp, E., & Suhre, C. (2007). Schoenfeld's problem solving theory in a student controlled learning environment. *Computers & Education*, *49*(3), 822–839.

Henderson, V. L., & Dweck, C. S. (1990). Motivation and Achievement. In S. S. Feldman & G. R. Elliott (Eds.), *At the threshold: The developing adolescent* (pp. 308–329). Cambridge, MA: Harvard University Press.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *American Educational Research Association*, *67*(1), 88–144.

It takes more than a major: Employer priorities for college learning and student success. (2013). *Hart Research Associates*, 1–14.

Job outlook 2013. (2012). *National Association of Colleges and Employers*.

Kienhues, D., Bromme, R., & Stahl, E. (2008). Changing epistemological beliefs: The unexpected impact of a short-term intervention. *The British Psychological Society*, *78*, 545–565.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgement: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey-Bass.

Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, *14*(4), 511–550.

Kuhn, D. (1991). *The skills of argument*. Cambridge, England: Cambridge University Press.

Leggett, E. L. (1985). *Children's entity and incremental theories of intelligence: Relationships to achievement behavior*. Paper presented at the annual meeting of the Eastern Psychological Association, Boston.

Lizarraga, M. L., Baquedano, M. T., Mangado, T. G., & Cardelle-Elawar, M. (2009). Enhancement of thinking skills: Effects of two intervention methods. *Thinking Skills and Creativity, 4*, 30-43.

Schraw, G., Crippen, K., & Hartley, K. (2006). Promoting self-regulation in science education: metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111-139.

Nadolski, R. J., Kirschner, P. A., & van Merriënboer, J. J. G. (2005). Optimizing the number of steps in learning tasks for complex skills. *The British Psychological Society*, *75*, 223–237.

National Research Council (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academies Press.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students* (Vol. 2). San Francisco, CA: Jossey-Bass.

Peeck, J., Van Den Bosch, A. B., & Kruepeling, W. (1982). The effect of mobilizing prior knowledge on learning from text. *Journal of Educational Psychology*, 74, 771–777.

Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart, and Winston.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeider (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego Academic Press.

The role of higher education in career development: Employer perceptions. (2012). *Marketplace and The Chronicle of Higher Education*, 1–86.

Schoenfeld, A. H. (1985). *Mathematical problem solving.* New York: Academic Press.

Schoenfeld, A. H. (1983). Beyond the purely cognitive: Belief system, social cognitions, and metacognitions as driving forces in intellectual performance. *Cognitive Science*, *7*, 329–363.

Schommer, M. (1990d). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, *82*(3), 498–504.

Schommer, M. (1992). *Predictors of epistemological beliefs: Comparing adults with only a secondary education to adults with post-secondary education*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago.

Schommer, M. (1993). Epistemological development and academic performance among secondary students. *Journal of Educational Psychology*, *85*, 406–411.

Schommer, M. (1994). An emerging conceptualization of epistemological beliefs and their role in learning. *Beliefs about Text and instruction with text*, 25–40.

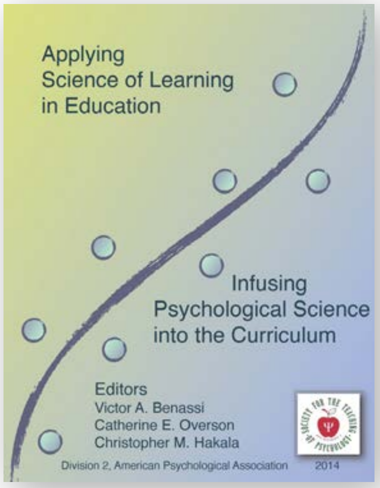
Schommer, M., Crouse, A., & Rhodes, N. (1992). Epistemological beliefs and mathematical text comprehension: Believing it's simple doesn't make it so. *Journal of Educational Psychology*, *84*, 435–443.

Schommer, M., & Walker, K. (1995). Are epistemological beliefs similar across domains? *Journal of Educational Psychology*, *87*(3), 424–432.

VanLehn, K. (1998). Analogy events: How examples are used during problem solving. *Cognitive Science,* 22(3), 347-388.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky & A. Graesser (Eds.), *Metacognition in educational theory and practice.* Mahwah, NJ: Erlbaum.

Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.) *Self-regulated learning and academic achievement* (2nd ed., pp. 1–38). Hillsdale, NJ: Erlbaum.



When and Why Introducing Difficulties and Errors Can Enhance Instruction

Courtney M. Clark and Robert A. Bjork
University of California, Los Angeles

Introduction

Should errors be avoided during the learning process, or are they an important component of effective learning? The merits of errorful and errorless learning have been debated intermittently throughout the history of psychology and education research. Both positions seem plausible: The argument that making and correcting errors is one key to learning seems compelling, but so does the argument that it is better to avoid letting students make mistakes lest their errors impede learning of corrective feedback and/or reemerge at a later time.

Virtues of Errorless Learning

Proponents of errorless learning suggest that learners should be safeguarded against making mistakes. Skinner (1958), for example, when recommending how technology should be used in the classroom, argued that learning should be carefully scaffolded to avoid having learners make mistakes. When learning how to spell the word 'manufacture,' for example, a student would first copy the word and then fill in missing letters in several iterations before writing the entire word from memory. Taking very small steps ensures that learners do not make errors, as each step is trivially more difficult than the last.

Perhaps the most adamant supporters of an errorless learning perspective are learners themselves. Part of the appeal of rereading—either entire texts or texts that have been previously marked—may be that there is no possibility for learners to make errors, given that they are able to review the source exactly as presented, often in whole. Various types of text marking, such as highlighting and underlining text materials, are also common—and errorless—learning strategies used by students (Kornell & Bjork, 2007; Hartwig & Dunlosky, 2012). As teachers, we may also be hesitant to create difficulties for our students. In fact, we may think that it is our responsibility to make learning easy for students and protect them from making mistakes.

Virtues of Errorful Learning

Research on introducing what Bjork (1994) labeled "desirable difficulties" for learners, questions, however, the virtues of errorless learning. Desirable difficulties are manipulations of the conditions for learning or practice that create difficulties and challenge for learners and tend to increase, not decrease, the frequency of errors during the acquisition process, but then often enhance long-term retention and transfer. Overcoming such difficulties, which include spacing rather than massing study opportunities, using tests rather than presentations as study events, and interleaving rather than blocking practice on separate to-be-learned tasks, appears to engage processes that support long-term retention and transfer.

Learning versus performance

To establish why eliciting difficulties is effective, it is useful to think of learning along two dimensions: current performance and "underlying" learning. Imagine a student who is able, on a Monday, to name all of the countries in Europe on a map. When he comes back to school on Tuesday, will he be able to fill in a map again, or will Luxembourg and Malta slip his mind? If he fails Tuesday's test, you would say that Monday's success represented current accessibility (perhaps he had just glanced at the answers seconds before), but not lasting learning. But if he could fill in the map with ease on Tuesday, then again for the final unit test a week later, we would agree that he had a good grasp on the material. Monday's performance, then, could be misleading, because it could reflect either lasting learning or merely current accessibility, which can be quite fragile.

High current accessibility is not, therefore, a reliable index of learning. It is also the case that low accessibility does not mean that learning has not occurred. Students can be learning even when a current test shows little or no progress. One consideration is that memory is cue dependent—that is, what you can remember depends on how many cues or "hints" you have. For instance, given a blank sheet of paper, drawing a map of Europe and labeling each country would be a very difficult task. After a study session on country names and locations in Europe, therefore, a student might do very poorly given a blank sheet, but do very well if the countries were outlined and the task was to fill in each country's name. So despite continued poor performance on a difficult test, improvement on an easier test could show that a student learned something from a lesson. In sum, when we seek long-term learning, current accessibility is not a reliable index of how effective a lesson has been.

The relationship between learning and accessibility

The distinction between accessibility and learning, or performance versus learning, is a time honored one in research on learning (for a review, see Soderstrom & Bjork, 2013). Estes (1955b), for example, pointed out that all learning theorists of that era had to distinguish between what he labeled *habit strength* and what he labeled *response strength*. In a theoretical framework Bjork and Bjork (1992) refer to as a "New Theory of Disuse," which led to the desirable-difficulties idea, they resurrect such a distinction—using the terms *storage strength* (learning) versus *retrieval strength* (accessibility). Bjork and Bjork argued among other things, that when accessibility is high it is difficult to make further gains in underlying learning (i.e., storage strength). For example, repeatedly rereading text or looking at a formula right before implementing that formula may keep accessibility high during the learning process, but such rereading is often in vain (see, e.g., Callender & McDaniel, 2009).

If accessibility can be reduced, however, learners can make bigger gains in underlying learning. Introducing difficulties into learning—thereby reducing accessibility—removes the learner from the experience of having the to-be-learned material at hand. Overcoming the difficulty through subsequent practice is then more effective and can create a new, higher, level of learning.

Decreasing accessibility need not involve errors per se. Accessibility lies on a continuum. Some state capitals, for instance, are likely very accessible for a given individual. If you are from California and someone asks you to name the capital of California, for example, “Sacramento” will pop to mind immediately. The capitals of other states, however, may be known but less accessible. You may know what the capital of Michigan is, for instance, even if it does not come to mind immediately. In searching your memory several cities may come to mind, such as Detroit, Ann Arbor, and Lansing. After pondering for a moment, you may realize that Lansing is in fact the correct city. Decreasing accessibility, then, only means making information harder to retrieve. When training conditions lower accessibility below whatever threshold is needed for a successful retrieval on a particular test, an error will occur, but learning in the storage-strength sense can still remain—as evidenced, for example, on a later test that provides a lot of cue support, or, perhaps, by accelerated relearning.

Three Manipulations that Introduce Desirable Difficulties: Spacing, Interleaving, and Testing

There are certainly ways of making learning more difficult that are not desirable or relevant to the learning process. Furthermore, as we discuss below, research has revealed boundary conditions for the application of techniques that typically create desirable difficulties. We cannot always predict, therefore, which difficulties will be desirable a priori for a given learner or class of learners. In general, for example, having learners generate an answer or solution, versus presenting the answer or solution, imposes a difficulty that is desirable and enhances long-term recall, but if a given learner does not possess the background knowledge to succeed at the generation, the need to generate becomes an undesirable difficulty. In short, the level of difficulty that is desirable can vary across learners and candidate techniques need to be evaluated through careful research both in the lab and in the classroom.

Three practice conditions that consistently produce desirable difficulties using diverse types of materials are spacing, interleaving, and testing. All are techniques that merit being explored in one’s own teaching. Other chapters in this volume elaborate on aspects of such practice conditions. Here, we highlight the ways in which introducing such learning conditions decreases accessibility and increases the probability of errors during learning.

Spacing

Distributing practice over time—that is, spacing study episodes out with breaks in between study sessions or repetitions of the same material—is more effective than massing such study episodes (for a review, see Cepeda, Pashler, Vul, Wixted & Rohrer, 2006; Carpenter, this volume). Massing practice is akin to cramming all night before the test. Teachers may suggest that students start studying earlier before the test so that they will have more time to devote to study, but studying more does not explain the benefits of spacing. Learners who space their practice outperform learners who mass their practice even if both groups spend the same time studying.

Benefits of spacing can be especially pronounced after a delay. That is, cramming can be effective for an immediate test, but spacing practice is usually better for long-term retention, sometimes by a large margin. In general, the optimal interval between study sessions increases as retention interval increases

(Cepeda, Vul, Rohrer, Wixted & Pashler, 2008). Very long spacing intervals, however, where the accessibility of the first study session approaches zero, eventually show limited benefits when compared to a single study session.

One example, among very many, of the benefits of spacing comes from an experiment by Rawson and Kintsch (2005). Participants read a long text passage either once, twice massed, or twice with one week intervening. Participants took a test either immediately after the last study session or two days after the last study session. On the immediate test, studying the passage twice massed led to best recall and comprehension. On the delayed test, however, studying the passage twice spaced out by one week led to superior performance.

How might spacing lead to errors? When learning is spaced, some information from a first study session is forgotten over the delay. When the learner returns to the task, you might imagine that he or she may have forgotten the definition of an unfamiliar symbol or how to implement a geometric formula. Errors may be of omission—merely forgetting what had been learned before—or perhaps of commission if old misconceptions return.

Interleaving

When you are trying to learn several types of information or different aspects of a skill, such as different swings in tennis, there are many ways to schedule your practice. You could block practice, for example, by practicing all of your forehands before moving on to backhands, then serves. You could also mix, or interleave, practice of the various to-be-learned swings. How might such different schedules affect learning?

The surprising answer to that question traces back to a classic study by Shea and Morgan (1979) in which participants learned three different movement sequences in either a blocked or interleaved fashion. The to-be-learned movement sequences each involved releasing a start button and then knocking down a specified set of three hinged barriers in a certain order with a tennis ball as quickly as possible. Participants in both the blocked and interleaved conditions were given 54 practice trials, 18 on each sequence, but in the interleaved condition the 18 trials on a given sequence were intermixed randomly with the 18 trials on each of the other two sequences, whereas in the blocked condition all 18 trials on given sequence were completed before moving on to the next sequence. During the learning phase, blocked practice appeared to enhance learning because the movement times decreased faster and remained faster across trials in that condition than in the interleaved condition. On the final criterion test administered 10 days later, however, participants who had practiced in the mixed condition performed faster, especially if the criterion test was carried out under interleaved conditions.

One important consequence of the fact that blocked practice appears to enhance learning is that instructors and trainers are susceptible to being fooled by students' or trainees' performance during the learning phase. That is, to the extent that performance during instruction or training is assumed to be a valid index of learning, instructors and trainers become prone to choosing blocked practice. Importantly, learners, too, are susceptible to being fooled by their own performance during an acquisition phase. In a study by Simon and Bjork (2001), for example, learners practiced completing specific sequences of key presses that required that a set of keys be hit in a given order and in a given total movement time. As with Shea and Morgan's study, performance on the key press task was better during training in the blocked case. When asked after the training phase how they expected to do on a later test, participants in the blocked condition—expecting their good performance during the training phase would translate into good post-training performance—predicted better performance on a criterion test to be administered 24 hours later than did participants in the interleaved condition. Performance on the delayed test, however, revealed just the opposite: Interleaved practice produced better performance.

Similar benefits of interleaving practice have been found in learning mathematics. After an introductory tutorial to formulas for calculating the volume of four geometric solids, participants in an experiment by Rohrer and Taylor (2007) completed blocked or mixed practice problems. Participants in the mixed condition did all of their tutorials before starting any practice problems, which were then intermixed, whereas participants in the blocked condition completed each tutorial immediately before completing the appropriate block of practice problems. Learners in the blocked condition achieved better performance during the practice phase; learners in the mixed condition, however, performed better on a final test a week later—and by a three to one margin!

Along with benefits of mixing or varying the practice of several different tasks, there seem to be benefits of varying the conditions of practice of a single task. For instance, participants in a massed condition learned more when they studied a lesson and then a paraphrased lesson instead of studying the lesson twice verbatim (Glover & Corkill, 1987). In addition, in the domain of multimedia learning, Yue, Bjork, and Bjork (2013) found that creating small discrepancies—that is, variations—between an auditory narration and on-screen text can be desirable. Participants studied a lesson about the life cycle of a star that included animation, narration, and on-screen text. When the narration was slightly different from the on-screen text, participants learned more than when the narration and text were identical. When asked to predict which conditions would lead to the best learning, however, participants said that having the on-screen text be identical to the narration would be best.

How might interleaved schedules and variation lead to errors? Of course, interleaved schedules also incorporate spacing, while blocked schedules incorporate massing. So in interleaved schedules, we can imagine that forgetting occurs between presentations of one task. We can also think of interleaving as producing "contextual interference" (for a review, see Lee, 2012). That is, when different tasks are mixed together, the learner might experience confusion between them, especially if they are very similar. When calculating the volume of a particular solid, for instance, a learner may use a formula appropriate for a different, but similar solid. When a single task or set of material is studied several times under varied conditions, interference might again come into play, requiring the learner to reconcile differences between the repetitions of material.

Testing

It is a commonly held belief that the purpose of testing is to assess what students do and do not know. Based on practice-test results, students decide what to keep studying, and based on test results, teachers can assign grades or decide what to cover in a follow up lesson. Although these are certainly useful benefits of testing, retrieval also has more direct benefits on learning (for reviews, see Roediger & Karpicke, 2006a; Pyc, Agarwal, & Roediger, in this volume). We can think of retrieval as a "memory modifier" (Bjork, 1975) in the sense that retrieving information strengthens that information, making it more likely to be recalled later.

In a particularly striking example of testing as a learning tool instead of as an assessment, participants in an experiment by Roediger and Karpicke (2006b) either studied a passage four times or studied it once and then took three free recall tests. Notably, no feedback was given after the tests. On an immediate test, participants in the study only condition outperformed those in the "study-test" condition. On tests delayed by one week, however, participants in the study-test-test-test condition performed much better than the study-study-study-study condition.

How might testing lead to errors? In most cases, it is unreasonable to expect that learners will achieve perfect performance on a test, especially if the test is taken early in the learning process. Sometimes learners will make errors of commission. Given that successful retrieval strengthens memory, we might expect that generating an error would strengthen it and lead it to be "resistant" to correction by feedback. Other errors on a practice test will be errors of omission. If feedback is not provided, learners are missing an opportunity to rehearse the complete set of correct information. Restudying, on the other hand, does not introduce errors or have learners produce errors, but—as demonstrated by Roediger and Karpicke, (2006b), and multiple other findings—also often produces poorer long-term recall.

Errors and Learning

Because spacing, interleaving, and testing result in a net positive gain in learning, we know that learners are, for the most part, able to overcome any detrimental effects of erring typically elicited by the techniques – or at least that the benefits of successful retrieval outweigh the costs of erring. But are there really any costs or risks associated with making errors? Or, might it be that some strategies are effective *because* they lead to errors?

Recently, researchers have gained some leverage on these questions by looking more directly at the role of errors and feedback in learning. This work is important because most work on erring suffers from a natural confounding between errors and difficulty. Imagine that a student gets six questions out of ten correct on quiz, and we want to see how the effects of making those four errors changes his ability to learn from feedback and perform on a final test. Those four errors are not randomly distributed. Instead, they represent the content that was most difficult for the student to understand. In effect, then, we are studying how errors operate on the learning of difficult material only, instead of the effect of errors themselves. In some of the following studies, all responses are errors, (which avoids confounding difficulty and erring). In other studies, participants give a mix of errors and correct responses, more closely mirroring classroom situations.

Pretesting and generation-enhanced encoding

Is guessing before receiving the correct answer better than just studying the correct answer? Participants in an experiment by Kornell, Hays and Bjork (2009) either studied associated word pairs intact (*whale-mammal*) for 13 seconds or were first required to produce a guess (*whale-__?__*) for 8 seconds before receiving feedback (*whale-mammal*) for 5 seconds. (The pairs were chosen to insure that participants' guesses would virtually always be wrong, and the 3% of instances when they guessed the correct to-be-learned association were eliminated from the analysis). On a final cued-recall test, pairs for which learners had to guess first were remembered better than pairs that had been studied intact. These results are surprising not only because guessing took up time that learners could have used to study the correct pairing, but also because by guessing incorrectly participants presumably strengthened the incorrect pairings in memory. That is, given the role of retrieval as a memory modifier, where retrieval strengthens information, we might have expected self-generated errors to be particularly harmful and to then create proactive interference when the participants were subsequently trying to retrieve the correct response. Participants, in fact, seem to adopt such reasoning: When Huelser and Metcalfe (2012) asked participants to say—at the end of an experiment that replicated Kornell et al.'s (2009) procedure and findings—whether guessing first or simply studying the correct answer was best for learning, the participants said, overall, that the study-only condition was superior for learning.

Even if they fail to predict benefits of answering pretest questions, students may think that pretests can provide good clues about what to attend to in the coming lesson. In one study, Richland, Kornell, and Kao (2009) sought both to assess the effect of pretesting and parse out its effectiveness above and beyond just letting students know what will be tested. Students either read text passages with key facts marked, took a pretest before study, memorized pretest questions before study, or studied the passage for a longer time. Despite getting nearly every pretest question wrong, students in the pretesting group outperformed all other groups on a final test, including those students who had memorized the pretest questions. Beyond just indicating what will be important, then, pretests can guide effective encoding.

Hypercorrection of errors

How might the type of error learners make affect their ability to correct it? One might imagine that errors made with high confidence would be especially resistant to feedback, but research on the *hypercorrection effect* suggests otherwise (e.g., Butterfield & Metcalfe, 2001). In this line of research, participants are asked a series of questions and also asked to give their confidence in each of their answers. Feedback is provided after incorrect responses. Participants are then given a final test for all facts. Surprisingly, participants are most likely to correct errors about which they were highly confident on the initial quiz. This pattern may be the result of learners devoting more attention to feedback when they discover that their confident answer was wrong (Fazio & Marsh, 2009). It may also be the case that when learners make errors with high confidence, the correct response was the learner's second guess (Metcalfe & Finn, 2011).

Productive failure

Recent research suggests that introducing difficulties and errors can be very effective in the classroom. In a study by Kapur and Bielaczyc (2012), for example, students were assigned to either a "productive failure" group or a "direct instruction" group. Students in the direct instruction group completed typical lessons on complex math problems. Their teacher helped them successfully solve problems along the way. In contrast, students in the productive failure group were given complex problems and then worked in groups with fellow classmates to attempt the problems. The problems were very difficult, and the productive failure groups were unable to solve them. During a final lesson, a teacher helped the productive failure group analyze their failed attempts and provided correct methods. On a final test, the productive-failure group outscored the direct-instruction groups on both complex problems as well as more straightforward problems. Another important finding is that groups of students who had been able to generate multiple approaches or ways to attempt the problems during learning showed larger benefits than did groups who did not generate multiple approaches.

Boundary Conditions and Cautions

Why might some difficulties fail to be desirable? Gaspelin, Ruthruff and Pashler (2013) recently tested the hypothesis that dividing attention during an interim test may serve as a desirable difficulty that could be easily implemented. Although the tone-counting task used to divide participants' attention slowed their ability to respond to a cued-recall test on Swahili-English word pairs, dividing attention did not improve memory on a final test two days later. Gaspelin et al. posited that participants may have been quickly switching back and forth between the tone-counting and retrieval tasks, but that the retrieval itself was not affected. If this is the case, their results support the notion that a desirable difficulty is one that reduces accessibility in memory. As mentioned earlier, spacing allows accessibility to decrease over time, and interleaving reduces accessibility through both spacing as well as the induction of contextual interference. Retrieving information does not reduce accessibility per se. The act of retrieval does, however, take place in the absence of the to-be-learned material, as compared to restudying,,where accessibility remains high and is perhaps even increased

It is important to emphasize, again, that, certain difficulties may also be undesirable if they are too difficult for a given learner. McDaniel and Butler (2010) offered a contextual framework for considering when difficulties are desirable. In addition to considering the learning technique itself, they considered characteristics of the learner, the materials at hand, and how the learner is tested. For instance, if the learner already has trouble decoding words, adding the additional difficulty of filling in missing letters in words may not be effective (McDaniel & Butler, 2010). Again, a basic point is that for a difficulty to be desirable, a given learner must, by virtue of prior learning or current cures, be able to overcome that difficulty.

Can desirable difficulties be combined? Although there is still much work to be done to answer that question in detail, some researchers have looked at the effects of simultaneously implementing more than one desirable difficulty on the same materials. Appleton-Knapp, Bjork, and Wickens (2005), for example, examined the extent to which participants' memory for advertising materials would be enhanced or impaired by combining spacing and variation. They found that up to a certain level of spacing, introducing variation added additional benefits, but that at long spacing intervals the effects of spacing and variation were sub-additive. It may be the case that to the extent that a second presentation, whether it be an exact or varied repetition, manages to trigger a retrieval or reminding of the first presentation, the learner will benefit. According to that view, the subadditivity Appleton-Knapp observed is a consequent of exact repetitions being able to trigger reminding at greater spacing intervals than can varied repetitions.

Concluding Comments on Bringing Difficulties and Errors to the Classroom and to Students' Self-regulated Learning

With so much to learn, students crave methods that make learning fast and easy. One basic message of this chapter is that methods that improve learning typically make learning seem harder, not easier. However nice it might be to have a magic wand to offer teachers and students, one that can get students to learn a foreign language in two weeks or become competent pianists after only a few lessons, what we offer, instead, are difficulties, but difficulties that promote durable and flexible long-term learning and transfer.

As indicated in the name itself, desirable difficulties tend to be difficult and require more effort and planning than students like to give. Furthermore, introducing difficulties can lead to errors. Metacognitive judgments often reflect that learners prefer to use less effective strategies—like cramming right before a test. Students need to know how to schedule their learning, especially when completing online classes that require extensive independent work. Students can improve their efficiency by knowing both what and how to study—and when to stop to studying when it is futile or no longer needed.

The desirable-difficulties framework can offer clear advice to teachers and students. When scheduling lessons for important material, teachers should plan to revisit the material several times in a spaced fashion. Although textbook problems are often blocked by concept, which the recent research on blocking versus interleaving suggests may be misguided, teachers can help students interleave practice by assigning a few problems from each section. In a subsequent lesson, all problem types can then be revisited. Low-stakes quizzing has also been shown to be effective in classroom settings (e.g., Roediger, Agarwal, McDaniel & McDermott, 2011). Among other things, such as the benefits of retrieval itself, taking interim tests may benefit learning by reducing mind wandering (Szpunar, Khan & Schacter, 2013).

Students should also implement techniques that induce desirable difficulties during their independent study. A recent analysis of study strategies identifies practice testing, interleaving practice, and distributing practice some of the most effective methods for learning (Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013). Students might, for instance, take a practice test without looking at any of the answers until the end, or try to write out their notes from memory. When the goal is long-term learning, as it usually is, cramming and passive learning are to be avoided. Also, as emphasized by Bjork, Dunlosky, and Kornell (2013) in a recent review of research on self-regulated learning, students need to become metacognitively sophisticated as learners, which includes knowing not only how to manage their own learning activities, but also knowing how to monitor the degree to which learning has or has not been achieved.

To what kinds of learning/materials do desirable difficulties apply? Students are required to learn information at various levels of analysis, from relatively simple facts to more complex, concept-laden passages. With appropriate modifications in implementation, the principles described tend to be quite general, applying to both more traditional school-based learning and even motor skills (Schmidt & Bjork, 1992).

Above all, the most basic message is that teachers and students must avoid interpreting current performance as learning. Good performance on a test during the learning process can indicate mastery, but learners and teachers need to aware that such performance will often index, instead, fast, but fleeting, progress.

References

Appleton-Knapp, S., Bjork, R. A., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes and their interaction. *Journal of Consumer Research, 32,* 266-276. doi: 10.1086/432236

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123-144). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp.185-205). Cambridge, MA: MIT Press.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64,* 417-444. doi: 10.1146/annurev-psych-113011-143823

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1491. doi: http://dx.doi.org/10.1037/0278-7393.27.6.1491

Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology*, *34*(1), 30-41. doi: http://dx.doi.org/10.1016/j.cedpsych.2008.07.001

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354-380. doi: http://dx.doi.org/10.1037/0033-2909.132.3.354

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning a temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095-1102. doi: 10.1111/j.1467-9280.2008.02209.x

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4-58. doi: 10.1177/1529100612453266

Estes, W. K. (1955b). Statistical theory of distributional phenomena in learning. *Psychological Review*, *62*, 369-377.

Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review*, *16*(1), 88–92. doi: 10.3758/PBR.16.1.88

Gaspelin, N., Ruthruff, E., & Pashler, H. (2013). Divided attention: An undesirable difficulty in memory retention. *Memory & Cognition*, *41*(7), 978-988. doi: 10.3758/s13421-013-0326-5

Glover, J. A., & Corkill, A. J. (1987). Influence of paraphrased repetitions on the spacing effect. *Journal of Educational Psychology*, *79*(2), 198-199. doi: http://dx.doi.org/10.1037/0022-0663.79.2.198

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, *19*, 126–134. doi: 10.3758/s13423-011-0181-y

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, *40*(4), 514-527. doi: 10.3758/s13421-011-0167-z

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45–83. doi: 10.1080/10508406.2011.591717

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 6,* 219-224. doi: 10.3758/BF03194055

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35,* 989-998. doi: http://dx.doi.org/10.1037/a0015729

Lee, T. D. (2012). Contextual interference: Generalizability and limitations. In N.J. Hodges & A.M. Williams (Eds.), *Skill acquisition in sport: Research, theory, and practice* (2nd ed) (pp. 79-93). London, UK: Routledge.

McDaniel, M. A., & Butler, A. C. (2010). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork* (pp. 175-199). New York: Psychology Press.

Metcalfe, J., & Finn, B. (2011). People's hypercorrection of high-confidence errors: Did they know it all along? *Journal of Experimental Psychology. Learning, Memory, and Cognition,* 37(2), 437–48. doi: http://dx.doi.org/10.1037/a0021962

Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology*, *97*(1), 70-80. doi: http://dx.doi.org/10.1037/0022-0663.97.1.70

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning?. *Journal of Experimental Psychology: Applied*, *15*(3), 243-257. doi: http://dx.doi.org/10.1037/a0016496

Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181-210. doi: 10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255. doi: 10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*(4), 382-395. doi: http://dx.doi.org/10.1037/a0026252

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*(6), 481-498. doi: 10.1007/s11251-007-9015-8

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(2), 179-187. doi: http://dx.doi.org/10.1037/0278-7393.5.2.179

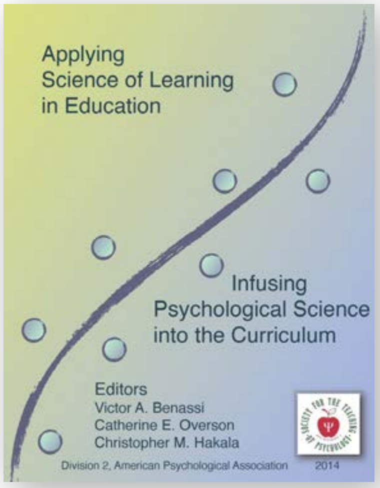
Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(4), 907-912. doi: http://dx.doi.org/10.1037/0278-7393.27.4.907

Skinner, B. F. (1958). Teaching machines. *Science*, *128*, 969 – 977.

Soderstrom, N. C., & Bjork, R. A. (2013). Learning versus performance. In D. S. Dunn (Ed.), *Oxford bibliographies online: Psychology.* New York: Oxford University Press. doi 10.1093/obo/9780199828340-0081

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, *110*(16), 6313-6317. doi: 10.1073/pnas.1221764110

Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology, 105* (2*),* 266-277. doi: http://dx.doi.org/10.1037/a0031971



Expertise Reversal Effect and Its Instructional Implications

Chee Ha Lee and Slava Kalyuga
School of Education, the University of New South Wales

Overview

One important concern about learning is not only whether a set of learning goals is achieved but also whether the learning environment is conducive to meeting goals efficiently–for example, with optimal expenditure of mental resources. Taking into account the contemporary view on human cognitive architecture, cognitive load theory has developed numerous empirically verified instructional principles and techniques for enhancing learning by optimizing learners' cognitive resources (known as cognitive load effects, see Sweller, Ayres, & Kalyuga, 2011, for an overview). The suggested instructional techniques and procedures prevent learners from overloading their working memory by eliminating wasteful cognitive load that disrupts learning or is not essential for learning, thus leaving sufficient working memory resources for essential learning activities that support learning.

Empirical studies conducted in a variety of learning disciplines have demonstrated that the effectiveness of a particular teaching technique depends on the characteristics of learners' cognitive processes that vary with the changes in their domain-specific knowledge base. An instructional technique that reduces working memory load for low-knowledge learners (e.g. by providing strong instructional support) might lose its relative effectiveness and may even bring negative consequences for more knowledgeable learners (*expertise reversal effect*). More knowledgeable learners may learn better in an environment without guidance. In contrast, an instructional procedure that is beneficial to advance learners could inhibit learning for low-knowledge individuals. The effect is explained by the cognitive overload that learners may experience with a particular technique at a specific expertise level.

In this chapter, we will first outline the basic assumptions of human cognitive architecture and cognitive load theory to provide a theoretical explanation for the expertise reversal effect. Then, specific examples that demonstrate how learner levels of expertise interact with some of the cognitive load effects will be briefly discussed. Last, we will discuss instructional designs implication of the effect for optimizing learning by considering learners' domain specific prior knowledge levels.

Human Cognitive Architecture

Our complex cognitive system consists of working memory as our major information processor, and long-term memory for permanent storage of organized knowledge. Information received from the environment via sensory memory that is within the focus of attention is processed consciously in working memory for possible storage in our long-term memory. Working memory is limited in duration and capacity and only several chunks of information are processed simultaneously (Cowan, 2001; Miller, 1956; Peterson & Peterson, 1959). We find it easy to perform simple multiplication like 12 x 12, but difficult to calculate 1212 x 1212 because working memory easily loses information necessary for complex mental calculations. Simultaneously processing too many interacting pieces of novel information can easily overload working memory (Sweller, 2010). If these pieces of information are essential for understanding, then meaningful learning might not occur.

The other essential component of human cognitive architecture is long-term memory which appears to have unlimited capacity and duration (Atkinson & Shiffrin, 1968; Baddeley, 2004). The organized knowledge structures (schemas) that are stored permanently in long-term memory effectively function as a means for overcoming working memory limitations and are central to human cognition (Chase & Simon, 1973; De Groot, 1965). Learning is a process of active reconstruction of existing schemas. During learning, relevant schemas stored in long-term memory are activated and integrated with incoming information in working memory resulting in the construction of more advanced knowledge and skills. The available schematic knowledge also directs and governs higher-level cognitive performance by playing an essential role in identifying a current problem situation, generating routine moves and problem solving strategies. For example, with available knowledge for 25 x 4 as a single processing unit, the mental calculation of 25 x 25 x 4 becomes much simplified. Multiple units of information that need to be processed individually in working memory by a less knowledgeable learner could be encapsulated into a single entity (a schema) for an advanced learner, thus leaving sufficient working memory resources for complex task operations. The working memory load experienced by a particular learner for a specific task depends on the available knowledge structures that effectively determine characteristics of working memory—that is, its content, capacity, and duration (Ericsson & Kintsch, 1995). Advanced learners might therefore experience less cognitive load than less knowledgeable learners in performing high-level cognitive tasks because of this chunking effect. When relevant schemas are unavailable, novice learners have to deal with many new elements of information and might approach the problem by using general search-based strategies such as means-ends analysis or trial-and-error attempts that demand excessive cognitive resources (Ayres, 1998; Kirschner, Sweller, & Clark, 2006). Processing limitations of working memory cause learners' cognitive overload and result in minimal learning.

Managing Cognitive Load for Optimal Learning Effects

Cognitive load is the theoretical concept that is critical to cognitive load theory in generating new instructional procedures (cognitive load effects) and providing the possible explanations for the effectiveness of these procedures (Sweller, 1994, 1999; Sweller, van Merriënboer, & Paas, 1998). Cognitive load is the demand for working memory resources of a particular learner by specific cognitive tasks or activities. The major concern of the theory is to optimize the consumption of learners' working memory resources. A most recent model of the theory has suggested two types of cognitive load (Kalyuga, 2011). The *intrinsic load* is productive cognitive load determined inherently by the intellectual complexity of learning materials or activities that are required for achieving a specific learning goal by a particular learner. This load is associated with essential learning activities like comprehending the learning tasks, establishing connections between essential interacting elements of information, and constructing new schemas incorporating novel elements of information etc. (Paas, Renkl, & Sweller, 2003).

The intrinsic load is also associated with learning strategies that are used for fostering learning to achieve higher-level goals. For instance, engaging self-explanations to enhance understanding solution steps (Atkinson, Renkl,& Merrill, 2003; Wittwer & Renkl, 2008), studying high variability worked examples to facilitate the acquisition of flexible schemas and transfer of knowledge (Paas & van Merriënboer, 1994), and, using imagining procedures to assist schema automation (Cooper, Tindall-Ford, Chandler, & Sweller, 2001). Engaging these activities might increase the complexity of a learning task by raising the number of interacting elements of information and thus render a higher intrinsic load.

As intrinsic load directly contributes to learning, it should possibly be maximized for motivating and guiding learners to higher-level skills. However, intrinsic load should also not exceed a learner's cognitive limits determined by working memory capacity. There are many ways to reduce intrinsic load if the learning goal is too difficult or too complex for a learner to achieve. For example, Pollock, Chandler, and Sweller (2002) demonstrated that novices benefited from a design in which the researcher initially presented complex materials as a sequence of smaller isolated units of information and delayed learning the relations between the units to a later stage. Clarke, Ayres, and Sweller (2005) suggested a similar technique that pre-trained technology-inexperienced learners in essential skills for using spreadsheets in learning mathematics concepts. This approach was more effective than learning the new skills concurrently with the mathematics concepts. The technique prevented too many elements of information to be processed simultaneously in working memory.

Contrary to the productive intrinsic cognitive load, extraneous load caused by undesirable instruction design requires learners to engage in cognitive activities that are not related to learning and should always be minimized (Kalyuga, 2011; Sweller et al., 1998). Processing redundant information is an example of a typical source of extraneous cognitive load that could be easily generated in the context of multimedia instruction. For example, presenting a diagram or picture with the same spoken and written explanatory verbal information concurrently might force learners to devote additional cognitive resources to cross-referencing this redundant verbal information which is extraneous to learning (Kalyuga, 2000; Kalyuga, Chandler, & Sweller, 1999; Mayer, Heiser, & Lonn, 2001). Likewise, captions may hinder the acquisition of essential listening comprehension skills by second language learners due to extraneous load (Diao, Chandler, & Sweller, 2007; Diao & Sweller, 2007).

Intrinsic and extraneous types of cognitive load are additive and their combination determines the required working memory resources. These two types of load compete for limited working memory resources during learning activities. Cognitive resources consumed by extraneous load might leave insufficient working memory capacity for dealing with intrinsic load that is essential for learning. Therefore, efficient instructional design should prevent extraneous load and leave as much working memory resources as possible for maximizing intrinsic load (see Sweller et al., 2011, for an overview of cognitive load effects).

The Expertise Reversal Effect

The expertise reversal effect is one of the instructional principles explained by cognitive load theory. However, rather than suggesting a specific instructional procedure or technique that may optimize working memory resources, the expertise reversal effect is unique in describing the interactions between other cognitive load principles and levels of learner expertise (Kalyuga, 2007; Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga & Renkl, 2010). It provides boundary conditions for cognitive load effects that occur when the effectiveness of a particular technique is influenced by domain-specific prior knowledge. The effect occurs when the relative effectiveness of different learning conditions reverse with changes in levels of learner expertise. For example, the instructional methods that provide best comprehensive guidance to novice learners may fail to keep their advantages or even have reversed effects in comparison with alternative formats when used with advanced learners or when the same learners gradually become more knowledgeable in the domain. Alternatively, advanced learners might perform relatively better in a learning environment that fosters deeper understanding (e.g., by using reflection prompts when performing a task) than in a learning environment without such reinforcements; however, the relative effectiveness of the two conditions reverses with novices.

It should be noted that the term *experts* is frequently used in discussing the expertise reversal phenomenon within the cognitive load framework. In most studies, the term actually refers to narrow task-specific expertise rather than genuine high-level professional expertise that requires extensive years of learning and deliberate practices (Ericsson, Krampe, & Tesch-Römer, 1993). In cognitive load theory, the term typically indicates relative standing in low-high knowledge continuum when referring to learners with more advanced knowledge base in a specific task domain (Kalyuga, 2007).

The expertise reversal effect was predicted and explained as a logical consequence of the features of human cognitive architecture. The available knowledge structures in long-term memory that depend on the level of learner expertise determine the content and, thus, the capacity and other characteristics of cognitive processing in working memory. Accordingly, the level of cognitive load and its classification as intrinsic or extraneous load are relative to the learner's expertise level. Therefore, the effectiveness of the same learning materials will vary as a function of different levels of prior knowledge. For instance, when experts deal with a familiar type of problem, relevant schemas and routine problem solving procedures are activated for governing the solution moves. Such activities would generate intrinsic load, which is essential for solving the problem. In addition, the benefit of strengthening and automating the acquired schemas or making appropriate generalizations for further application of the knowledge also occurs under these conditions.

In contrast, in the absence of relevant knowledge, novices might be forced to use such methods as random search followed by testing of effectiveness in dealing with problem solving tasks (Kirschner et al., 2006; Sweller & Sweller, 2006). The random search for solutions is a cognitively inefficient procedure that imposes wasteful extraneous load on novices. Although learners may occasionally obtain a correct answer by chance, this would unlikely result in meaningful learning of solution schemas. Learning may be more effective when novices are provided with direct and complete instructional support, for example, in the form of worked examples with detailed solution steps (Renkl, 1999, 2005; see Renkl, in this volume, for a review of research in worked examples).

For novices, learning with worked examples may prevent extraneous search-based activities and generate productive intrinsic load essential for constructing missing schemas ("worked examples effect"). However, support may become redundant for experts due to available domain-specific knowledge from their long-term memory. Studying worked examples may require experts to cross-reference external support materials with already learned procedures. Learners could therefore be distracted from fluent cognitive processing by taking full advantage of their existing knowledge structures and could have working memory overloaded by additional, extraneous activities. From this perspective, the expertise reversal effect can be understood as a form of the redundancy effect. It occurs by co-referencing the learners' internal available knowledge structures with redundant, external forms of support. Thus, the effect could be predicted with an increase in learner levels of expertise for the instructional procedures that provide comprehensive guidance or support for novices (Kalyuga, 2007).

Empirical studies on the effectiveness of different instructional formats for presenting multiple sources of information have provided a number of examples of the expertise reversal phenomenon. An integrating information technique may eliminate or reduce extraneous cognitive load associated with searching for or matching spatially or temporally separated sources of information, such as diagrams and related textual explanations or mutually referring pieces of text (split-attention effect). Searching, matching, mentally holding and integrating multiple split-source materials that need to be co-referred for full understanding may generate additional extraneous cognitive load. As a result, insufficient working memory resources are left for making essential inferences and enhancing understanding. However, the effect diminishes for relatively more advanced learners. For experts, these interdependent sources of information may become intelligible on their own and the explanatory components become unnecessary and redundant. An integrated format of learning materials that includes guidance may force experts to waste cognitive resources on processing redundant sources of information and also on cross-checking them with their own understanding. In this situation, a split-attention format might be more suitable for experts because it allows them to ignore the unnecessary information and avoid extraneous load. For example, Kalyuga, Chandler, and Sweller (1998) found that inexperienced electrical trainees learned better in a condition with diagrams of electrical circuits integrated with essential textual explanations than in a format with the diagram and text separated. However, the relative advantage of the integrated-diagram-and-text format diminished as the same group of learners became more knowledgeable after a short intensive training. When the participants reached a level of expertise at which the diagram was self-contained to be understood, they performed better under a diagram-only learning condition.

A similar expertise reversal effect was also found using animated diagrams with auditory explanations in learning how to use specific types of diagrams (nomograms) in mechanical engineering (Kalyuga et al., 1999). Oksa, Kalyuga and Chandler (2010) found an expertise reversal effect caused by co-referring text with another piece of text when comprehending complex literary text. Novices learned better in a condition that provided original Shakespearean play extracts with embedded line by line modern English translation and explanatory interpretations than in a condition that provided the original text only. The relative effectiveness of the two conditions reversed with highly skilled Shakespearean experts. The Shakespearean experts in the integrated explanatory notes group reported higher subjective ratings of cognitive load. Retrospective verbal reports indicated that the experts were unable to avoid cross-checking the accuracy and validity of the explanations with their internalized interpretations and those activities distracted them from directly analysing the original text.

The above examples showed that the positive effect of using integrated formats for novices may reverse for more advanced learners, thus demonstrating an expertise reversal effect. If a single representation of information is sufficient for more knowledgeable learners, the additional information should better be removed. For example, Lee, Plass and Homer (2006) demonstrated that using a single symbolic representational format for gas law simulations in chemistry (e.g., using words *temperature*, *pressure* and *volume* with corresponding numerical values) was beneficial for more knowledgeable learners than using a multiple-representation format with added icons (e.g., images of burners for temperature, weighs for pressure etc.). In contrast, less experienced learners learned better from simulations using representations with words and the added icons than from single representations with words only.

A longitudinal study which investigated the long-term effectiveness of using cognitive and metacognitive learning strategies prompts for assisting university students in writing learning journals in a psychology course provides another example of the advantage of guidance for novices that diminishes as learners become more knowledgeable (Nückles, Hübner, Dümer, & Renkl, 2010). Journal writing is useful in organizing learners' newly acquired knowledge and deepening understanding. Results indicated that novice journal writers benefited from the instructional prompts that were used for stimulating their cognitive strategies (e.g., organization and elaboration skills) and metacognitive strategies (knowing about and the ability to control one's own cognitive processes) at the initial stage of learning in comparison with a control condition without prompts. However, when the learners became relatively more knowledgeable in the domain at the second half of the semester, no differences were found in learning outcomes between the two conditions. Results also demonstrated that more experienced learners were less motivated to write the learning journals when prompted and invested less effort than those in the no-prompt condition. It was suggested that the prompting procedures might have reduced extraneous cognitive load involved in effortful but unproductive search processes and thus benefited novice learners. However, as learners became familiar with or internalized the writing strategies, the external guidance became redundant and interfered with learners' tendency to apply the strategies.

Expertise reversal effect occurs not only when techniques that aim at reducing extraneous load for novices are used with advanced learners, but also when techniques that aim at reducing intrinsic load for novices are used with advanced learners. Several techniques such as *pre-training essential skills* and *isolating interactive elements of information* were suggested to assist novices in learning complex tasks. Blayney, Kalyuga and Sweller (2010) found that breaking down complex spreadsheet formulas into a number of intermediary working entries helped in alleviating excessive working memory load for less experienced undergraduate students in learning accounting. In contrast, as anticipated, reverse patterns of instructional effectiveness were found when using this method with more knowledgeable learners. These students benefited more from a fully interactive task format because intermediary entries could be redundant for them resulting in additional wasteful cognitive load. For novice learners, pre-training in some initial procedures could assist in constructing partial schemas that would effectively reduce the intrinsic load imposed by the complex task. However, it is the learners' available knowledge base that determines the levels of complexity of a task that can be processed in working memory. For advanced learners who have relevant schemas in long-term memory, few elements of information could be encapsulated into a single unit. Therefore, a task that is complex for less knowledgeable learners (due to numerous interacting elements of information that may exceed their working memory capacity) could be easy for advanced learners. Techniques that reduce the complexity of learning materials might result in negative learning consequences for these learners.

The learner levels of prior knowledge also determine if there are any benefits in learning from well-structured materials as a kind of learning support. Although not investigated within a cognitive load framework, McNamara, Kintsch, Songer and Kintsch (1996) found that the effects of textual details and text coherency interacted with learners' levels of expertise in learning from biology texts. More knowledgeable learners obtained higher outcomes when studying with original minimally coherent text while less knowledgeable learners benefited from a format with added details to increase the text coherence. Kalyuga, Law, and Lee (2013) also found an expertise reversal pattern in comprehending narrative texts. The authors suggested that the expertise reversal effect was caused by differential levels of intrinsic cognitive load rather than by extraneous cognitive load. The results of the study demonstrated that less proficient readers benefited from a format with embedded causal words that demonstrated explicit causal relations between the events in the text for enhancing comprehension. Conversely, more proficient readers benefited from the original text without causal words. Measures of cognitive load using subjective ratings suggested that the original text prompted the more proficient readers to make deeper inferences using their available linguistic knowledge structures that increased intrinsic cognitive load. Without the support from the added causal words, the less proficient readers could only engage in surface processing. The embedded causal words assisted these readers in generating deeper inferences that resulted in higher levels of intrinsic load than in the no-causal words conditions.

A number of other studies that investigated hypertext and hypermedia designs also suggested that lower knowledge learners benefited more from well-structured materials than unstructured materials. For instance, Amadieu, Tricot, and Mariné (2009) investigated the effectiveness of two types of concept map structures for presenting hypertext and its interaction with learners' prior knowledge levels in learning multiplication cycle of coronavirus in the domain of virology using graduate students. Results demonstrated that novices performed better on a free recall task when learning with the hierarchical structure format that demonstrated an explicit hierarchy between a set of relations than learning with a network structure format that was built up with relational links (e.g., causal and temporal relations). For more advanced learners, there were no differences between the learning conditions. The hierarchical structure format that provided organisational cues might have assisted novice readers in constructing a coherent reading sequence and limiting disorientation, and thus enhanced learning. More advanced learners might be able to use their existing knowledge structures to construct a coherent mental model that guided their reading paths without any organisational cues. Amadieu, van Gog, Paas, Tricot, and Mariné (2009) conducted a similar investigation in learning the infection process of the retrograde virus HIV. The results were in accord with the expertise reversal expectations and were explained by the traditional cognitive load mediation. Novices who learned under the hierarchical structure condition reported lower ratings of mental effort than the novices who learned under the network structure condition for the performance tests that evaluated their factual and conceptual knowledge acquisition. For more advanced learners, no differences were found for mental effort ratings for both types of performance tasks. (See also Calisir & Gurel, 2003, for another similar expertise reversal pattern in the effectiveness of different hypertext structures in learning productivity using undergraduate students with different levels of prior knowledge.)

Self-explanation is a learning activity that might foster deeper structural processing of information for relatively more knowledgeable learners (see Chiu & Chi, this volume, for a review). Leppink, Broers, Imbos, Vleuten, and Berger (2012) conducted a study investigating the effectiveness of using self-explanations in learning statistics for undergraduate psychology students with different levels of prior knowledge in statistical reasoning. Students in both the worked-examples condition and self-explanation condition were initially instructed to read a basic inferential statistic text. Participants in the self-explanation condition were required to formulate an argument that integrated their judgement of a true/false statement and their answers to a few open-ended questions for each of the learning tasks. Participants under the worked-out examples condition were provided with all the solutions and also the arguments. Results demonstrated that advanced students benefited more from the self-explanations condition than from the worked-out examples conditions on a post-test that evaluated their conceptual understanding of the learning materials. However, the relative effectiveness of the two learning conditions reversed with less experienced learners; they performed better under the worked- examples condition than the self-explanations condition. The authors suggested that novices who had insufficient knowledge base might have experienced high demands of cognitive resources when confronted with complex task and thus the self-explanation activities might have overloaded their working memory. Novices learned better with the provided instructional explanations. In contrast, advanced learners were able to self-explain and thus the external instructional explanations became redundant and rendered extraneous load that impeded learning.

Learning usually does not involve only constructing new schemas but also includes automating at least some of the newly acquired schemas (Cooper & Sweller, 1987). Automated schemas could interact with novel information and be processed in working memory with minimum conscious attention and without consuming cognitive resources (Kotovsky, Hayes, & Simon, 1985; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). It is therefore important for learners to be equipped with relevant automated basic knowledge structures so that working memory is freed up for achieving higher-level learning goals. However, expertise reversal effects were found in studies that investigated the effectiveness of techniques that aim at facilitating knowledge automation. For example, when using an imagination technique, learners are instructed to imagine the acquired procedures (or a set of relations) that may facilitate schemas automation. However, it was demonstrated that the techniques only benefited learners who already acquired the knowledge structures which were essential for imagination. No benefits or even negative outcomes were obtained when the technique was used with novice learners in early stages of learning. Novices learned better by studying worked examples rather than following supplementary imagining instructions (e.g., Cooper et al., 2001; Ginns, Chandler, & Sweller, 2003; Leahy & Sweller, 2005).

Educational Implications of the Expertise Reversal Effect

The expertise reversal effect may raise the awareness of instructors and instructional designers that learning materials should not be provided in a one-for-all fashion. They should not also assume that the designs which provide additional guidance to novices are also suitable for advanced learners. Plenty of available instructional materials have been designed without taking into account learners' cognitive characteristics based on their prior achievements in a domain. Of course, in many realistic situations, the effect may not necessarily reverse entirely, especially if the novice-expert differences are not wide (e.g., Kalyuga et al., 2013). For example, higher-level learners may perform equally in the compared conditions while novices demonstrate significantly better learning outcomes in one of the conditions (e.g., Calisir & Gurel, 2003; Clarke et al., 2005; Mayer & Gallini, 1990). In other words, some techniques that are beneficial for novices may not harm more knowledgeable learners. However, in a number of cases, techniques that enhanced learning for novices had real negative learning impacts on more knowledgeable learners, even following a relatively short training time for increasing the level of expertise (Rey & Fischer, 2013). Still, the expertise reversal effect does not imply a diminution of the actual level of expertise for expert learners following the instruction (as the name of the effect may suggest). Even in the cases of full reversals, the experts still usually perform better than novices in all the compared conditions. What actually reverses is the relative effectiveness (or relative standing) of different instructional conditions at different levels of learner expertise.

The expertise reversal effect also implies that for optimizing learning effects and learning efficiency, it is essential to develop procedures for diagnosing learners' current knowledge levels (see Ambrose & Lovett, this volume) and adjusting appropriate instructional conditions accordingly as learners' expertise gradually changes during learning. For example, Nückles et al. (2010) suggested a *gradual and adaptive fading-out of the prompts method* to overcome the expertise reversal effect when using instructional prompts for enhancing competency in writing learning journals. The prompts were gradually faded out over a few learning stages for each individual student according to her/his performance in the previous stage. Results demonstrated that students in the fading-out group outperformed the permanent-prompts group. This dynamic approach can be realised and practiced in real-time using advanced computer-based learning environments such as intelligent tutors (see Millis, Graesser, & Halpern, this volume). In order to successfully monitor learners' expertise levels in real time, rapid online assessment techniques have also been suggested within the framework of cognitive load theory (Kalyuga, 2006a; Kalyuga & Sweller, 2004). It is based on the idea that when exposed to a problem for a short time, experts are usually able to rapidly retrieve appropriate higher-level solution schemas from long-term memory, while novices may only be able to identify some lower-level representations or commence a random search for solution steps. Accordingly, with the first-step diagnostic method, learners are requested to rapidly indicate their first step in working out the solution for a provided problem within a limited time. Different first-steps will indicate different levels of prior knowledge in a particular task domain (Kalyuga, 2006a, Kalyuga & Sweller, 2004). The *rapid verification method* is an alternative technique that briefly provides learners with a series of possible steps at various stages of the solution procedure, and requests them to verify the correctness of these steps in an immediate response (Kalyuga, 2006b; 2006c). This format is easier to use in online environments, including learning tasks that require drawing graphical representations that may generate too many possible solution routes.

A number of empirical studies demonstrated that learners obtained better outcomes in real-time tailored adaptive learning conditions that implemented the above rapid assessment techniques than in non-adaptive learning conditions (e.g., Kalyuga & Sweller, 2004; Kalyuga, 2006b). In the learner-tailored environments, learners are initially rapidly assessed and allocated to different instructional conditions according to their pretest scores. For learners with lower levels of expertise, direct and detailed instructional guidance is provided to replace their missing knowledge structures that are essential for achieving a particular learning goal. For learners with higher levels of expertise, more problem solving tasks are used to strengthen their ability in applying the acquired knowledge or automate the routine problem solving procedures. Ongoing rapid assessment is conducted at some stages of the learning process. Then, the learners are either allowed to proceed to the next learning stage if the rapid tests indicate that they have mastered the required partial schemas, or they continue to study at the same stage if they fail the rapid tests. As learners' expertise gradually increases during learning, they transition from learning with full guidance to receiving less support. At these stages, fading worked examples with gradually reduced step-by-step guidance could be an optimal instructional procedure (Atkinson, Derry, Renkl, & Wortham, 2000; Kissane, Kalyuga, Chandler, & Sweller, 2008; Renkl & Atkinson, 2003; Renkl, Atkinson, & Groβe, 2004; Renkl, Atkinson, Maier, & Staley, 2002).

Conclusion

The level of learner expertise is a major factor that moderates the relative effectiveness of a particular learning technique or procedure. The critical role of this factor has become clear with recent advances in our knowledge of human cognitive architecture, and it is reflected in the expertise reversal effect. The effect has been consistently replicated in abundant empirical studies using a wide range of learning materials including well-structured tasks in science, mathematics, engineering, programming, and other technical domains as well as ill-structured tasks in language learning. Participants ranged from young school children to adult undergraduates. The research has also been extended to non-academic learning areas like sport training (Khacharem, Spanjers, Zoudji, Kalyuga, & Ripoll, 2013) and professional workplace training of air traffic controllers (Salden, Paas, Broers, & van Merriënboer, 2004; Salden, Paas, & van Merriënboer, 2006). The main instructional implication of the effect is the need to develop adaptive learning environment in which learning materials are tailored dynamically in accord with learners' current levels of achievement. It is therefore essential for instructors to be able to identify the prerequisite schematic knowledge structures required for performing a particular class of tasks and develop tests items that diagnose different levels of learner prior knowledge. In future research, such techniques need to be extended to the domains with poorly defined tasks, for example, in language and social sciences learning.

References

Amadieu, F., Tricot, A., & Mariné, C. (2009). Prior knowledge in learning from a non-linear electronic document: Disorientation and coherence of the reading sequences. *Computers in Human Behavior, 25*, 381-388. doi: 10.1016/j.chb.2008.12.017

Amadieu, F., van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction, 19*, 376-386. doi: 10.1016/j.learninstruc.2009.02.005

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (ed.), *The psychology of learning and motivation: Advances in research and theroy*, Vol. 2 (pp.89-195). New York, NY: Academic Press.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*, 181-214. doi: 10.2307/1170661

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-outs steps. *Journal of Educational Psychology, 95*(4), 774-783. doi: 10.1037/0022-0663.95.4.774

Ayres, P. (1998). The effectiveness of a no-goal approach in solving Pythagorean problems. In C. Kanes, M. Goos & E. Warren (Eds.). *Published Proceedings of the Mathematical Education Research Group of Australasia (MERGA 21)* (pp. 68-73). Gold Coast, Australia.

Baddeley, A. D. (2004). The psychology of memory. In A. D. Baddeley, M. D. Kopelman & B. A. Wilson (eds). *The essential handbook of memory disorders for clinicians* (pp.1-14). Hoboken, NJ: John Wiley & Sons.

Blayney, P., Kalyuga, S., & Sweller, J. (2010). Interactions between the isolated–interactive elements effect and levels of learner expertise: Experimental evidence from an accountancy class. *Instructional Science, 38*, 277-287. doi: 10.1007/s11251-009-9105-x

Calisir, F., & Gurel, Z. (2003). Influence of text structure and prior knowledge of the learner on reading comprehension, browsing and perceived control. *Computers in Human Behavior, 19*, 135-145. doi: 10.1027//1618-3169.49.2.109

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55-81. doi: 10.1016/0010-0285(73)90004-2

Clarke, T., Ayres, P., & Sweller, J. (2005). The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. *Educational Technology Research and Development, 53*, 15-24. doi: 10.1007/bf02504794

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*, 347-362. doi: 10.1037/0022-0663.79.4.347

Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagining. *Journal of Experimental Psychology: Applied, 7*, 68-82. doi: 10.1037/1076-898x.7.1.68

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185. doi: 10.1017/s0140525x01003922

De Groot, A. (1965). *Thought and choice in chess.* The Hague, Netherlands: Mouton. (Original work published 1946).

Diao, Y., & Sweller, J. (2007). Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learning and Instruction, 17*, 78-88. doi: 10.1016/j.learninstruc.2006.11.007

Diao, Y., Chandler, P., & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology, 120*, 237-261.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245. doi: 10.1037/0033-295x.102.2.211

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363-406. doi: 10.1037/0033-295x.100.3.363

Ginns, P., Chandler, P., & Sweller, J. (2003). When imagining information is effective. *Contemporary Educational Psychology, 28*, 229-251. doi: 10.1016/s0361-476x(02)00016-4

Kalyuga, S. (2000). When using sound with a text or picture is not beneficial for learning. *Australian Journal of Educational Technology, 16*, 161-172.

Kalyuga, S. (2006a). Rapid cognitive assessment of learners' knowledge structures. *Learning and Instruction, 16*, 1-11. doi: 10.1016/j.learninstruc.2005.12.002

Kalyuga, S. (2006b). Assessment of learners' organised knowledge structures in adaptive learning environments. *Applied Cognitive Psychology, 20*, 333-342. doi: 10.1002/acp.1249

Kalyuga, S. (2006c). Rapid assessment of learners' proficiency: A cognitive load approach. *Educational Psychology Review, 26*, 735-749. doi: 10.1080/01443410500342674

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509-539. doi: 10.1007/s10648-007-9054-3

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*, 1-19. doi: 10.1007/s10648-010-9150-7

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science, 38*, 209-215. doi: 10.1007/s11251-009-9102-0

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*, 558-568. doi: 10.1037/0022-0663.96.3.558

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31. doi: 10.1207/s15326985ep3801_4

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*, 1-17. doi: 10.1518/001872098779480587

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371. doi: 10.1002/(SICI)1099-0720(199908)13:4<351::AID-ACP589>3.0.CO;2-6

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*, 126-136. doi: 10.1037/0022-0663.92.1.126

Kalyuga, S., Law, Y. K., & Lee, C. H. (2013). Expertise reversal effect in reading Chinese texts with added causal words. *Instructional Science, 41*(3), 481-497. doi: 10.1007/978-1-4419-8126-4

Khacharem, A., Spanjers, I. A. E., Zoudji, B., Kalyuga, S., & Ripoll, H. (2013). Using segmentation to support the learning from animated soccer scenes: An effect of prior knowledge. *Psychology of Sport and Exercise, 14*(2), 154-160. doi: http://dx.doi.org/10.1016/j.psychsport.2012.10.006

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*. doi: 10.1207/s15326985e04102_1

Kissane, M., Kalyuga, S., Chandler, P., & Sweller, J. (2008). The consequences of fading instructional guidance on delayed performance: The case of financial services training. *Educational Psychology, 28*(7), 809.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17*, 248-294. doi: 10.1016/0010-0285(85)90009-x

Leahy, W., & Sweller, J. (2005). Interactions among the imagination, expertise reversal, and element interactivity effects. *Journal of Experimental Psychology: Applied, 11*, 266-276. doi: 10.1037/1076-898x.11.4.266

Lee, H., Plass, J. L., & Homer, B. D. (2006). Optimizing cognitive load for learning from computer-based science simulations. *Journal of Educational Psychology, 98*, 902-913. doi: 10.1037/0022-0663.98.4.902

Leppink, J., Broers, N. J., Imbos, T., van der Vleuten, C. P. M., & Berger, M. P. F. (2012). Self-explanation in the domain of statistics: an expertise reversal effect. *Higher Education, 63*(6), 771-785. doi: 10.2307/41477913

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82,* 715-726. doi:10.1037/0022-0663.82.4.715

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*, 187-198. doi: 10.1037/0022-0663.93.1.187

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43. doi: 10.1207/s1532690xci1401_1

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97. doi: 10.1037/h0043158

Nückles, M., Hübner, S., Dümer, S., & Renkl, A. (2010). Expertise reversal effects in writing-to-learn. *Instructional Science, 38*(3), 237-258. doi: 10.1007/s11251-009-9106-9

Oksa, A., Kalyuga, S., & Chandler, P. (2010). Expertise reversal effect in using explanatory notes for readers of Shakespearean text. *Instructional Science, 38*, 217-236. doi: 10.1007/s11251-009-9109-6

Paas, F., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive load approach. *Journal of Educational Psychology, 86*(1), 122-123.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1-4. doi: 10.1207/s15326985ep3801_1

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193-198. doi: 10.1037/h0049234

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*, 61-86. doi: 10.1016/s0959-4752(01)00016-0

Renkl, A. (1999). Learning mathematics from worked-out examples: Analyzing and fostering self-explanations. *European Journal of Psychology of Education, 14*, 477-488.

Renkl, A. (2005). The worked-out examples principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 229-245). New York, NY: Cambridge University Press

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*, 15-22. doi: 10.1207/s15326985ep3801_3

Renkl, A., Atkinson, R. K., & Groβe, C. S. (2004). How fading worked solution steps works—A cognitive load perspective. *Instructional Science, 32*, 59-82. doi: 10.1023/B:TRUC.0000021815.74806.f6

Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *Journal of Experimental Education, 70*, 293-315. doi: 10.1080/00220970209599510

Rey, G. D., & Fischer, A. (2013). The expertise reversal effect concerning instructional explanations. *Instructional Science, 41*(2), 407-429.

Salden, R. J. C. M., Paas, F., & van Merriënboer, J. J. G. (2006). Personalised adaptive task selection in air traffic control: Effects on training efficiency and transfer. *Learning and Instruction, 16*, 350-362. doi: 10.1016/j.learninstruc.2006.07.007

Salden, R. J. C. M., Paas, F., Broers, N. J., & van Merriënboer, J. J. G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instructional Science, 32*, 153-172. doi: 10.1023/B:TRUC.0000021814.03996.ff

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1-66. doi: 10.1037/0033-295x.84.1.1

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127-190. doi: 10.1037/0033-295x.84.2.127

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295-312. doi: 10.1016/0959-4752(94)90003-5

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.

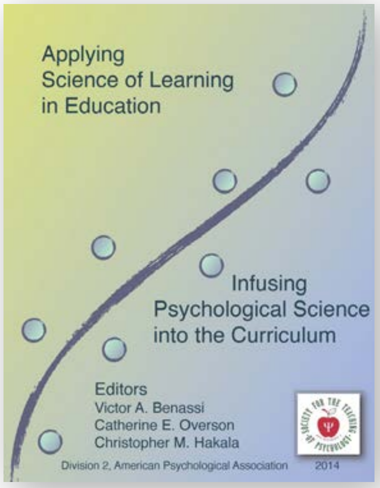
Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*, 123-138. doi: 10.1007/s10648-010-9128-5

Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology, 4*, 434-458.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory (Explorations in the learning sciences, instructional systems and performance technologies)*. New York, NY: Springer.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296. doi: 10.1023/a:1022193728205

Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist, 43*(1), 49-64. doi: 10.1080/00461520701756420



Using Feedback to Promote Learning

John A. C. Hattie
University of Melbourne

Gregory C. R. Yates
University of South Australia

The term 'feedback' emerged around the start of the 20th century. It was coined to describe mechanisms that would allow a well-engineered system to regulate itself (e.g., when a heating system exceeded a certain temperature it turned itself off, and could restart when the temperature got below a certain level). There is an inherent sense of control and automation within its original conceptualization. In reviewing the history of the social sciences, it is apparent that early behaviourists used feedback ideas extensively when attempting to analyse alterations in behavioural rates and probabilities. They conceived feedback as stemming from reinforcers (positive feedback) or in avoidance of punishment (negative feedback). The consequence of present actions would regulate changes in the rate of future action, in accord with the goal of servicing an organism's motivational requirements, be they biological or socially defined. Irrespective of such uses, the term 'feedback' began to be used widely, becoming an aspect of everyday vernacular well aside from its original derivation in engineering theory and practice.

Today, feedback has come to have many more expansive meanings. For instance, there are distinctions between feedback and feed forward, and it can refer not only to actual consequences but can also be self-rewarding (i.e., feedback may inspire further actions and be intrinsically rewarding within itself). Certainly, our responsiveness to feedback is seen as a key driver of behavioural adaptation. This notion of sensitive and adaptive responsiveness was one of the initial ideas investigated through psychological science methods, often researched and documented under the term ‘knowledge of results’.

In broad perspective, it is not hard to locate convincing evidence pointing to feedback as a key process implicated in behavioural change. For instance, military gunners become more accurate and less wasteful in using ammunition when provided with detailed instruction coupled with feedback, both factors being essential (Gagne, 1962), young drivers become more careful after accidents or near accidents (Mayou, Simkin, & Threlfall, 1991), and entrepreneurs often report certain positive learning effects associated with analysing business failures (Politis & Gabrielsson, 2009). Consider, as an example, the famous Hawthorne studies where workers were asked to be part of an experiment; no matter how different independent variables were manipulated, worker productivity increased. For example, when the light in the factory was reduced, rest pauses were introduced, and group payment was changed to individual payment, still the production increased. The claim, now known as the Hawthorne effect, was that because the workers knew they were part of an experiment and being observed, their performance increased. An alternative explanation for the increased output is the feedback workers received regarding their performance. Parsons (1974), reviewed original data from the Hawthorne studies. He noted that the workers were provided with regular (half-hourly plus accumulated total) records of the work; this not only provided information about their performance relative to their targets, it also provided goal setting information that led to increased wages.

Irrespective of diverse theories and terminology, feedback has come to be regarded as a vital part in every individual's development and learning progress. To experience feedback presents opportunity for an individual student to map progress toward his or her providence. Indeed, the commonly used definition of feedback, within behavioral science, is that it constitutes information allowing a learner to reduce the gap between what is evident currently and what could or should be the case (Sadler, 1989). Within educational programs, any feedback, so afforded, informs and guides an individual as to the next step to be undertaken. At least, that is the ideal.

Feedback in the Classroom Context: Is There an Empathy Gap?

Classroom-based research has disclosed three curious facets about feedback. First, teachers allege they dispense much helpful feedback to their students at relatively high levels and they claim they do so routinely. Second, trained classroom observers, in studies into interaction patterns in classrooms, disclosed that feedback occurred in most classrooms at relatively low levels, and barely at all in many locations. Even studies with highly expert teachers indicated strangely low levels of feedback observed during classroom instruction (Ingvarson & Hattie, 2008). Third, when students ware asked what feedback they experienced in classrooms, many reported the level as being low to virtually nil. For instance, Nuthall (2007) tracked young children across extended periods and found they received considerably higher levels of feedback from their peers than their teachers (and most of the peer feedback was incorrect). For many students, feedback from teachers is indexed in terms of only several seconds per day.

In all this, there is the making of the classic empathy gap. The evaluation of a supposedly similar event depends upon who you are, the role or position you occupy, and the nature of the relationship within the interaction. Empathy gaps have been studied a good deal through social psychological research. For instance, we underestimate others' pain when we have never had a similar experience (Nordgren, Banas, & MacDonald, 2011), and when we are in love it is difficult to understand what is like for one not to be. Our biases can lead to us to underestimate the influences of others' experiences. The notion of empathy gaps between teachers and their students within the classroom, however, remains a matter yet to be fully explored. One area empathy gap has been examined is in relation to classroom management. For example, Lewis, Romi, and Roache (2012) found that students were inclined to blame teachers for punishments they received, even when teachers strove to be fair and just.

Teachers' statements concerning feedback is that the their intentions are honest and open. They believe they are doing a sound job in this area and providing much feedback. For example, many excellent teachers realise that individual feedback is time consuming and will spend time instructing their class in how past assignments can be improved. However, in interviewing students, it is noted that group-level feedback is largely irrelevant to those that have mastered an objective, and often is ignored by those that have not. The teacher believes he or she has provided feedback, but many within the class are bored, tuned out, or simply focussing on other things in their life more important at the time. That students 'tune out' when teachers administer generic or group-level feedback is well known (Hattie & Gan, 2011).

A teacher may act as though he or she believes giving praise is a valuable form of feedback. Indeed, a low level of praise is welcomed and valuable in establishing relationships within a positive or benign classroom climate. Praise, however, quickly loses appeal, and is what classroom researchers call a threshold variable. You need to provide some praise, but not too much. If you praise a good deal, students learn you are a teacher that praises a good deal, and that is all. This is not a commodity you can increase and expect the effects of praise will increase. Psychologically, praise within the classroom can become problematic in that it fails to convey any genuine feedback information. Even worse, it can shift the students' attention onto irrelevant, even destructive, factors, such as excessive attention to the self or one's ability, thus discouraging further effort or listening to feedback about the task(Dweck, 1999; Skipper & Douglas, 2012).

Recognising the Power of Feedback

Several meta-analyses have established that the average effects attributable to feedback are among the highest we know in education, but also that feedback effects are among the most variable in their influences (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008). From a review of 12 meta-analyses that have included specific information on feedback in classrooms (based on 196 studies and 6972 effect-sizes), the average effect-size was *d* = .79 (see Bertsch & Pesta, this volume, for a discussion of effect sizes and meta-analysis), which is twice the average effect (Hattie, 2009). This gives credence the common aphorisms that feedback can double the rate of learning, and that feedback is among the top 10 influences on achievement. It certainly is a common denominator of many of the top influences, and largely absent from those averaging below .40.

As highlighted in Kluger and DeNisi's (1996) review, the variance of feedback effects, however, is considerable, indicating that some types and circumstances of feedback are more powerful than others. It appears that merely prescribing high dosages of feedback does not imply that learning will take place, as one has to take into consideration differential effects on learners. It seems we know much about the power of feedback, but too little about how to harness this power and make it work more effectively in the classroom.

One of the hidden effects of feedback appears to be in influencing how much effort students allocate to an activity. In one controlled study, college students were asked to work on several tasks within a class period (Northcraft, Schmidt, & Ashford, 2011). It was found students devoted considerably more time and effort to those tasks upon which they discovered that specific and timely feedback was available. The availability and quality of feedback available to students can readily signal the importance placed on the learning activity by their teachers.

Conceptions of Feedback and the Three Key Questions

To appreciate feedback as a psychological process, it is helpful to draw attention to two rich models by way of analogy. These are (a) the computer video game, and (b) the motorist's GPS device. Computer games provide a player with immediate feedback on tasks, which provide individual challenges within a stimulating context. The machine monitors or 'knows' the player's past history. This knowledge is used to ensure that the current problem is presented slightly beyond past accomplishment – that is using the Goldilocks principle – not too hard and not too easy. Feedback is then administered which, through applied effort, enables the prescribed challenge to be met. Such a process may continue for hours at a time.

In the case of your automobile's GPS, travelling through an unknown city becomes a breeze. The device yields feedback information about where you are, and provides active directions and immediate feedback to guide you to a specified destination. The device assumes you know where you want to go, although you lack specific knowledge concerning how to get there. The GPS will not growl at you for deviations, but politely awaits your corrective action. If you fail to apprehend the supplied information, it is forgiving of any error you make, maintaining a patient unemotional disposition until you finally garner sufficient resources as to locate the desired pathway.

The computer game analogy is valuable in that it highlights the level of individual tailoring that has to take place and the critical importance of knowing where you currently are, having clear understanding of the criteria of success – and then much feedback is provided aimed to reduce this gap. The GPS analogy appears valuable in that it maps the key questions that underpin the successful applications of feedback in student-teacher interactions: The three key inevitable questions are always: (a) where is the student going? (b) Just how is the student getting on right now? (c) Just what is the next step?

Where is the student going? The role of goals

The first question fixes upon the role of clearly articulated goals. Feedback does not work in a vacuum. Instead, it registers discrepancy between the current state and a known objective. The student can be expected to be sensitive to experienced feedback only once there is a clearly known objective in place. This has to be seen as the necessary starting point.

There exists a substantial literature base concerning the value of goals and goal setting as base-level ingredients in motivation, effortful striving, and in achieving self-control (Locke & Latham, 2002; 2013). It has been shown that setting challenging goals is more motivating than either possessing easy goals or urging people to 'do your best'. Available studies with adults suggest, however, this proposition is valid only when individuals are provided with feedback as to their progress towards the challenging goal. In the language of goal-setting theory, the presence of feedback moderates the impact of goal difficulty on personal striving. People work hard on difficult goals once they can perceive the availability of salient and supportive feedback. On the other hand, such feedback is unimportant if a goal is facile. When goals are easy, people do not depend on externally administered feedback, and providing such feedback adds no additional motivational value.

The critical goal underpinning effortful striving is not to appear 'busy', but to achieve a known and attainable standard or performance. Students need to be able to recognise successful products such that these products represent worthwhile goals attainable through personal effort. Although long-term goals are important, motivating goals require a short-term focus. When dealing with students, it is thus important for them to map a series of proximal sub goals as this helps considerably in maintaining effortful momentum. Teachers need to invest time ensuring that students are aware of immediate lesson goals and performance standards, and how these relate to the broader distal goals.

Hence, effective goals are expressed, not in terms of personal enjoyment or mere participation, but in terms of indexed outcomes, generally aligned within a sequence of ordered steps. Without such an understanding of the desired outcomes and the necessary sub goals, feedback could be disorienting, confusing, or interpreted as something about the student rather about than the current task that both parties (student and teacher) must focus upon.

Teachers profit immensely from being able to present motivating goals to students in terms of social modelling exemplars or worked examples. Students use such exposure to comprehend what success looks like. These models can provide not only information as to what is the desired end state, but will also convey knowledge as to steps and processes undertaken on route. Indeed, the feedback process assumes that such sources of motivation and striving are firmly set into place as a base plan. Such motivation is shared by both teacher and student, and is often expressed, either explicitly or implicitly, in terms of a desirable modelling stimulus or completed product. To effect, there needs to be a definable entity to look at, value, dissect, and emulate. For example, providing worked examples of the goal is a major contributor to students putting in effort and being receptive to feedback (see Renkl, this volume for a discussion on worked-examples).

What progress has been made? Or how am I going?

The second pertinent question – *what is the level of progress*? - hinges upon at least one of the two agents (teacher or student) being in a position to assess the student's strengths and weaknesses leading to the success criteria. For goals to be effective, students need information about current achievements. There is anticipated progress, and intermediate sub goals must be attained. Progress is generally along a prescribed track, which is not necessarily unique to the individual. There may be a necessary sequence or skills hierarchy to be mastered. An overall goal will imply a sequence of sub goals which may impose excessive load and make progress problematic. Feedback needs to take the form of realistic assessments as to how far along the journey the student has come, and can serve to reduce cognitive load by showing the students where in this sequence they need to exert their thinking and effort. Studies with high school students indicate that feedback becomes more effective when it reflects progress made in terms of actual accomplishments, rather than normative standards such as letter grades (Harks, Rakoczy, Hattie, Besser, & Klieme, 2013). Such process-oriented feedback is rated by students as more useful than grade information, and is linked to enhanced interest in the material being taught.

What direction to take now? What is the next step?

The third question *Where to next*? builds directly on answers elicited by the previous question. Students are disinterested in past errors and post-mortems, but clamour for guidance as to what to do in the future, as defined in terms of the next few minutes. The teacher's role is now to enable resources, help, instruction, and scaffolds to be in place to assist the student to know *where to next?* A clear direction has to be identified. Through such guidance, progress toward the initial goal can be maintained with enthusiasm and effort. Students who receive feedback with little guidance as to how to approach the task to make revisions may feel frustrated, often leading to further negative affect (Lipnevich & Smith, 2009).

In connection with this third point, the GPS analogy becomes poignant. The device is programmed to only look forward. It remains unconcerned about any past streets, suburbs, or erstwhile errors passed through on route. Further, the same routes must apply to all individuals sharing the same goal. We may get to a location through using different roads and we make take longer or shorter to get there. But such differences are superficial since options are constrained severely by reality. At some point, all routes to one goal will converge, even though starting points vary. Despite the obvious gulf in technologies, we can see the GPS as providing a heuristic analogy as to how students can use teacher-sourced feedback in the classroom context. The analogy also suggests aspects of the process that teachers can find instructive, such as the need to move forward through non-emotional gestures. It also suggests that we may need to attend more closely to developing curricula progressions that model how students learn (whereas most curricula are based on how teachers would like students to progress).

Feedback Adapted to Student Needs: Responding to Levels of Mastery

It is well established in the literature that generic feedback in the form of teacher praise fails to encourage actual classroom learning in the manner suggested by earlier theories of self-esteem and of behavioural reinforcement (Brophy, 1981). Within the world of psychology, such theories were popular from the 1960s through to the present day. All-to-often, it was assumed that within the classroom then, (a) students needed high levels of affirmative feedback to bolster their self-esteem which then would have overall positive effects, and that (b) learners needed contingent positive reinforcement in the form of approval and praise, which would cement learned associations into place and enable mastery to be achieved. Unfortunately, both views were based on incomplete notions about human motivation, which in turn gave rise to insufficient conceptualizations about both (a) the feedback process itself, and (b) the role classroom teachers play in stimulating students' urge to learn and acquire knowledge.

Modern cognitive theories focus on learners' need to build their knowledge bases. Cognitive theory assumes that learning is goal-oriented, where comprehension and understanding are inevitable goals. Each student's fundamental problem is to understand his or her world, and to feel able to control, or at least make more predictable, key elements that impact directly on adjustment within that world. One of the basal problems facing each student sitting in a classroom is assessing what to learn, how much to learn, and how much – and where to expend effort. Biologically and psychologically, we are resource-limited creatures. Although educators thrive on messages such as 'learning for its own sake', or 'intrinsic motivation', the reality is that students arrive at school with the need to be cautious and economical in cognitive effort expenditure (Willingham, 2009).

As stressed earlier, the feedback process moderates the impact of perceived goal difficulty. Difficult goals motivate striving once a learner can see that efforts will be acknowledged. If the goal is difficult, long term, or remote, then students need instruction, scaffolding, and feedback to operate in a seamless process as they continue to invest effort and attention to meet the goal. We suggest that that learners' needs surface in three distinct guises as related to level of knowledge and expertise. These three levels are: (a) initial knowledge acquisition, with a focus on building knowledge in new areas, (b) applying knowledge and skills acquired earlier, and (c) exhibiting strong levels of mastery or expertise. Below is a brief review of what is known about the effective feedback process in relation to these levels.

Initial knowledge acquisition

Beginners need feedback based on content knowledge while striving to build basic knowledge and vocabulary. They need assurance and corrective feedback, often in the form of discriminations such as correct versus incorrect, or right versus wrong. They need to know they are heading in the right direction, and that they are using the right vocabulary within the right context. They will appreciate also a level of encouragement in confidence building. They require teachers' recognition that they are genuinely getting somewhere along the path to greater understanding. Their mental efforts will involve a good deal of memory load, which can be a huge source of stress. Of major importance to beginners is (a) the feeling that they have been able to recall a relatively complete set of knowledge elements, and (b) that they will not be caught out displaying glaring omissions relative to what is known by other people.

Applying knowledge

Intermediate learners have acquired basic concepts but need help linking ideas together, seeing relationships, and in extending and elaborating upon the basic ideas. They need assurance that they are applying the right methods and strategies in the right place. Since they are starting to move beyond the information given, they need to know if the inferences they have been making are seen as valid ones by a more knowledgeable or senior agent such as a teacher. They appreciate positively phrased suggestions for alternative strategies (For instance, "*strong use of adjectives in just the right spots* , "*good use of the acceleration principle, but other principles could play a role here", or "a beautifully well-constructed argument, but have you thought of what this implies for the future?"*).

At this stage, there is less stress on straight memory recall, but a marked shift into describing relationships between interacting elements, and knowing valid procedures to apply. These students lap up worked examples, which demonstrate how basic materials, as previously mastered, will provide powerful means of understanding broader areas of functioning. The student needs to appreciate that knowledge elements do not reside as isolated islands of limited or useless facts. Instead, knowledge is vital and is implicated in much broader pictures. Hence, the role of feedback is now one of assuring the student that, when expressed in conceptual, schematic, and procedural forms, then securely acquired knowledge is being applied well beyond the time and place of the original learning.

Advanced mastery level

At more advanced levels, helpful feedback takes the form of supporting the self-regulated learner such that sincere efforts to extend and apply knowledge even further are actively recognised. The recognition here is that the outcome or product achieved is of such quality as to approach the highest level of achievement possible. Hence, the teacher takes time to acknowledge both the effort expended and the product achieved. In essence, different types of feedback work best depending on the individual's phase of learning -- corrective feedback is suited for novices, process feedback is needed as the learner develops proficiency, and elaborated conceptual feedback becomes appropriate with highly competent students. One of the outcomes from this level is more automaticity in what has been learned, which is powerful when starting on the next series of knowledge building, and so on. This is a virtuous cycle as it makes the next series somewhat more fluid, whereas those who do not advance much past the first level are bound to be more cognitive resource intensive and this can revert into a less virtuous cycle leading to giving up, reducing effort, and seeking easier goals.

The Learner's Need to Self-Correct: How Does Delay in Feedback Play a Facilitating Role?

Beginning learners profit from feedback close in time to their responses – particularly during the initial knowledge acquisition phase. Indeed, the traditional finding in this area has been that with any simple learning, or with fine-grain discriminations, or in motor sports areas, then effective feedback takes place immediately, sometimes within seconds. Immediate feedback assists the initial knowledge or skill acquisition process when dealing with novices; with increasing skill level, immediately applied feedback is less critical. Once the products of effort become more visible, then the key element is for recognition of the quality and value of those outcomes. Such an aspect has been referred to as the IKEA effect (Hattie & Yates, 2014). Students value what they produce and will expect you to recognise the inherent value of their work, and possibly devalue if you fail to do this (Norton, Mochon, & Ariely, 2012).

Once learners become knowledgeable and relatively autonomous, then immediate feedback becomes less critical. Immediately experienced feedback can even discourage the emerging self-correction process (Mathan & Koedinger, 2005). As mastery is acquired, students develop strategies for monitoring, review, and revision. Since it takes time for reflection, and to access and apply one's knowledge base, it can be unhelpful to receive gratuitous feedback too early in the piece.

In essence, poorly timed external feedback may subtract from the opportunity to engage in self-corrective personal feedback. This aspect becomes more pertinent when feedback is targeted upon lower level activities that fail to relate well to more advanced product-oriented goals harboured within the recipient's mind. Furthermore, several studies indicate that when people are exposed to continuous monitoring, even when applied by the self, then on-going performances in complex situations can be disrupted (Shapira, Gundar-Goshen, Liberman, & Dar, 2013). A modicum of monitoring is helpful, but too much becomes destructive.

It is important for the person administering feedback to recognise the unit of work being produced by an individual, and the mental processes and demands so involved. Recognition of the individual's cognitive load and the value being placed on the resultant outcomes, become key elements underpinning feedback effects when dealing with clever and knowledgeable individuals. One of the purposes of feedback is to provide information or strategies to reduce cognitive load such that the student expends thinking power onto specific aspects of the task that they are working on. On the other hand, aspects such as timing become less important at this level.

Evidence from laboratory studies indicates that, when learning clearly does take place, immediate feedback may be unnecessary, but feedback, when delayed, can still contribute to learning through a beneficial spacing effect. That is, a delay of several hours, or even a day, between the initial acquisition of complex information, and receiving feedback (for instance on a test) may serve to provide valuable practice in retrieval and use of the initial learning (Metcalfe, Kornell, & Finn, 2009; Roediger & Butler, 2011).

It defies common sense to think that learning is enhanced when feedback is deliberately withheld. But that reality becomes possible provided learning is secure in the first place, and when immediately applied feedback serves little purpose. Such a notion cannot be applied easily in classrooms. Teachers cannot control all the factors that account for learning in the way a laboratory study might operate. It would be churlish to withhold feedback on this rationale. In general, teachers have to operate on the heuristic that feedback is best applied soon after a task is completed to the learner's satisfaction. On the other hand, it is comforting to know that when a teacher does give delayed feedback on tests, it may assist students through active retrieval practice.

The Issue of Positive and Negative

The impact of negative experience considerably outweighs positive experience. This notion is captured by the phrase 'bad is stronger than good', which sums up a considerable body of research (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). On the other hand, many classroom 'process-product' studies conducted in the 1970s repeatedly found high levels of teacher criticism to have a negative impact on many classroom indices, including student achievement test scores (Brophy, 1981, 1986). Punitive teachers contribute to negative classroom climate, which is linked with a range of undesirable outcomes, often in a manner suggesting snowball or cascade effects (Demanet & Van Houtte, 2012).

Such findings, however, cannot be interpreted to mean that feedback needs to be set at some universally positive level. Indeed, the findings cited in the previous paragraph most often refer to instances where teacher criticism was associated with student conduct matters. When it comes to receiving feedback for goal directed efforts, however, its purpose is to indicate deficiencies, to correct for errors, or to suggest better ways of proceeding. This type of feedback is ‘negative’ only in the sense that its purpose is to redirect effort in a different direction (i.e., akin to the customary meaning of the term negative feedback as used in engineering systems).

Corrective feedback can be carried out within the context of positive teacher-student relationships, especially when the teacher displays a positive and respectful attitude toward student efforts, and communicates that such errors are natural steps along the path to mastery. Such positive relationships (not only between teacher and student, but among students) is often a precursor to building the trust needed to not merely tolerate, but welcome feedback that is specific and understood. The notion that excellent teachers express expectations in accord with incremental learning approaches, rather than treat fumbling efforts as indexing fixed ability, has been well documented under the rubric of teacher expectancy effects (Cooper, Hinkel, & Good, 1980; Rattan, Good, & Dweck, 2012; Rubie-Davies, 2007).

When teachers dispense feedback in the sense of disconfirmation and error correction, the critical facet will lie in the student's perception of the teacher's purpose, and not in terms of any constant need for positive affirmation. As indicated earlier, the impact of feedback is bound up with recipients' understandings of the nature of the objectives and respective goals being pursued by all parties. In a study with 11-year old students, Autin and Cloizet (2012) showed that the effects of failure on a difficult problem solving task were offset by conveying to students the idea that experiencing difficulty is a perfectly normal and expected part of the learning process (see also Clark & Bjork, this volume for a discussion on desirable difficulties and errorful learning). Students who heard this explicit message before failing performed better on a subsequent series of memory and comprehension tasks, than those who had not heard the message, and also better than a third group that had experienced only success on easy problem solving tasks beforehand. Being able to “wallow in the pit” of problem solving is learned skill, often needing high levels of feedback (but not necessarily immediately).

A child's tendency to attribute lack of immediate success to personal capability rather than to natural task difficulty is one central aspect of the learned helplessness syndrome (Au, Watkins, & Hattie, 2010; Diener & Dweck, 1978). However, the average child possesses sufficient resilience to avoid functioning in self-defeating ways, and steadfastly avoids such premature attributions. Nevertheless, it is worth noting that helplessness symptoms have been shown to be elicited in students when teachers are consistently paired with students’ failure experiences (Dweck & Reppucci, 1973), or when teachers convey the notion that failure on tests is an acceptable outcome, even one worthy of sympathy (Graham, 1984).

Once students are aware that they have some distance to travel, then 'errors' are no longer perceived as negative. Errors are tentative trials rather than the endgame. They can represent significant efforts, achieved with some cost, which have some resemblance to a product. But such efforts do not represent or account for the final version. What may be rated as 'negative' by a casual observer may be 'positive' to a learner. This becomes pertinent when a student is aware that substantive goal progress is being made and is being acknowledged by a teacher appreciative of both (a) the approximations attained, and (b) the genuine level of psychological load and effort experienced by the student at the moment of production.

Appreciating When Feedback Is, and Is Not, a Critical Component of Classroom Learning

Feedback is not an isolated or unitary process. Instead it is embedded within a sequence of instructional moves that enable a teacher to convey to a student that specific outcomes are attainable. Teachers naturally employ motivational strategies, convey aspirations, create desires, and outline new possible selves. But remote goals are vacuous in the absence of explicit scaffolding, direct instruction, modelling stimuli, and appropriate feedback. Certainly, students view those teachers who provide feedback as more committed to their progress (Lipnevich & Smith, 2009).

Feedback becomes a key determinant of classroom-based learning when tasks are difficult, the student lacks knowledge, and unassisted efforts toward a remote goal will occasion likely disappointment. Many aspects of classroom life and behaviour may not require any explicit feedback on the part of a teacher. Once an individual has acquired a skill, its maintenance requires little external support. However, students arrive at school deficient in their knowledge of areas such as science, geography, social history, and mathematics. Developing mastery in such curriculum areas will require the teacher to manage learning sequences with sensitivity and precision.

For many students, school learning is hard, mentally demanding, and potentially energy sapping. At the individual level, the problem is to perceive a knowledge chasm as a knowledge gap to be bridged. Such a view is aided by two other necessary perceptions: (a) that one's achievement is not limited by inherently fixed attributes such as intelligence, and (b) outcome is correlated with effort. If such incremental beliefs are to supplant an entity view of personal capacity, then supportive instructional components need to be in place. This is the essential classroom context within which teachers' use of the feedback principle will determine how their students' will elect to deploy their limited energies.

The Place of Error

Feedback thrives on errors, but for many educators errors have been something they tend to minimise, and students too often see errors as indicators of their failure to material. Indeed, Skinner (1953), the father of behaviourism, equated errors with punishment that inhibits behaviour. Since this time, many researchers have tried to classify errors. Reason (1990) saw errors as part of three cognitive stages. The first stage relates to the formation of an intention and planning of the subsequent actions, and an error relates to an incorrect intention—such that while the action proceeds the intention may be inappropriate. The second stage involves the storage and retention of task-salient information, such that an error is a ‘lapse’ usually related to retaining information appropriate for the task. The third stage consists of executing the specified action, and an error, or ‘slip’ refers to the incorrect implementation of a correctly formulated intention and plan. Hence, feedback can be oriented to each of these stages, perhaps by assisting the student to be aware of the intention or success criteria, by ensuring that the progression through the actions to the intention are well scripted (e.g., by providing a rubric), that there are many opportunities to pick up ‘lapses’ and relearn or learn anew task salient information, or to provide students the answers so they can concentrate on the process of getting to the answers.

By considering an environment where errors are welcomed, feedback can be more effective. From this error-perspective, feedback can be considered as ‘disruptive’ in that its role is to move the student back onto the path of moving to the intention or success criteria. Of course, it helps if the student is adept at error-detection, but often this is the primary skill needed in learning (and for feedback to be sought and received).

With adult participants, there has been recent research on error management training. Such training is based on three premises: there is only minimal guidance and students are encouraged to actively explore and experiment with the concepts being learning; students are told to expect errors and that they should be considered positive to learning; and when encountering errors, students are encouraged to think ahead and try something new. Keith and Frese (2008) completed a meta-analysis of 24 studies that compared error management training with an alternative training. The overall effect was .44 which led them to conclude that EMT leads to better training outcomes compared with training methods that do not encourage errors during training. The effects were higher when the major training goal is to transfer learned skills to novel problems (adaptive transfer, *d* =.80) than where the goal is to learning and apply just one particular procedures (analogical transfer, *d* =.20).

In Perspective: When Feedback is Effective

As stressed earlier, feedback is not an isolated classroom procedure. Its impact depends on adequacy of the instructions the teacher has conveyed, the nature of the goals and sub goals set, and the students' conception of the assessment to be applied. To complete this chapter we list nine key points that help to describe when feedback becomes an effective instructional component.

1. The feedback process resides in that what is received and interpreted by a student, rather than what a teacher gives or believes has taken place.
2. Feedback can work best when criteria for success are known to the learner in advance, and where the goal to achieve such success is shared by students and teacher alike.
3. Feedback can best cue attention onto a task, with a known goal or sub goal, and away from self-focus.
4. Feedback must engage a learner at, or just above, the current level of functioning.
5. Feedback must challenge the student to invest effort in moving forwards, and assure the learner that it is perfectly natural to experience difficulties when mastering difficult tasks.
6. Feedback is powerful when the classroom climate is one of welcoming errors and seeing disconfirmation as a natural and positive part of developing and exercising new skills.
7. It is optimal when feedback matches the content of the instruction given earlier, and is in accord with available social modelling stimuli identified as worthy of emulation.
8. Feedback thrives in an environment where errors are welcomed, and error training may be a worthwhile adjunct to increase the power of feedback.
9. Feedback is received and used by students when teachers themselves are prepared to adapt and adjust their methods and priorities in response to the outcomes indexed through their students' success on various assessment outcomes.

References

Au, R. C. P., Watkins, D. A., & Hattie, J. (2010). Academic risk factors and deficits of learned hopelessness: a longitudinal study of Hong Kong secondary school students. *Educational Psychology, 30*(2), 125-138. doi: 10.1080/01443410903476400

Autin, F., & Croizet, J.-C. (2012). Improving working memory efficiency by reframing metacognitive interpretation of task difficulty. *Journal of Experimental Psychology: General, 141*(4), 610-618. doi: 10.1037/a0027478

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323-370. doi: 10.1037/1089-2680.5.4.323

Brophy, J. (1981). Teacher praise: A functional analysis. *Review of Educational Research, 51*(1), 5-32. doi: 10.3102/00346543051001005

Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist, 41*(10), 1069-1077. doi: 10.1037/0003-066x.41.10.1069

Cooper, H. M., Hinkel, G. M., & Good, T. L. (1980). Teachers' belief about interaction control and their observed behavioral correlates. *Journal of Educational Psychology, 72*(3), 345-354. doi: 10.1037/0022-0663.72.3.345

Demanet, J., & Van Houtte, M. (2012). Teachers' attitudes and students' opposition: School misconduct as a reaction to teachers' diminished effort and affect. *Teaching and Teacher Education, 28*(6), 860-869. doi: 10.1016/j.tate.2012.03.008

Diener, C. I., & Dweck, C. S. (1978). An analysis of learned helplessness: Continuous changes in performance, strategy, and achievement cognitions following failure. *Journal of Personality and Social Psychology, 36*(5), 451-462. doi: 10.1037/0022-3514.36.5.451

Dweck, C. S. (1999). Caution: Praise can be dangerous. *American Educator, 23*(1), 1-5.

Dweck, C. S., & Reppucci, N. D. (1973). Learned helplessness and reinforcement responsibility in children. *Journal of Personality and Social Psychology, 25*(1), 109-116. doi: 10.1037/h0034248

Gagne, R. M. (1962). Military training and principles of learning. *American Psychologist, 17*(2), 83-91. doi: 10.1037/h0048613

Graham, S. (1984). Teacher feelings and student thoughts: An attributional approach to affect in the classroom. *Elementary School Journal, 85*(1), 91-104. doi: 10.2307/1001621

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2013). The effects of feedback on achievement, interest and self-evaluation: the role of feedback’s perceived usefulness. *Educational Psychology, 33*, 1-22. doi: 10.1080/01443410.2013.785384

Hattie, J.A.C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* London: Routledge.

Hattie, J.A.C., & Gan., M. (2011). Instruction based on feedback. In R. Mayer & P. Alexander (Eds.), *Handbook of Research on Learning and Instruction*. (pp. 249-271). New York: Routledge.

Hattie, J.A.C., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. doi: 10.3102/003465430298487

Hattie, J.A.C., & Yates, G. C. R. (2014). *Visible learning and the science of how we learn*. London: Routledge.

Ingvarson, L., & Hattie, J. A. C. (Eds.). (2008). *Assessing teachers for professional certification: The first decade of the National Board for Professional Teaching Standards*. Amsterdam: Elsevier Press.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology, 93*(1), 59-69. doi: 10.1037/0021-9010.93.1.59

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284. doi: 10.1037/0033-2909.119.2.254

Lewis, R., Romi, S., & Roache, J. (2012). Excluding students from classroom: Teacher techniques that promote student responsibility. *Teaching and Teacher Education, 28*(6), 870-878. doi: 10.1016/j.tate.2012.03.009

Lipnevich, A. A., & Smith, J. K. (2009). 'I really want to learn': Students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability, 21*(4), 347-367. doi: dx.doi.org/10.1007/s11092-009-9082-2

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705-717. doi: 10.1037/0003-066x.57.9.705

Locke, E. A., & Latham, G. P. (Eds.). (2013). *New developments in goal setting and task performance*. New York: Routledge.

Mathan, S. A., & Koedinger, K. R. (2005). Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist, 40*(4), 257-265. doi: 10.1207/s15326985ep4004_7

Mayou, R., Simkin, S., & Threlfall, J. (1991). The effects of road traffic accidents on driving behaviour. *Injury, 22*(5), 365-368. doi: http://dx.doi.org/10.1016/0020-1383(91)90095-V

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition 37*(8), 1077-1087. doi: 10.2307/ 1170349

Nordgren, L. F., Banas, K., & MacDonald, G. (2011). Empathy gaps for social pain: Why people underestimate the pain of social suffering. *Journal of Personality and Social Psychology, 100*(1), 120-128. doi: 10.1037/a0020938

Northcraft, G. B., Schmidt, A. M., & Ashford, S. J. (2011). Feedback and the rationing of time and effort among competing tasks. *Journal of Applied Psychology, 96*(5), 1076-1086. doi: 10.1037/a0023221

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology, 22*, 453-460.

Nuthall, G. A. (2007). *The hidden lives of learners*. Wellington: New Zealand Council for Educational Research.

Parsons, H. M. (1974). What happened at Hawthorne?: New evidence suggests the Hawthorne effect resulted from operant reinforcement contingencies. *Science, 183*(4128), 922-932. doi: 10.1126/science.183.4128.922

Politis, D., & Gabrielsson, J. (2009). Entrepreneurs' attitudes towards failure: An experiential learning approach. *International Journal of Entrepreneurial Behaviour and Research, 15*(4), 364-383. doi: 10.1108/13552550910967921

Rattan, A., Good, C., & Dweck, C. S. (2012). “It's ok — Not everyone can be good at math”: Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology, 48*(3), 731-737. doi: http://dx.doi.org/10.1016/j.jesp.2011.12.012

Reason, J. (1990). *Human error*. Cambridge: Cambridge University Press.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20-27. doi: http://dx.doi.org/10.1016/j.tics.2010.09.003

Rubie-Davies, C. M. (2007). Classroom interactions: Exploring the practices of high- and low-expectation teachers. *British Journal of Educational Psychology, 77*(2), 289-306. doi: 10.1348/000709906x101601

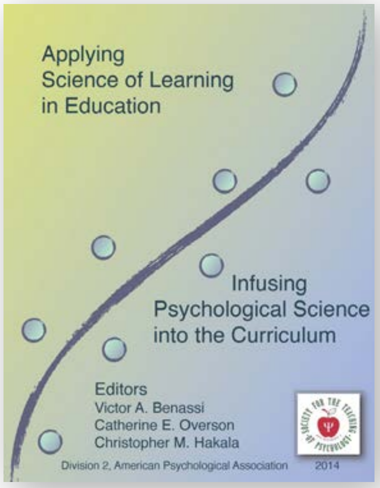
Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144. doi: 10.1007/bf00117714

Shapira, O., Gundar-Goshen, A., Liberman, N., & Dar, R. (2013). An ironic effect of monitoring closeness. *Cognition & Emotion, in press*, 1-9. doi: 10.1080/02699931.2013.794771

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189. doi: 10.2307/40071124

Skipper, Y., & Douglas, K. (2012). Is no praise good praise? Effects of positive feedback on children's and university students' responses to subsequent failures. *British Journal of Educational Psychology, 82*(2), 327-339. doi: 10.1111/j.2044-8279.2011.02028.x

Willingham, D. T. (2009). *Why don't students like school? A cognitive scientist answers questions about how the mind works and what it means for your classroom*. San Francisco: Jossey-Bass.



Research-Based Principles for Designing Multimedia Instruction

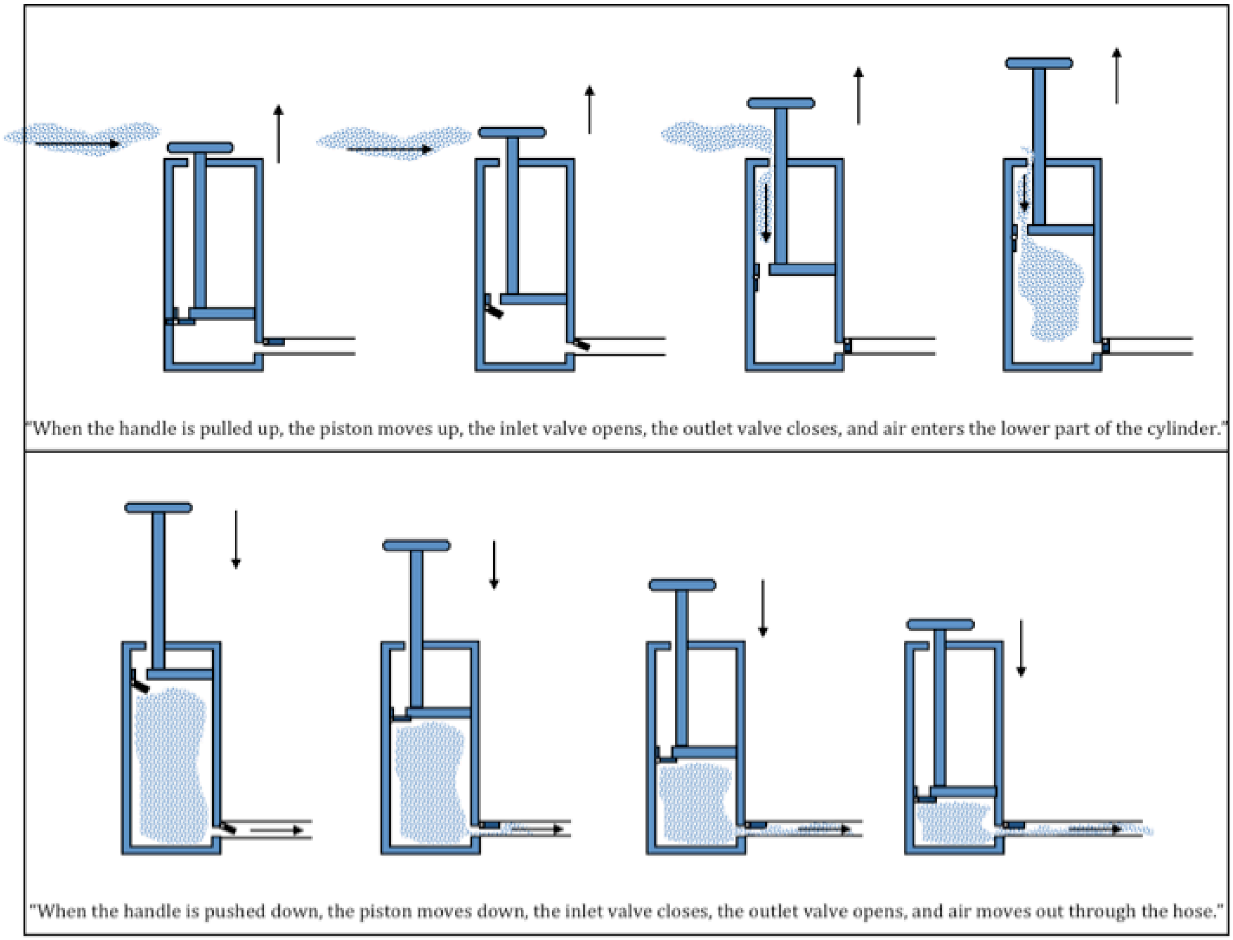
Richard E. Mayer
University of California, Santa Barbara

Overview of Multimedia Instruction

People learn more deeply from words and graphics than from words alone. This assertion can be called the *multimedia principle*, and it forms the basis for using *multimedia instruction*—that is, instruction containing words (such as spoken text or printed text) and graphics (such as illustrations, charts, photos, animation, or video) that is intended to foster learning (Mayer, 2009).

For example, Figure 1 shows frames from a narrated animation on how a tire pump works. In this case the words are spoken and the graphics are presented as an animation. Other examples include textbook lessons presented on paper, slideshow presentations presented face-to-face, captioned video presented via computer, or educational games and simulations presented on hand-held devices. Regardless of presentation medium, what makes all these examples of multimedia instruction is that they use words and graphics to promote learning.

Figure 1: Frames from a Narrated Animation on How a Tire Pump Works



What is the evidence for the multimedia principle? In a series of 11 experimental comparisons my colleagues and I have found that students perform much better on a transfer test when they learn from words and graphics than from words alone (e.g., narration and animation versus narration alone, or text and illustrations versus text alone), yielding a median effect size of $d = 1.39$ (Mayer, 2009). For example, students are better able to answer troubleshooting questions about tire pumps after viewing the narrated animation shown in Figure 1 than after solely hearing the narration (Mayer & Anderson, 1991). In short, research shows that multimedia instruction has the potential to greatly improve student understanding of how things work, including tire pumps, brakes, generators, and lightning (Butcher, in press; Mayer, 2009).

Although instruction has traditionally emphasized verbal modes of presentation (such as lectures and printed books), recent advances in graphics technology now allow more widespread incorporation of visual modes of presentation including illustrations, charts, photos, animation, and video in presentations and in interactive venues such as games and simulations. However, not all graphics are equally effective so careful research is needed to pinpoint principles of multimedia instructional design. The goal of this chapter is to provide a brief overview of 12 research-based principles for how to design effective instruction that uses words and graphics.

An important starting point is to examine principles that are based on an understanding of how people learn from words and graphics. Figure 2 summarizes the cognitive theory of multimedia learning (Mayer, 2009, in press-a), which is based on three core assumptions based on the science of learning (Mayer, 2011): *dual channel assumption*—people have separate channels for processing visual and verbal material (Paivio, 1986); *limited capacity assumption*—people can process only a limited amount of material in a channel at any one time (Baddeley, 1999); and *active processing assumption*—meaningful learning occurs when learners select relevant material, organize it into a coherent structure, and integrate it with relevant prior knowledge (Wittrock, 1989; Mayer, 2009).

Figure 2 represents memory stores as rectangles: *sensory memory,* which temporarily holds incoming images and sounds; *working memory*, which allows for mentally manipulating a small amount of the incoming visual and verbal material; and *long-term memory,* which is the learner's permanent storehouse of knowledge. Figure 2 represents cognitive processing as arrows: *selecting*, which transfers some of the incoming images and sounds to working memory for additional processing; *organizing,* which organizes the images into a pictorial model and the words into a verbal model in working memory; and *integrating,* which connects the models with each other and with relevant knowledge activated from long-term memory. A multimedia message enters the cognitive system through the learner's ears and eyes. The top row represents the verbal channel (into which spoken words and sounds enter) and the bottom row represents the visual channel (into which graphics and printed words enter), although in working memory printed words can be converted into sounds and images can be converted into spoken words.

Figure 2: Cognitive Theory of Multimedia Learning

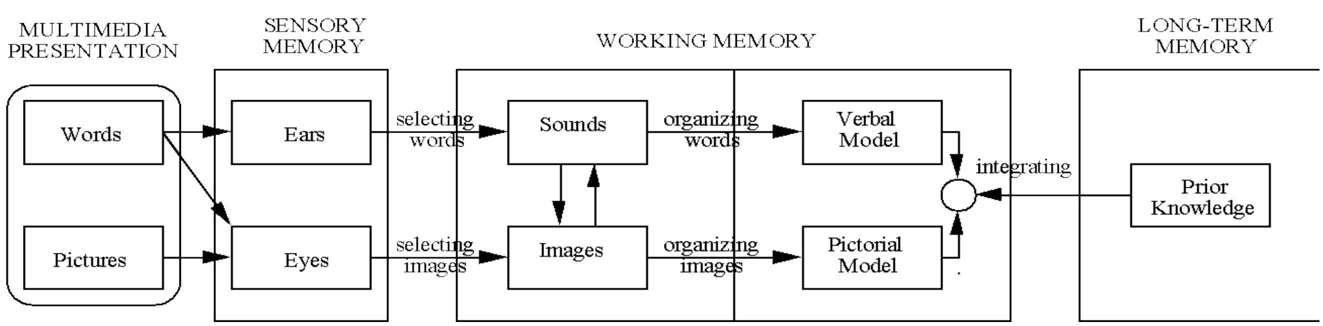


Table 1 summarizes three kinds of processing that can occur during multimedia instruction—*extraneous processing,* which drains limited cognitive processing capacity without contributing to learning; *essential processing,* which involves selecting relevant information and organizing it as presented in working memory; and *generative processing*, which involves making sense of the material by reorganizing it into a coherent structure and integrating it with relevant prior knowledge. This analysis is similar to that proposed in cognitive load theory (Sweller, Ayres, & Kalyuga, 2011) and suggests the need for three kinds of instructional design goals—*reducing extraneous processing*, when extraneous processing and required essential processing exceed the learner's cognitive capacity; *managing essential processing*, when the required essential processing exceeds the learner's cognitive capacity; and *fostering generative processing*, when the learner has processing capacity available but chooses not to exert the effort to use it for making sense of the material. These three types of goals form the basis for three kinds of instructional design principles for multimedia learning, which are presented in the next section.

Table 1: Three Kinds of Cognitive Processing During Learning

Cognitive processing	Description	Instructional goal
Extraneous	Not related to instructional goal, caused by poor instructional design	Reduce extraneous processing
Essential	Aimed at representing essential material, caused by complexity of material	Manage essential processing
Generative	Aimed at making sense of essential material, caused by learner's effort	Foster generative processing

Research on Design Principles for Multimedia Instruction

In this section I examine 12 research-based principles for how to design multimedia, including five principles aimed at reducing extraneous processing, three principles aimed at managing essential processing, and four principles aimed at fostering generative processing. For each principle, we conducted a meta-analysis (Ellis, 2010) using published articles that included transfer tests and we computed the median effect size based on Cohen's d (Cohen, 1988; see Bertsch & Pesta, in this volume, for a definition of the meaning of d and a brief description of meta-analysis). Following Hattie (2009), we consider any effect size greater than *d* = 0.40 to be educationally important.

Reduce Extraneous Processing

Table 2 summarizes five principles aimed at reducing extraneous processing: the coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. The table specifies the number of experimental tests in which positive results were obtained and provides the median effect size based on a meta-analysis by Mayer and Fiorella (in press). A learner experiences *extraneous overload* when essential cognitive processing required to understand the essential material in a multimedia message and extraneous cognitive processing required to process extraneous material exceeds the learner's cognitive capacity. These five principles are intended to address the problem of extraneous overload.

Table 2: Five Research-Based Principles Based on Reducing Extraneous Processing

Principle	Description	Tests	Effect size
Coherence	Delete extraneous material	22 of 23	0.86
Signaling	Highlight essential material	25 of 29	0.41
Redundancy	Don't add onscreen captions to narrated graphics	16 of 16	0.86
Spatial contiguity	Place printed words near corresponding part of graphic	22 of 22	1.10
Temporal contiguity	Present spoken words at same time as corresponding graphics	9 of 9	1.22

The Coherence Principle

People learn more deeply from a multimedia message when extraneous material is excluded rather than included. The rationale for the coherence principle is that people are better able to focus on the essential material if we eliminate extraneous material that could distract them. This principle was supported in 22 out of 23 experimental tests, yielding a median effect size of 0.86. For example, students who learned from a multimedia lesson on how a virus causes a cold performed better on a transfer test if the lesson did not contain seductive details—sentences that gave interesting but irrelevant facts about viruses (Mayer, Griffith, Jurkowitz, & Rothman, 2008). Similarly, students who learned from a narrated animation on lightning formation performed better on a transfer test if the lesson did not also contain short video clips depicting lightning strikes (Mayer, Heiser, & Lonn, 2001). Concerning boundary conditions, reviews of the coherence principle suggest the effects may be strongest for learners with low rather than high working memory capacity, when the lesson is system-paced rather than learner paced, and when the extraneous material is highly interesting rather than neutral (Mayer & Fiorella, in press; Rey, 2012).

The Signaling Principle

People learn more deeply from a multimedia message when cues are added that highlight the organization of the essential material. The rationale for the signaling principle is that people will learn more efficiently if the lesson is designed to call their attention to the important material in the lesson and how it is organized. This principle was supported in 25 out of 29 experimental tests, yielding a median effect size of 0.41. Signaling of the verbal material includes using an outline, headings, highlighting (such as underlining) and pointer words (such as first, second, third). Signaling of visual material includes arrows, flashing, and spotlighting. For example, in a narrated animation on how an airplane achieves lift, students performed better on a transfer test if the narration included an initial outline, headings, and voice emphasis on key words (Mautone & Mayer, 2001). Concerning boundary conditions, Mayer and Fiorella (in press) report that the signaling principle may apply most strongly when the learner might otherwise be overwhelmed with extraneous processing—such as, for low-knowledge learners rather than high-knowledge learners, for complex material rather than simple material, and when it used sparingly rather than excessively.

The Redundancy Principle

People learn more deeply from graphics and narration than from graphics, narration, and on-screen text. The rationale is that with redundant presentations people may waste precious processing capacity by trying to reconcile the two verbal streams of information or may focus on the printed words rather than the relevant portions of the graphics. This principle was supported in 16 out of 16 experimental tests, yielding a median effect size of 0.86. For example, Moreno and Mayer (2002) reported that students performed better on a transfer test about lightning if they received a narrated animation about lightning formation rather than the same narrated animation with concurrent onscreen text inserted at the bottom of the screen. Concerning boundary conditions, the redundancy effect can be diminished or even reversed when the learners are experienced, the on-screen text is short, or the material lacks graphics (Mayer, 2009; Mayer & Fiorella, in press).

The Spatial Contiguity Principle

People learn more deeply from a multimedia message when corresponding printed words and graphics are presented near rather than far from each other on the page or screen. The rationale is that spatial contiguity helps learners build connections between corresponding words and graphics. This principle was supported in 22 out of 22 experimental tests, yielding a median effect size of 1.10. For example, Moreno and Mayer (1999) found that students performed better on a transfer test after viewing an animation about lightning in which printed words were placed next to the part of the lightning system they described than when printed words were placed at the bottom of the screen as a caption. Similar results are reported in a meta-analysis by Ginns (2006). Mayer and Fiorella (in press) report there is preliminary evidence that the spatial contiguity principle may be strongest for low prior knowledge learners, non-redundant text and pictures, complex lessons, and interactive formats.

The Temporal Contiguity Principle

People learn more deeply from a multimedia message when corresponding graphics and narration are presented simultaneously rather than successively. The rationale is that temporal contiguity helps learners build connections between corresponding words and graphics. This principle was supported in 9 out of 9 experimental tests, yielding a median effect size of 1.22. For example, students performed better on a transfer test when they received a narrated animation on how a tire pump works than when they heard the narration before or after the animation (Mayer and Anderson, 1991). Researchers have noted that the temporal contiguity principle may apply most strongly for high rather than low spatial ability learners, when lessons are long rather than short, and when the lesson is system paced rather than learner paced (Mayer, 2009; Mayer & Fiorella, in press).

Manage Essential Processing

Table 3 summarizes three principles aimed at managing essential processing: the segmenting, pre-training, and modality principles. The table specifies the number of experimental tests in which positive results were obtained and provides the median effect size based on a meta-analysis by Mayer and Pilegard (in press). These principles are intended to address the instructional problem of *essential overload,* which can occur when a fast-paced multimedia lesson contains material that is complicated for the learner. A learner experiences *essential overload* when the amount of essential cognitive required to understand the multimedia instructional message exceeds the learner's cognitive capacity.

Table 3: Three Research-Based Principles Based on Managing Essential Processing

Principle	Description	Tests	Effect size
Segmenting	Break lesson into learner-paced parts	10 of 10	0.79
Pre-training	Present characteristics of key concepts before lesson	13 of 16	0.75
Modality	Use spoken words rather than printed words	52 of 61	0.76

The Segmenting Principle

People learn more deeply when a multimedia message is presented in learner-paced segments rather than as a continuous unit. The rationale is that segmenting allows people to fully process one step in the process before having to move onto the next one. This principle was supported in 10 out of 10 experimental tests, yielding a median effect size of 0.79. For example, Mayer and Chandler (2001) found that students performed better on a transfer test if a narrated animation about lightning was broken into 16 10-second segments in which students could press a “continue” button to go on to the next segment. A review revealed that some potential boundary conditions are that segmenting may have stronger effects for learners with low rather than high working memory capacity and for low achieving rather than high achieving learners (Mayer & Pilegard, in press).

The Pre-training Principle

People learn more deeply from a multimedia message when they have learned the names and characteristics of the main concepts. The rationale is that pre-training allows students to focus on the causal connections in the multimedia explanation because they already know the names and characteristics of the key elements. This principle was supported in 13 out of 16 experimental tests, yielding a median effect size of 0.75. For example, students performed better on a transfer test based on a narrated animation on how brakes work if before the lesson they were introduced to the names and characteristics of key components mentioned in the lesson such as the piston and brake shoe (Mayer, Mathias, & Wetzell, 2002). However, an important boundary condition is that the pre-training principle may not apply to high prior knowledge learners, perhaps because they are less likely to experience essential overload (Mayer, 2009; Mayer & Pilegard, in press).

The Modality Principle

People learn more deeply from a multimedia message when the words are spoken rather than printed. The rationale is that the modality principle allows learners to off-load some of the processing in the visual channel (i.e., the printed captions) onto the verbal channel, thereby freeing more capacity in the visual channel for processing the animation. This principle was supported in 53 out of 61 experimental tests, yielding a median effect size of 0.76. For example, Moreno and Mayer (1999) found that students performed better on a transfer test after receiving a narrated animation on lightning formation than after receiving the same animation with on-screen captions that contained the same words as the narration. Similar results are reported in a meta-analysis by Ginns (2005). As the most studied principle in the list, research shows that the modality principle should not be taken to mean that spoken words are better than printed words in all situations. Some important boundary conditions reported in some studies are that printed word may be effective when the verbal material contains technical terms, is in the learner’s second language, or is presented in segments that are too large to be held in the learner’s working memory (Mayer, 2009; Mayer & Pilegard, in press).

Foster Generative Processing

Table 4 summarizes four principles aimed at fostering generative processing: the personalization, voice, embodiment, and image principles. The table specifies the number of experimental tests in which positive results were obtained and provides the median effect size based on a meta-analysis by Mayer (in press-b). These principles are intended to use social cues to prime the learner’s motivation to exert effort to make sense of the material. Social cues in a multimedia message such as conversational style, voice, and gesture may prime a sense of social presence in learners that leads to deeper cognitive processing during learning and hence better test performance.

Table 4: Four Research-Based Principles Based on Fostering Generative Processing

Principle	Description	Tests	Effect size
Personalization	Put words in conversational style rather than formal style	14 of 17	0.79
Voice	Put words in human voice rather than machine voice	4 of 5	0.69
Embodiment	Have onscreen agent use human-like gestures and movements	11 of 11	0.36
Image	Do not necessarily put static image of agent on the screen	9 of 14	0.20

The Personalization Principle

People learn more deeply when the words in a multimedia presentation are in conversational style rather than formal style. The rationale for this technique is that conversational style can prime a sense of social presence in the learner, which causes the learner to try harder to make sense of what the instructor is saying by engaging in appropriate cognitive processing during learning, leading to learning outcomes that are better able to support problem-solving transfer. This principle was supported in 14 out of 17 experimental tests, yielding a median effect size of $d = 0.79$. For example, Mayer, Fennell, Farmer and Campbell (2004) found that students performed better on a transfer test after receiving a narrated animation on how the human respiratory system works when conversational wording was used (e.g., “your lungs,” or “your nose”) rather than formal style (e.g., “the lungs” or “the nose”). Similar results are reported in a meta-analysis by Ginns, Martin, and Marsh (in press). Some important boundary conditions are that the personalization principle may not apply for high-achieving students or long lessons.

The Voice Principle

People learn more deeply when the words in a multimedia message are spoken in a human voice rather than in a machine voice. Human voice is intended to prime a sense of social presence in learners. For example, Mayer, Sobko, and Mautone (2003) found that students performed better on a transfer test after receiving a narrated animation lightning that used a human voice rather than a machine synthesized voice. This principle was supported in 4 out of 5 experimental comparisons, with a median effect size of $d = 0.69$. Research shows that a possible boundary condition is that the voice principle may not apply when there are negative social cues such as low embodiment.

The Embodiment Principle

People learn more deeply when onscreen agents display human-like gesturing, movement, eye contact, and facial expression. Human-like action is intended to create a sense of social presence with the instructor. In 11 out of 11 experimental comparisons, people performed better on transfer tests when they learned from a high-embodied agent than from a low-embodied agent, yielding a median effect size of $d = 0.36$. For example, Mayer and DaPra (2012) found that students performed better on a transfer test after viewing an online slideshow that was narrated by an onscreen agent who used human-like gesture, facial expression, eye gaze, and movement than an onscreen agent who did not move, gesture, gaze, or show expression. A possible boundary condition is that the embodiment principle may not apply when there are negative social cues such as machine voice.

The Image Principle

People do not necessarily learn more deeply from a multimedia presentation when the speaker's image is on the screen rather than not on the screen. Having a static image may cause distraction that detracts from any social benefits. For example, Mayer, Dow, and Mayer (2003) found that adding the image of an onscreen character did not improve learning much (*d* = 0.19) from a narrated animation on how electric motors work. This principle is based on 14 experimental tests in which half produced negative or negligible effects, yielding a median effect size of *d* = 0.20.

Practical Application of Design Principles for Multimedia Instruction

The principles summarized in this chapter are based on research and grounded in cognitive theory, but more work is needed to better delineate the boundary conditions under which the principles apply. In particular, most of the supporting research involves short-term laboratory studies, so it is useful to determine the degree to which the principles apply in more authentic learning situations such as in schools or work training. For example, a promising step in this direction involves a recent finding by Issa et al. (2013) showing that redesigning a medical school classroom slideshow lecture on shock based on multimedia principles resulted in improvements on an immediate transfer test (*d* = 0.76) and a delayed transfer test (*d* = 1.17).

The principles presented in this chapter are intended to apply to a range of instructional scenarios ranging from textbooks to face-to-face slide show presentations to computer-based lessons to interactive games and simulations. Within a classroom, these principles apply to the design of classroom printed materials, computer-based exercises and simulations, and face-to-pace instruction including slideshow presentations.

For example, suppose you wished to apply research-based multimedia design principles to improve a short slideshow on how a solar cell works for presentation in an environmental science class. First, in deference to the coherence principle, you might decide to prune interesting but irrelevant material that you had downloaded from the Internet, including photos of a solar cell installation in a desert in southern California and a short video clip you found in which Al Gore envisions the coming environmental disaster. In short, you work to weed out images and words that are not essential to explaining how a solar cell works, including eliminating entire slides or parts of slides.

In deference to the signaling principle, you place a heading at the top of the remaining 10 slides, which succinctly describe the step in the process being depicted. In addition, you use arrows to show the movement of electrons within the solar cell, corresponding to the description in the text.

In line with the segmenting principle, you break the description in each slide into a few very short descriptions of actions rather than one long paragraph.

In line with the spatial contiguity principle, you place the text describing the action next to the corresponding part of the graphic (such as putting “Electrons move across the barrier” next to an arrow from one side of the barrier to another) rather than at the bottom of the graphic.

Based on the pre-training principle you begin with a slide that that depicts the key elements (such as the “positive side” and “negative side” of the solar cell) along with a verbal label next to each key element, perhaps connected with a line.

Corresponding to the redundancy principle, you do not simply read the text on the slides, but rather elaborate on it, using complete sentences.

In this case, the strictest interpretation modality principle and the redundancy principle can be modified by including a minimal number of words on the screen—mainly, to help highlight the main points and to concretize technical or unfamiliar terms.

In line with the personalization principle, you use first and second person style (such as saying, "Let's see how a solar cell works." rather than "This lesson tells how a solar cell works.").

Consistent with the embodiment principle, you stand next to the slide as you talk, pointing out what you are talking about and maintaining a smile and eye contact with the audience.

Based on temporal contiguity, you are careful to choreograph your speech so that it corresponds to what you are pointing to in the graphic.

In line with the voice principle, you practice to make sure you use a pleasant, flowing voice that exudes confidence and warmth.

In line with the image principle, you remove a distracting logo from each slide which shows Al Gore's face looking down along with the slogan: "LIVE GREEN."

All in all, this example demonstrates how to accomplish your goal of removing material that primes extraneous processing, helping learners understand a complex system by using techniques such as segmenting and pre-training, and fostering deeper processing by creating a feeling of social conversation. I will consider this chapter to be a success to the extent that instructors and instructional designers are able to improve student learning by adapting the principles presented in this chapter and other evidence-based principles (see Mayer, in press-c).

Acknowledgement

Preparation of this chapter was supported by a grant from the Office of Naval Research.

References

Baddeley, A. D. (1999). *Human memory*. Boston: Allyn & Bacon.

Butcher, K. (in press). The multimedia principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning.* New York: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Mahwah, NJ: Erlbaum.

Ellis, P. D. (2010). *The essential guide to effect sizes.* New York: Cambridge University Press.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313-332.

Ginns, P. (2006). Integrating information: A meta-analysis of spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16,* 511-525.

Ginns, P., Martin, A.J., & Marsh, H.M. (in press). Designing instructional text in a conversational style: A meta-analysis. *Educational Psychology Review.*

Hattie, J. (2009). *Visible learning.* New York: Routledge.

Issa, N., Mayer, R. E., Schuller, M., Wang, E., Shapiro, M. B., & DaRosa, D. A. (2013). Teaching for understanding in medical classrooms using multimedia design principles. *Medical Education, 47,* 388-396.

Mautone, P. D. & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology, 93*, 377-389.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed). New York: Cambridge University Press.

Mayer, R. E. (2011). *Applying the science of learning*. Upper Saddle River, NJ: Pearson.

Mayer, R. E. (in press-a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning.* New York: Cambridge University Press.

Mayer, R. E. (in press-b). Principles based on social cues: Personalization, voice, image, and embodiment principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning.* New York: Cambridge University Press.

Mayer, R. E. (Ed.). (in press-c). *The Cambridge handbook of multimedia learning.* New York: Cambridge University Press.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83*, 484-490.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*, 390-397.

Mayer, R. E., & DaPra, C. S. (2012). An embodiment effect in computer-based learning with animated pedagogical agent. *Journal of Experimental Psychology: Applied, 18,* 239-252.

Mayer, R. E., Dow, G., & Mayer, R. E. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology, 95,* 806-813.

Mayer, R. E., Fennell, S., Farmer, L., & Campbell, J. (2004). A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology, 96*, 389-395.

Mayer, R. E. , & Fiorella, L. (in press). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. New York: Cambridge University Press.

Mayer, R. E., Griffith, E., Jurkowitz, I. T. N., & Rothman, D. (2008). Increased interestingness of extraneous details in multimedia science presentation leads to decreased learning. *Journal of Experimental Psychology: Applied, 14,* 329-339.

Mayer, R. E., Heiser, H., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93,* 187-198.

Mayer, R. E., Mathias, A., & Wetzell, K. (2002). Fostering understanding of multimedia messages through pre-training: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied, 8*, 147-154.

Mayer, R. E. , & Pilegard, C. (in press). Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. New York: Cambridge University Press.

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology, 95*, 419-425.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358-368.

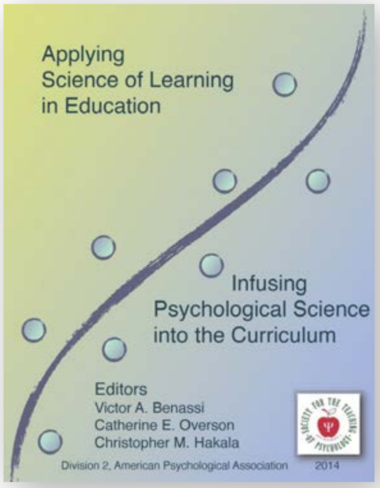
Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology, 94,* 598-610.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

Rey, G. D. (2012). A review and a meta-analysis of the seductive detail effect. *Educational Research Review, 7,* 216-237.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* New York: Springer

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist, 24*, 345-376



Generating Active Learning

Sharon Bertsch
University of Pittsburgh at Johnstown

Bryan J. Pesta
Cleveland State University

The idea that learning is more effective when one spends effort on the material instead of using more passive types of review is well recognized in the field of education (e.g., Thomas & Rhower, 1986). What is often less clear, however, is exactly what is means by “effortful” or “active” versus “passive” learning, and what types of active learning, if any, are more effective educational tools than others. In the field of psychology, one way these concepts are described and tested is as an experimental finding called the generation effect or generation advantage. In its simplest form, the generation effect is the finding that, in most cases, information is more likely to be remembered when it was studied under conditions that required learners to produce (or generate) some or all of the material themselves, versus reading what others had prepared. In this chapter, we will summarize the research on the generation effect and how strong an impact it has on memory performance, describe what conditions make it more or less effective (experimentally), and offer some suggestions for how these experimental designs might be adapted to college classroom experiences.

In experimental research on the memory advantage of generating material participants use a particular researcher-provided rule to produce information that is later tested in one of several ways. These rules vary widely and include things like creating list of words that are the opposite of (or that rhyme with) provided words, completing the last word in a sentence, or even the completion of addition or multiplication problems. These generated words or numbers are then tested using different types of measures including recognition (e.g., multiple-choice) or some type of recall (e.g., essay) format. The percent correct on the test for these participants is compared to the percent correct of a similar group who read prepared lists of words (or numbers) but produced none of the information themselves. The difference between these percentages is the size of the generation effect (Slamecka & Graf, 1978).

As happens in any scientific field, individual experiments conducted on the same topic often find different results. This inconsistency makes trying to apply those results to real-life circumstances difficult, as one experimenter’s results sometimes appear to conflict with others’. This has happened in work on the generation effect as well, as researchers use different materials, study slightly different aspects of the same phenomenon, and use different experimental designs and different groups of participants (who are diverse themselves in many ways). Underlying all this variability is the true size of the benefit of generating material over reading it. The apparent change in how effective generating is over reading is largely because each of these individual studies pulls out one aspect of the effect for analysis, distorting the complete picture. One way to get more of a "bottom-line" feeling about any particular effect is to use a statistical method called a meta-analysis. This technique converts all the results of individual experiments into a common measure (called an effect size) that summarizes the seemingly disparate experiments. Effect size measures give a more practically useful estimate of how much generation of information boosts memory performance over reading only, and it can also be used compare groups of results that are different based on a category of interest (e.g., the memory benefits of solving addition problems versus multiplication problems).

In 2007, we conducted a meta-analysis with our colleagues Richard Wiscott and Michael A. McDaniel. In it, we combined the results of 86 studies on the generation effect. In addition to summarizing these studies to determine (on average) how beneficial the generation of material is across the board, we also carved these studies up into categories. This allowed us to investigate what kinds of effects were larger or smaller—what types of study, testing or individual person characteristics made producing your own answers more beneficial for memory performance over reading them as compared to others. Before we get to the results of our study, we pause here for some brief statistical commentary. First, remember that people's scores on all the different types of memory tests fall on a continuum, representing very poor to excellent performance—regardless of the kind of studying they do. Although the best estimate of a score in the group is the average one, individuals will be scattered over the whole scale. This means that while the average score for generators might be higher than for average readers, some generators will score lower than some readers, and some readers will score better than some generators. So while we can make recommendations to improve memory performance that are likely to work for many (or even most) students, we cannot promise they will work for everyone (see Lee & Kalyuga, this volume, for context on this point). The second comment has to do with interpreting the size of the generation effect itself, or the differences among generation benefits depending on how they are produced. For example, we found in our meta-analysis that producing an entire word improved memory performance more than generating only part of that word (such as adding the final letter). But how useful or practical is this difference? How many generators in these two conditions did better than those who only read?

Although guidelines exist that allow determinations of whether these average differences are small, medium or large, those descriptions may not be meaningful for application in a practical context. For the following discussion, we will provide the statistical effect sizes that we found in our meta-analysis (these are represented as something called Cohen's d, which is translated as how many standard deviation units one average score is from another), but we will also describe what that effect size means in terms of how much the memory scores of the individual people taking the tests in the different conditions overlap. For example, across all the studies we evaluated, the average size of the generation effect was $d = .40$. In psychology, this is referred to as a medium-sized effect. Not very helpful, we know. However, a little more useful is the translation of this effect size to the percent of overlap in the distribution of scores—about 66% of people who generated information had better test scores than the average score of those who only read the information. Thus, in general, the larger the effect size is, the less the reading and generating distributions overlap and the more generators who score better than the average reader. To summarize some of the most potentially useful information we found in our meta-analysis:

Variable	Effect size (*d*)	% of generators scoring better than average reader score
Recognition test	.46	68
Free recall test	.32	63
Older Adults tested (65+)	.50	69
Younger adults tested	.41	66
Rhyming rule used to generate words	.46	68
Sentence completion used to generate words	.60	73
Synonym rule used to generate words	.41	66
Letter rearrangement used to generate words	.37	64
25 or fewer words generated or read	.60	73
More than 1 day retention interval between generating/reading and testing	.64	74

Based on these results, we can expect certain types of test preparation to be (on average) more effective than others. For example, the effect is stronger for multiple-choice types of tests as compared to essay testing, and is very effective among older adult learners as well. Of the different kinds of generation rules used, completing sentences appears to lead to the biggest advantage over reading only, but even relatively minor types of generation (like switching two underlined letters) improves test scores over reading alone. Finally, the effect is strong for small blocks of information, and to leads to relatively long-term retention of the material.

These results apply to participants in the kinds of experiments that are typically conducted in psychology. These experiments usually follow the same format, where participants are given a list of words to either read over or to create themselves using a particular rule. Does this same generation effect principle work in more real life situations where students are attempting to learn lecture or textbook material? We believe the answer to this is yes, although investigations of the usefulness of generating information in more realistic situations is less common than laboratory-based research (DeWinstanley, 1995; Peynircioglu & Mungan, 1993). Considering potential applications of the generation effect to improve memory performance, two have been found to be practically useful (although we will mention a possible third later on). The first of these has to do with how students take lecture notes in the first place (or how they take notes from their textbooks)The second has to do with how students study material from lecture notes they have taken, or from a text source they have read In reviewing the experiments done in these situations, there are indications that having students produce at least part of the information for which they will later be held responsible leads to better test scores both when used as a study technique (King, 1992) and also when used as a note-taking strategy (Russell, Caris, Harris, & Hendricson, 1983). This research has also indicated some potential limitations and ways that the technique should be supplemented to make it more effective.

On their own, students chose a wide variety of study techniques, but outlining information, highlighting notes or text material, summarizing notes/text, and writing practice questions are among the most common. Research on whether these techniques produce more learning (i.e., higher test scores) than more passive forms of studying (such as reading over notes, text, or study guides produced by an instructor) tends to agree that students who produce information have higher average test scores than those who only re-read information they are given (e.g., Paris, Wixson & Palincsar, 1986; Van Blerkom, Van Blerkom, & Bertsch, 2006). One of the most commonly investigated study techniques is writing practice questions. In these experiments, students are asked either to write questions in advance of reading information with which they are familiar in order to help them organize what they read (e.g., perhaps asking "What are some of the causes of the Civil War?" before they start reading a passage on it for an American History class), or in cases of unfamiliar material, to create multiple-choice or essay type questions based on lecture or text material after they have read it). These questions can then be used by the student later in preparation for the test. In experimental examinations of the effectiveness of these study techniques, students' later test scores are compared to those students who either did not see practice questions at all, or those who read over practice questions and answers created by the class instructor. For example, Van Blerkom et al. (2006) compared test scores covering a passage on locus of control among four groups of students: the first read the text passage and copied it, the second read and highlighted, the third read and took notes, and the fourth read and wrote potential test questions and answers in the margins as they read. Two results were clear: students who created practice questions on their own had the highest scores on test questions based on that information and students are generally not good at targeting what is considered to be important information from notes or texts. In the Van Blerkom et al study, an average of only 40% of the ideas about locus of control that were on the test were also targeted by the students writing practice questions. In another example, Foos, Mora and Tkacz (1994, Experiment 2) found that students who created their own practice questions got 86% of similar content test questions correct, compared to students who were provided with practice questions (who got 72% of similar test questions correct). The problem was that students creating their own questions only targeted about 20% of the important concepts they were asked to learn from the passage of text.

This difficulty with discerning more important from less important information appears to be a general problem, particularly among poorer academically performing students. This issue is not only with what students choose to study but also in the notes they take from lectures or text sources in the first place, as students' lecture notes typically contain a low percentage of important ideas. King (1992) found that only about 15% of important ideas were recorded in students' notes, which unfortunately is a not an unusually low number. Clearly, if students are not taking adequate notes, their test scores will suffer. Kobayashi (2005) combined the results of 57 studies comparing the learning outcomes of students who did or did not take notes and found an average effect size of $d = .22$ (this translates to only slightly more than half of note takers with better outcomes over the average non-note taker). Basically, this means that the average students' note taking skills are so poor that it is not surprising they do not do well on tests. Although the generation effect appears to be viable in classroom settings, it has empirically highlighted an important problem about which most class instructors were already aware: many students fail to tell the difference between more and less important information.

Although many different interventions exist to help students learn effective study techniques, if they are not working with adequate information in the first place, no amount of study intervention is likely to produce a large effect. In today's technology-assisted learning environments, some instructors are attempting to solve the problem of poor note taking by providing students with either hard copy or electronic versions of their lecture notes. This practice is problematic in two ways. First, as previously indicated, even when students have a complete record of the information they are supposed to learn (e.g., a textbook chapter), many of them are woefully inadequate at parsing the core ideas or most important points from more shallow details when they study. Second, providing students with a complete set of lecture notes enables a more passive type of initial learning of the material. Having a complete set of lecture notes does appear to result in higher test scores for students when compared to those of un-assisted student note-takers: as the un-assisted students are likely only recording a low percent of relevant information, this result is not surprising. The work on the generation effect, however, would predict higher scores (over longer retention intervals) for students who only receive general outlines of lecture or text information, but have to complete the full information themselves. This pattern has been found by several different groups of researchers (e.g., Barnett, 2003). Katayama and Crooks (2003) gave their participants a text passage to read along with either a complete or partial set of notes about the content of the passage. The partial condition provided participants with column and row headers to help guide and organize the notes the students took. After time to read the passage and study either the complete notes or the ones they had taken (in the partial condition), the participants took a test that consisted of several types of questions. Not only did students in the partial condition perform better on the test overall (particularly for application types of questions), but their higher scores were more consistent over a one week retention interval. Russell et al (1983) compared multiple-choice test scores of students based on three types of note taking: one group was given comprehensive notes about the lecture content; one group partial notes which contained an outline of the lecture material, tables and figures; and a skeleton notes group which received only a brief outline of the lecture information. The skeleton notes groups had higher test scores than the other two groups on both immediate and delayed (2 or 4 week) testing.

Educational Implications

In reviewing what educational practices are predicted by research on the generation effect to lead to higher test scores, we have seen evidence that this basic empirical finding does transfer to more practical situations. Making students responsible for the production of information both during note taking and also during study does appear to improve test scores, with the caveat that many students appear to need some external structure (e.g., outlines) during note taking to ensure that they are recording appropriate information. What other areas of education might be improved through the use of these generation principles? Although research evidence is lacking on the following suggestions, they proceed directly from the same principles above, which have provided evidence of effective transfer. The first suggestion has to do with a particular tendency we have noticed among our own students and has been the topic of discussion with other faculty as well: we find that students often want answers to questions in class that they could figure out for themselves. It seems to be either questions about information they have already received (perhaps in an earlier lecture), or questions about examples or applications of principles or theories they have learned. In many cases, these kinds of application questions are ones that students could answer on their own with a little cognitive effort (possibly instructor-guided). While it is certainly easier and more efficient for the class instructor to answer these questions (sometimes repeatedly) and move on to the next block of material, doing so may hinder students' learning. Clearly, the traditional Socratic method of using questions and answers to lead individuals to self-discovery of knowledge is not an efficient way to teach a classroom of 50 Introduction to Psychology students. However, helping them attempt to figure out the answers to some of their own questions either by having them look back through their notes or by asking questions that get them to think about how a theory explains behavior are likely to produce better learning than providing complete answers.

In addition to attempting to generate the answers to their own classroom questions, students who modify preferred study techniques to make them more generation-based are also likely to learn more. For example, making and using notecards is a common way students study course material. In preparing their cards, they often copy terms or definitions from their notes or textbook onto the cards, putting the answer on the back. Although this method is likely to benefit from the copying process compared to re-reading the original information, it is not particularly generative in that no manipulation of the information occurs. Teachers commonly tell students to put the information into their own words, which is a more generative process, but one with which our students often struggle. They seem to have difficulty translating technical information into language with which they are familiar. The first author of this chapter has found success in helping students to make this conceptual leap by telling them to try to limit the number of words that can be written on any card (say, to perhaps half of the source original ones). For example, instead of copying the complete definition of *schema* as “an expectation about objects, events, others’ or our own behavior based on past experiences,” students might read this definition in their notes and write “we think things happen because they have before”, or “expectations based on something’s history”, or even “expectation, things, people, based on the past” (full sentences are not necessary). Limiting word counts forces students to re-word or paraphrase more effectively. Is this more work than copying the original definition? Of course it is. Is it likely to lead to better memory for the concept of a schema over time? Yes.

Reading text and copying that information onto notecards is less work than creating study questions and answers. Reading over complete lecture slides received from an instructor or downloaded from a website is easier than taking one’s own notes in class. Answering a students’ question in class is easier than leading them through the process of coming up with that answer on their own. At some point, we need to value the work that goes into effective learning instead of trying to find the easiest path to knowledge (see Clark and Bjork, this volume, for an extended discussion on desirable difficulties).

References

Barnett, J. E. (2003). Do instructor-provided on-line notes facilitate student learning? Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois, April 21–25.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35,* 201-210.

DeWinstanley, P. A. (1995). A generation effect can be found during naturalistic learning. *Psychonomic Bulletin & Review, 2,* 538-541.

Foos, P. W., Mora, J. J. & Tkacz, S. (1994). Student study techniques and the generation effect. *Journal of Educational Psychology, 86,* 567-576.

Katayama, A. D. & Crooks, S. M. (2003). Online notes: Differential effects of studying complete or partial graphically organized notes, *Journal of Experimental Education, 71,* 293–312.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal, 29,* 303-323.

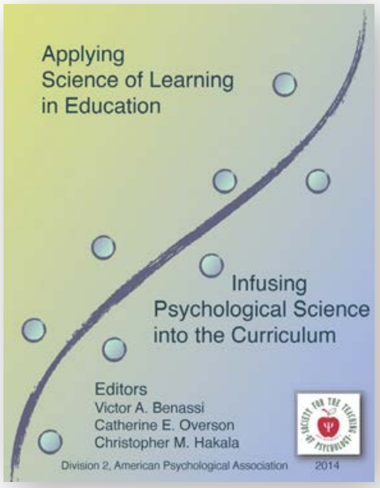
Kobayashi, K. (2005). What limits the encoding effect of note-taking?: A meta-analytic examination. *Contemporary Educational Psychology, 30,* 242–262.

Paris, S. G., Wixson, K. K., & Palincsar, A. S. (1986). Instructional approaches to reading comprehension. *Review of Research in Education, 13,* 91-128.

Russell, I.J., Caris, T. N., Harris, G. D., & Hendricson, W. D. (1983). Effects of three types of lecture notes on medical student education. *Journal of Medical Education, 58,* 627-636.

Thomas, J. W. & Rhower, W. D. (1986). Academic studying: The role of learning strategies. *Educational Psychologist, 21,* 19-41.

Van Blerkom, D. L., Van Blerkom, M. L., & Bertsch, S. (2006). Study strategies and generative learning: What works? *Journal of College Reading and Learning, 37,* 7-18.



Test-enhanced Learning

Mary A. Pyc, Pooja K. Agarwal, &
Henry L. Roediger, III
Washington University in St. Louis

Overview

Test-enhanced learning refers to the finding that practice tests during study increase the likelihood that information will be recalled at a later time. This view contrasts with the traditional idea that tests merely measure prior learning without changing it. To the contrary, a wealth of research has shown that tests are potent learning events (for recent reviews see Rawson & Dunlosky, 2011; Roediger & Butler, 2011). Compared to simply rereading a given piece of information (or doing nothing at all), actively retrieving information from memory improves the likelihood that the information will be recalled at a later time. The benefits of practice tests have been documented hundreds of times and the benefits of tests may last up to nine months after information has been initially learned (see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Practice tests have been shown to improve long-term retention across the lifespan, from children (e.g., for a review see Agarwal, Bain, & Chamberlain, 2013) to older adults (e.g., Logan & Balota, 2008) and have been documented using a wide range of materials and experimental designs, from word lists and foreign language word pairs to expository texts and maps (e.g., for a review see Roediger & Karpicke, 2006a).

Test-enhanced learning has been referred to by a number of different terms over the years; testing effects, retrieval practice effects, retrieval-enhanced learning, low stakes quizzes, and practice tests, to name a few. Research has shown that the benefits of practice tests are not dependent upon the type of test implemented (e.g., multiple choice, short answer, essay questions) (e.g., Foos & Fisher, 1988; Kang, McDermott, & Roediger, 2007; McDaniel, Roediger, & McDermott, 2007; McDermott, Agarwal, D'Antonio, Roediger, & McDaniel, in press). The important point here is that any type of practice testing enhances long-term retention.

Practice tests can be used both in the classroom by instructors and during self-regulated study by students. In the classroom the procedure can be as simple as frequently asking students questions, requiring all students to respond in some way (discussed further below) and providing them with feedback (i.e., the correct answer). For students, it can be as simple as using flashcards to study information or trying to recall information they have recently read or been lectured on. Many textbooks already incorporate practice testing into learning by asking questions at the end of chapters. However, students may not use these end-of-chapter questions optimally without help from the teacher (e.g., Karpicke, Butler, & Roediger, 2009). In particular, students need to practice retrieval several more times after they think they know the issue at hand well (Karpicke, 2009).

The idea of additional testing in education often suffers a bad reputation, but this is due largely to issues surrounding the use of standardized tests that have little to do with our recommendation of retrieval-enhanced learning. In the current chapter, we advocate that retrieval practice via testing should be incorporated into the classroom and into students' self-regulated learning as frequently as possible. Below we describe specific examples of how testing has been experimentally evaluated both in the laboratory and in actual classroom settings, and we offer suggestions for how instructors and students may easily incorporate testing into classroom and self-regulated learning situations.

Recent Research on Test-Enhanced Learning

The memorial benefits of practice tests have been documented across a wide range of materials including foreign language word pairs (e.g., Pyc & Rawson, 2009), text passages (Roediger & Karpicke, 2006b), statistics (e.g., Lyle & Crawford, 2011), general knowledge facts (e.g., Carpenter, Pashler, Wixted, & Vul, 2008), world history (e.g., Roediger, Agarwal, McDaniel, & McDermott, 2011), science (McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011) and map learning (e.g., Carpenter & Pashler, 2007).

The benefits of retrieval practice are more pronounced as the delay between encoding and a final test increases. For example, Roediger and Karpicke (Experiment 2, 2006b) had students initially read a text passage. Students then had three more practice trials in which they either read the passage three more times (SSSS), read the passage two more times and then took one free recall test (SSST), or did not read the passage again and took three free recall tests (STTT). Participants then completed a final free recall test either five minutes or one week later. The test involved just giving the students the name of the passage and asking them to recall as much of it as they could. Researchers scored the number of idea units (simple content phrases the text was parsed into for scoring purposes) recalled. Results are reported in the figure below. For the groups that had the final test five minutes after learning, the greatest benefit came from reading the text multiple times, and the single study-multiple test condition showed poorest performance. In other words, on an immediate test, cramming works. However, when the delayed final test was given one week later, the opposite pattern of results were obtained: Recall was best when students had been tested three times during practice and worst when they had read the passage four times. This outcome occurred despite the fact that students rereading the text were exposed to 100% of the material each time whereas students being tested were only exposed to what they could recall. These results nicely illustrate the powerful influence of testing on long-term retention and highlight the interaction between learning condition and retention interval (the delay between learning and final test).

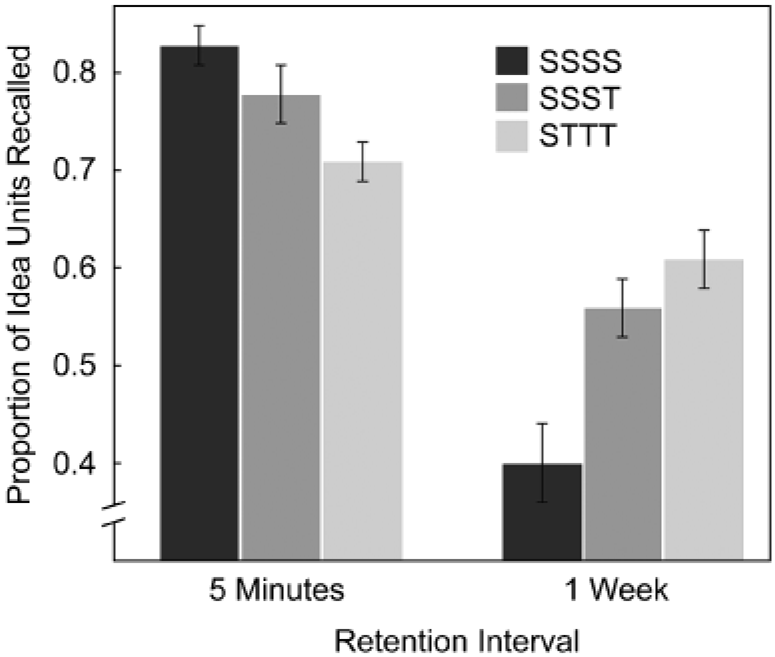


Figure 1. From Roediger & Karpicke (2006, *Psychological Science*, volume 17, issue 3, p 253, Copyright ©2006 Association for Psychological Science, reprinted by permission of SAGE publications). On an immediate final test, performance is highest when practice involves more study trials versus test trials. However, on a delayed final test, performance is highest when practice involves more test trials versus study trials.

The benefits of practice tests also obtain when retrieval from memory is covert (thinking of, but not explicitly stating a response). For example, Carpenter and Pashler (2007) had students learn maps so that they could recreate the map on a later test (see a sample map below). After initial study, students either studied the map again or engaged in covert test practice. For test practice, features (e.g., house) were removed and students were asked to think about which feature was missing. Thirty minutes later, students were asked to recreate the map with all of the features they could recall. Results showed that students who engaged in covert practice tests were more accurate in their final drawing than were those who received study practice. These results demonstrate that the memorial benefits of practice tests are not dependent upon overt retrieval of information. Additional research has shown that covert retrieval practice is as good as overt practice in benefitting later retention (Smith, Roediger, & Karpicke, in press) and, further, that it does not matter whether students using overt practice type or speak their answers on the initial tests – both methods produce a robust testing effect (Putnam & Roediger, 2013).

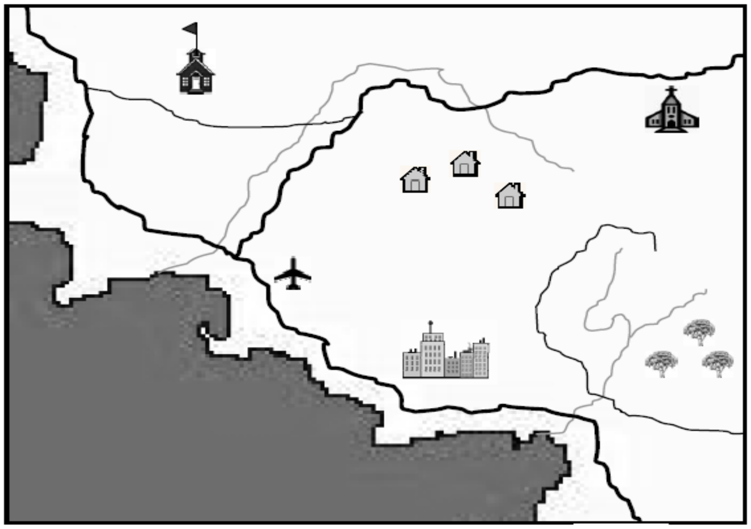


*Figure 2***.** Carpenter and Pashler created two maps, one of a town (upper panel), and one of a recreational area (lower panel). Each map contained four roads, two rivers, and six additional features such as a school or golf course. The original maps contained colors such as blue (for the lake, ocean, and rivers), green (for the trees and golf course), and red (for the school and first aid station). Each subject learned both of the maps, one through completing a test with feedback (test/study) that lasted 120 sec, and the other through engaging in additional study time (study) for 120 sec. Each subject was randomly assigned to one of four different counterbalancing conditions to learn both of the maps.

With kind permission from Springer Science and Business Media, Springer and the Psychonomic Society (Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review, 14* (3)*,* 474-478. Psychonomic Bulletin & Review, Figure 1, Copyright 2007 Psychonomic Society, Inc) given to the publication in which the material was originally published.

The benefits of practice tests even obtain when information is not recalled, provided that the correct answer is given after the test (e.g., Hays, Kornell, & Bjork, 2013; Kornell, Hays, & Bjork, 2009). Across a series of experiments, Kornell et al. (2009) evaluated the extent to which the benefits of practice tests are dependent upon correctly retrieving information. Students were asked to learn weakly associated items (e.g., whale – mammal) so that they could recall them on a later test. Students were assigned to either study the items or to receive a test trial (followed by feedback) with the items. For study trials, the entire pair was presented together for students to study. For test trials, only the first item was presented and students were to guess the correct answer and were then provided with the correct answer. Given that the items were weakly associated, almost all items were incorrectly recalled on test trials. After a delay, students received a final cued recall test in which they were provided with half of the pair (e.g., whale) and asked to generate the word associated with it (e.g., mammal). Results are presented below for four of their experiments and show that students who received the practice test trial recalled a greater proportion of items on the final test compared to students who received the practice study trial. Thus, even when retrieval is unsuccessful, practice tests are beneficial for memory.

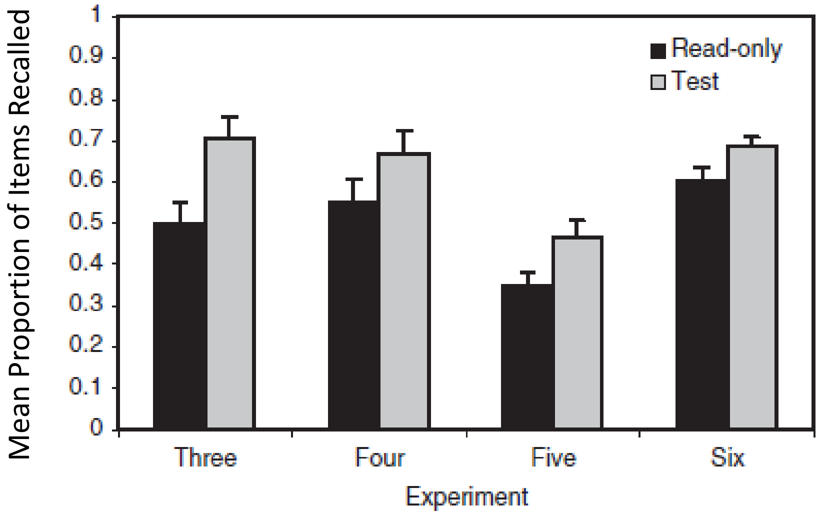


*Figure 3***.** Reprinted with permission (from Kornell, N., Hays, M. J., & Bjork, R. A., 2009. Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *35*, 989-998, published by APA). Final test performance is greater after test versus study practice even when items are initially not retrieved during encoding.

Although the majority of research has been conducted on college-aged students, a growing body of work has shown that the benefits of testing extend to all age groups from young children to older adults (e.g., Logan & Balota, 2008; Metcalfe & Kornell, 2007). For example, Rohrer, Taylor, and Sholar (2010) had children in the fourth and fifth grades learn the names and locations of cities on a map. The children learned the map by either studying the map or having a practice test (followed by feedback) of the cities on the map. They assessed performance on a final test one day later and the test was either the same as the prior test (standard) or a transfer test (designed to evaluate how information generalizes from one context to another). The figure below (adapted from Rohrer et al., 2010) shows accuracy for each practice group as a function of the final test type. For both types of final test, performance was substantially greater when the maps had been learned via practice tests versus concentrated study.

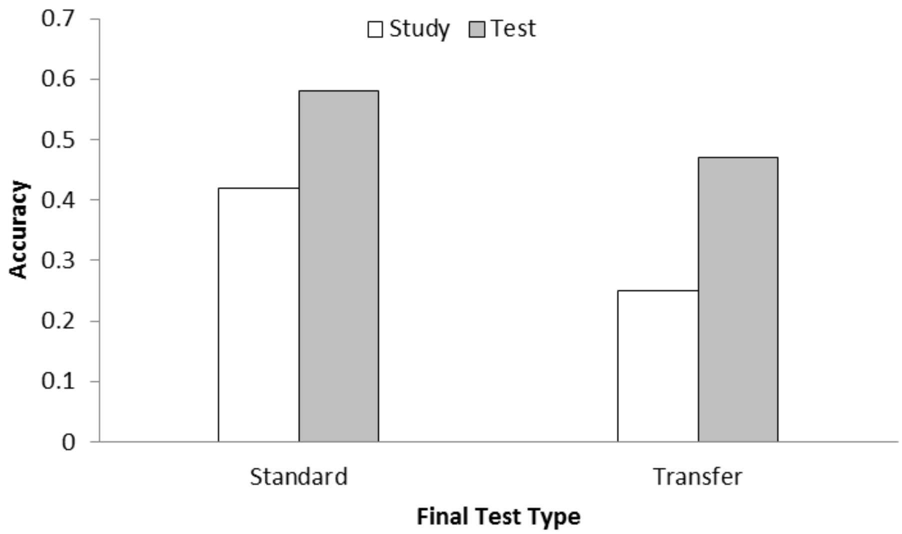


*Figure 4***.** Replotted from Rohrer et al., 2010. In a study with young children, final test accuracy was greater when the children engaged in test versus study practice. The effect was robust for both standard and transfer final test formats.

Although prior research has firmly established the memorial benefits of practice tests (compared to study), research has shown that the benefits of practice tests are even greater when tests are followed by feedback (e.g., Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Cull, 2000), especially for items that are incorrectly recalled on test trials (Pashler, Cepeda, Wixted, & Rohrer, 2005). For example, Pashler et al. (2005) had students learn Luganda vocabulary items (e.g., leero – today) so that they could recall the English word when prompted with the foreign language word. After items were initially studied two times, students had two test trials with each item in which they were provided with the foreign language word (e.g., leero) and asked to retrieve the English translation (e.g., today). Of interest for present purposes, students either received no feedback after test trials or were provided with the correct answer. One week later they completed a final test with each item and they were again provided with the foreign language word and asked to recall the English translation. Results showed that performance was greatest on the final test when students were provided with the correct answer after test trials during practice. Butler, Karpicke, & Roediger (2008) showed that, for more complex materials, feedback can also help students even when they get the answer correct (and especially when they provide a correct answer with low confidence).

In addition to research demonstrating that the memorial benefits of practice tests are greater than practice involving restudy of information, research has also shown that the amount and timing of tests influences the likelihood that information will be recallable at a later time. For example, Pyc and Rawson (2009) had students learn Swahili-English word pairs (e.g., wingu-cloud) until items could be recalled to a certain criterion level of performance during practice (between 1 and 10 correct recalls). They manipulated the amount of time between correct retrievals during practice, such that some items had a short lag (6 other items) between test trials and others had a long lag (34 items) between tests. Students were given a final recall test either 25 minutes after learning items or one week later. The figure below shows performance on the final test as a function of the number of times items were recalled (i.e., criterion level) and the amount of time between correct retrievals (short or long lag, which was about 1 or 6 minutes between retrievals) during practice. Results showed that performance increased as the number of correct recalls during practice increased (see the figure below, adapted from Pyc & Rawson, 2009). Results also showed that the amount of time between correct recalls substantially influenced the likelihood that information could be recalled on a later test. Performance was always greatest when there was a long lag versus a short lag between correct recalls. In fact, when the final test occurred one week later and items were learned with a short lag, performance was on the floor. Even ten successful retrievals a week before did not help learning the material (relative to one retrieval). Thus, for the short lag condition, the extra time spent practicing retrieval was essentially wasted for these students. This outcome points to the importance of having spaced retrieval. When retrieval is made too easy, retrieval practice fails to benefit recall (see too Karpicke & Roediger, 2007).

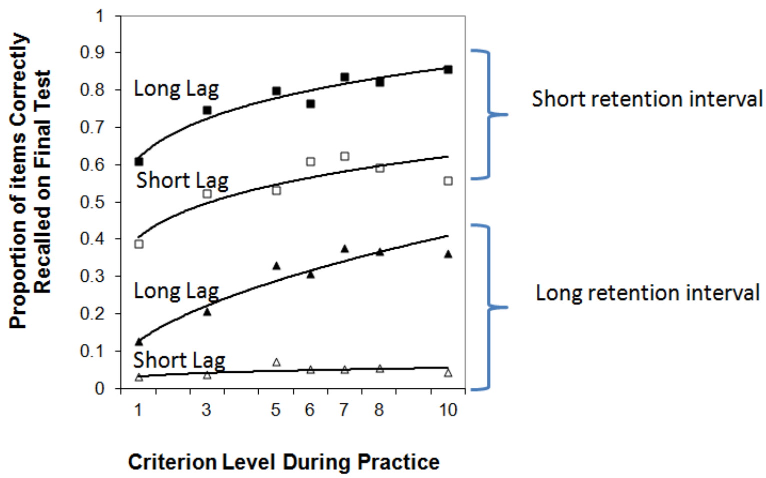


*Figure 5***.** Replotted from Pyc & Rawson (2009). Performance improves as criterion level increases and is greater for long lag versus short lag items. The memorial benefits of lag are particularly robust at a long retention interval.

Importantly, research has shown that the memorial benefits of practice tests are not dependent upon the match between the format of the practice test and final test. For example, Carpenter and DeLosh (2006) had participants learn lists of words. After an initial study trial with items, participants either studied the list again, had a multiple choice test, a cued-recall test, or a free recall test. After a short retention interval participants were asked to recall the list of words they had just learned. Depending upon the group to which each participant was assigned, the format of the final test was multiple choice, cued-recall, or free recall. Performance was greatest on the final test, regardless of the final test format, when participants received a free recall test during practice. These results suggest that when students exert more effort during encoding (i.e., free recall vs. cued-recall or multiple choice), they perform better on a final test, regardless of the format of the final test (free recall, cued-recall, or multiple choice formats). However, more recent work has shown that the format of practice tests does not influence the size of testing effects (McDermott et al., in press).

How Can Test Practice Be Incorporated Into the Classroom?

Research has shown that implementing frequent low-stakes practice tests in the classroom can lead to a substantial boost in performance on exams in actual classroom settings (e.g., Leeming, 2002; Lyle & Crawford, 2011; McDaniel et al., 2011). There are many different ways that instructors can incorporate test practice into everyday activities in the classroom. Here we cover four different strategies, but of course, others may also work well depending upon the structure of the class and preferences of the instructor.

Frequent questioning during course instruction is one way for instructors to incorporate testing into the classroom. This involves the instructor asking a question , waiting for students to generate a response, and then randomly choosing any student in the class to answer the question. This requires that all students generate an answer, so all students will benefit from testing. It is important that instructors provide feedback to students, so that they may correct any wrong answers they may have generated. We realize that some instructors may not want to put students on the spot, and would prefer to have volunteers provide answers. This could also work successfully, but we encourage instructors to emphasize the importance of all students generating (and perhaps writing down) their own answers to a question, before any student answers out loud. In this way all students should reap the benefits of engaging in test practice.

A number of other strategies incorporate test practice into classroom instruction and do not isolate one student (e.g., using student response systems or "clickers," giving quizzes at the beginning or the end of class and others). Clickers are remote devices for personal responding, and they have received increasing interest in recent years. For a clicker test, the instructor asks a question and students make their response by choosing between options to determine which is the most appropriate answer (i.e., multiple choice test) or by typing in their answer via texting (i.e., short response tests). Clickers engage *all* students, by requiring everyone to respond to a given question. The clicker results are also informative to instructors because they can provide information about how many students understand a given concept or topic. This can serve to modify instruction by highlighting information that may need to be covered in more detail. One downside to clicker use is that they do involve some cost (although they can be used for many students across years). A reminder: feedback should always be giving during or just after the clicker question or quiz in class.

A very similar and nearly zero cost method for incorporating quizzing into classroom instruction would involve providing students with a set of colored index cards at the beginning of a course with instructions to bring the card to class every day. Each card could represent an answer (e.g., true/false or a, b, c, d options for a multiple choice question). Throughout the class period, whenever an instructor asks a question, students would choose the colored flashcard representing the answer they believe is correct and hold it up. Like clickers, this strategy also provides instructors with immediate feedback about the number of students who understand a given concept yet costs very little to implement in the classroom.

Another simple strategy for incorporating frequent test practice into the classroom is to give quizzes at the beginning or end of each class. The basic idea behind each of these is to test students during every class period. A few minutes at the beginning or end of each class is devoted to test practice. Instructors ask students to answer a couple of questions either before class begins (covering topics covered in previous classes and the reading assignment) or at the end of each class (covering topics covered during the class) and students write their answers on a piece of paper and hand them in to the instructor. Students are encouraged to study frequently because they know they will be tested on information.

For each of the strategies just described, we highly recommend that they involve low stakes testing for students so that they do not significantly influence grades (which would be largely determined in the usual ways – papers, final exams, etc.). The most important aspect of each of these strategies is that they involve frequent testing in the classroom. Although some additional time will be required of the instructor to generate the questions and grade the answers, we believe this is time well spent. Frequent practice testing will enhance learning and long-term retention, which may ultimately save instructors time because they will not have to re-teach information as often. These quizzes also offer instructors immediate feedback regarding how much information students are learning. If all students perform poorly on a particular set of questions, this may serve as a cue to the instructor to review the concept or topic. Finally, we emphasize the importance of instructors providing feedback after students have engaged in test practice so that they may correct any errors that may have occurred during test practice. One point we have not emphasized here is that quizzing has many benefits besides the direct effect of retrieval practice (see Roediger, Putnam & Smith, 2011, who discussed ten benefits of testing).

How Can Testing Be Incorporated Into Student's Self-Regulated Learning?

Although there has been a surge in programs that students can buy that claim to improve learning and long-term retention, these can often be quite costly. Below we cover a number of strategies that students can use during self-regulated learning that cost nothing.

Students may already incorporate some forms of testing into their self-regulated learning. One common study strategy that students report is the use of flashcards during self-regulated learning (e.g., Hartwig & Dunlosky, 2012; Wissman, Rawson, & Pyc, 2012) of basic facts and concepts. Flashcards naturally incorporate test practice by having a question on one side and the answer on the other. So, students are prompted to try generating an answer and then immediately checking to see if they were correct. Students also report practicing with flashcards until items can be correctly recalled (e.g., Wissman et al., 2012). As discussed above, research has shown that the long-term benefits of retrieval practice improve as the number of correct recalls during practice increases. Therefore, students can incorporate a powerful learning strategy into their self-regulated learning through use of flashcards, provided that they continue to review information until they can correctly recall items multiple times at spaced intervals. In other words, students should be encouraged to keep cards in their flashcard "deck" for as long as possible and not to drop any cards out, to maximize the amount of information that is tested (e.g., Karpicke & Roediger, 2008).

Another way that students may already incorporate test practice into their learning is by answering the questions that are often found at the end of textbook chapters. We would recommend that students first answer each of the questions to the best of their ability without looking up the answers and then check the accuracy of their answers. Students can also periodically stop while they are reading a chapter and try to recall what they have read (and then check back for any important information they may have missed).

Are There Benefits of Test-Enhanced Learning Beyond Promoting Long-Term Retention?

Although we believe the long-term memory benefits from practice tests are reason enough to widely incorporate testing into classroom settings, there are a number of other benefits of testing (see also Roediger, Putnam, & Smith, 2011). As we have alluded to above, one important indirect benefit of testing is that it helps both teachers and students immediately identify concepts or topics that are not well known. This may be particularly beneficial for teachers because it will highlight information that students are struggling with and may suggest that more time needs to be spent on a given topic. For students studying on their own, it can highlight which information they need to concentrate on more heavily.

Research has also shown that open-ended practice tests (like free recall of what students know about a topic) improve student's organization of information. Compared to simply restudy, testing leads to enhanced organization of information, which aids later retrieval (e.g., Zaromb & Roediger, 2010). The benefits of testing extend beyond the particular item that is quizzed, showing transfer to other information that was not specifically tested during encoding (e.g., Butler, 2010; McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013). As noted above, researchers have become increasingly interested in the influence of practice tests on subsequent learning (e.g., Arnold & McDermott, 2012; Pyc & Rawson, 2010, 2012). That is, does the act of taking a test influence how well students learn information on study opportunities that follow the tests? Arnold and McDermott (2012) showed that as the number of practice tests increases, the amount of information learned during a subsequent study trial also increases. Similarly, Pyc and Rawson (2010) showed that the effectiveness of strategies that students use to learn items is greater when students engage in test practice versus study practice.

Summary

A wealth of research has established that practice tests enhance long-term retention. This chapter reviewed a small portion of the research that has been done in the area and focused primarily on studies done with educationally relevant materials. We offer some suggestions for how instructors and students can easily incorporate practice tests into the classroom and self-regulated study. We believe it is imperative to incorporate practice quizzes and tests as much as possible into the classroom and self-regulated learning to promote long-term student learning and retention. For a recent guide designed for instructors see "How to use retrieval practice to improve learning" (Agarwal, Roediger, McDaniel, & McDermott, 2014).

References

Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2013). The value of applied research: Retrieval practice improves classroom learning and recommendations from a teacher, a principal, and a scientist. Educational Psychology Review, *24*, 437-448. doi: 10.1007/s10648-012-9210-2

Agarwal, P. K., Roediger, H. L., McDaniel, M. A., & McDermott, K. B. (2014). How to use retrieval practice to improve learning. Available from http://www.retrievalpractice.org

Arnold, K. M., & McDermott, K. B. (2012). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 940-945. doi: 10.1037/a0029199

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61,* 213-238. doi: 10.2307/1170535

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1118-1133.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 918-928. doi: 10.1037/0278-7393.34.4.918

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34,* 268-276. doi: 10.3758/BF03193405

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review, 14,* 474-478. doi: 10.3758/BF03194092

Carpenter, S. K., Pashler, H., Wixted, J. T., Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*, 438-448. doi: 10.3758/MC.36.2.438

Cull, W.L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14,* 215-235. doi: 10.1002/(SICI)1099-0720(200005/06)14:3<215

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4-58. doi: 10.1177/1529100612453266

Foos, P. W., & Fisher, R. P. (1988). Using tests as learning opportunities. *Journal of Educational Psychology, 80,* 179-183. doi: 10.1037/0022-0663.80.2.179

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review, 19,* 126-134. doi: 10.3758/s13423-011-0181-y

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, Cognition, 39,* 290-296*.* doi: 10.1037/a0028468

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19,* 528-558. doi: 10.1080/09541440601056620

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*, 469-486. doi: 10.1037/a0017341

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory, 17,* 471-479. doi: 10.1080/09658210802647009

Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33*, 704-719. doi: 10.1037/0278-7393.33.4.704

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319,* 966-968. doi: 10.1126/science.1152408

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *35*, 989-998. doi: 10.1037/a0015729

Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology, 29,* 210–212. doi: 10.1207/S15328023TOP2903_06

Logan, J. M. & Balota, D. A. (2008). Expanded vs equal spaced retrieval practice in healthy young and older adults. *Aging, Cognition, and Neuropsychology,* 15, 257-280. doi: 10.1080/13825580701322171

Lyle, K. B. and Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology, 38*, 94-97. doi: 10.1177/0098628311401587

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology, 103*, 399-414. doi: 10.1037/a0021782

McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14*, 200-206. doi: 10.3758/BF03194052

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. K. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology, 27,* 360-372

McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., & McDaniel, M. A. (in press). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*.

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review, 14*, 225-229. doi: 10.3758/BF03194056

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 3-8. doi: 10.1037/0278-7393.31.1.3

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60,* 437-447. doi: 10.1016/j.jml.2009.01.004

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 333*, 335. doi: 10.1126/science.1191465

Pyc, M. A., & Rawson, K. A. (2012). Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 737-746. doi: 10.1037/a0026166

Putnam, A. L., Roediger, H. L. (2013). Does response mode affect amount recalled or the magnitude of the testing effect? *Memory & Cognition, 41,* 36-48. doi: 10.3758/s13421-012-0245-x

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General, 140,* 283-302. doi: 10.1037/a0023956

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied, 17*, 382-395. doi: 10.1037/a0026252

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20-27. doi: 10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J.D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210. doi: 10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249-255. doi: 10.1111/j.1467-9280.2006.01693.x

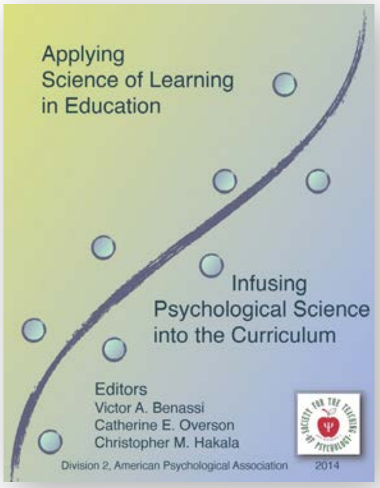
Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation: Cognition in education* (pp. 1-36). Oxford: Elsevier. doi: 10.1016/B978-0-12-387691-1.00001-6

Rohrer, D., Taylor, K. & Sholar, B. (2010). Tests enhance the transfer of learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36, 233-239. doi: 10.1037/a0017678

Smith, M. A., Roediger, H. L., & Karpicke, J. D. (in press). Covert retrieval practice benefits retention as much as overt retrieval. Journal of Experimental Psychology: Learning, Memory, and Cognition. doi: 10.1037/a0033569

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2012). How and when do students use flashcards? Memory, 20, 568-579. doi: 10.1080/09658211.2012.687052

Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organization processes. Memory & Cognition, 38, 995-1008. doi: 10.3758/MC.38.8.995



Supporting Self-Explanation in the Classroom

Jennifer L. Chiu
University of Virginia

Michelene T.H. Chi
Arizona State University

Overview

Self-explaining, or making sense of new information by explaining to oneself, can greatly enhance learning for students across ages, domains, and instructional types. Over two decades of research have demonstrated that explaining concepts while learning results in more effective problem-solving, more robust conceptual understanding, and better monitoring of learning. But how can a teacher apply these findings to his or her own instruction? In what settings and with what kinds of students can self-explanations be most effective? This chapter reviews self-explanation research to give teachers practical advice on how to incorporate self-explanation most effectively into their own classrooms.

What are self-explanations?

Classrooms often present information to be learned through text and/or words from an instructor. In order to learn, students must take information from various sources and construct meaning or understanding. Self-explaining, or making sense of new information by explaining to oneself, helps learners construct new knowledge by elaborating upon presented information, relating them to existing knowledge, making inferences, and making connections among given information. For example, in the seminal paper finding the *self-explanation effect* (Chi, Bassok, Lewis, Reimann & Glaser, 1989), students learned to solve basic mechanics problems by studying a worked-out example solution from a college physics textbook. One problem consisted of a written description of a weight hanging from three strings accompanied by a diagram (taken from Chi et al., 1989). Examples of self-explanations generated by students working this problem follow in quotations:

> The diagram shows an object of weight W hung by strings. Consider the knot at the junction of the three strings to be the body.
>
> ***"Why should this [the knot] be the body? I thought W was the body"***

> The body at rest under the action of the three forces shown…
>
> ***"I see. So, the W will be the force and not the body. OK…"***
>
> …of the three forces shown in the diagram.
>
> ***"Uh huh, so…so they refer to the point as the body."***

Chi et al. (1989) found that successful students spontaneously generated more self-explanations as compared to less successful students, that is, they uttered statements that indicated they refined their understanding of the problem and made connections to relevant principles. Even with no instructor, coach, tutor, or feedback on the explanation, just the act of engaging with the information through generating explanations improved learning. This *self-explanation effect*, or students learning better when they explain material to themselves, has replicated across domains such as biology (Chi et al., 1994), computer science (Pirolli & Recker, 1994; Recker & Pirolli, 1995), probability (Renkl, 1997) electricity and magnetism (Ferguson-Hessler & de Jong, 1990), and history (Wolfe & Goldman, 2005). More than two decades of research demonstrate the benefits of self-explanation with wide ranges of ages (e.g. Calin-Jageman & Ratner, 2005; Griffin, Wiley, & Thiede, 2008; Kastens & Liben, 2007; Pillow, Mash, Aloian, & Hill, 2002; Pine & Messer, 2000) and instructional formats (e.g. Ainsworth & Loizou, 2003; Roscoe & Chi, 2008; Trabasso & Magliano, 1996). The self-explanation effect is one of the twenty-five principles of learning from the Association for Psychological Science (Graesser, Halpern, & Hakel, 2007) and is one of the Institute for Educational Science's seven recommended learning strategies in the 2007 practice guide (Pashler et al., 2007).

Why are self-explanations helpful?

Generally, self-explaining is a constructive activity requiring students to actively engage in their learning process. Active participation is better than passive participation for learning (Bransford, Brown, & Cocking, 1999; Chi, 2008). How does self-explanation, in particular, help people learn? Two mechanisms have the most empirical support (Chi, 2000). First, self-explanations help learners identify gaps in their understanding and fill in missing information from the instructional material. For instance, when students are trying to digest information in a textbook, on the computer, or while solving a problem, there are inherent omissions in the information. That is, even the best materials cannot possibly contain all the information that is needed for every learner for the specific topic. Self-explanations help learners generate inferences to fill in those missing gaps in their own understanding of the topic.

The second way that self-explanations can help learning is by revising what a student already knows about the topic. Learners can come in with their own ideas, or their own mental models of a concept. These mental models are typically flawed. When a learner encounters instructional material that conflicts with their existing mental models, self-explaining helps repair and revise their understanding.

Supporting Self-Explanation: Description and Discussion

Although good learners typically self-explain more, not all self-explanations are equal. A closer look at the results from Chi et al. (1989) reveals qualitative differences between good and bad problem solvers. Good learners tended to spend more time studying examples, give explanations that were more principle-based, elaborate on the conditions and goals, and monitor comprehension. Poorer learners made less of these productive types of self-explanations and instead paraphrased and re-read materials. Additionally, there were no differences in prior domain knowledge or GPA between the good and poor learners (Chi & VanLehn, 1991). Other studies have replicated the findings that successful learners more often use particular kinds of self-explanations (Ferguson-Hessler & deJong, 1990; Pirolli & Recker, 1994) independent of prior domain knowledge. Studies have also reproduced the self-explanation effect with different media, such as studying with text and diagrams (Ainsworth & Loizou, 2003) or learning from multiple sources on the web (Goldman, Braash, Wiley, Graesser, & Brodowinska, 2012). Unfortunately, however, studies also show that only a small percentage of learners naturally generate productive self-explanations (Renkl, 1997).

Building upon the finding that successful learners self-explain, Chi et al. (1994) investigated whether *prompting* self-explanations could have similar benefits for learning. That is, if good learners naturally generate productive self-explanations, could instructional prompts elicit self-explanations from students to help them learn more deeply? Chi et al. (1994) asked eighth-grade students to self-explain after reading each sentence of a passage on the human circulatory system. A comparison group of students read the same passage twice but were not asked to self-explain. The students who were prompted to self-explain understood more about the circulatory system than the comparison group. Within the self-explain group, the students who generated more self-explanations learned better than students who did not self-explain as often. High self-explainers had more accurate mental models and understood complexities of the circulatory system that were not explicit in the text.

A wealth of studies has replicated this finding that explicit prompts for self-explanations encourages students to generate explanations and improves learning in various settings (e.g. Bielaczyc, Pirolli & Brown, 1995; McNamara, 2004; Renkl, 1997). For example, Griffin, Wiley, and Thiede (2008) compared three groups of college students: one group read a complex text once "as though studying for an exam" (p. 95), the second group read and then re-read the same text as if there was a test, and the third group read the text but was also instructed to self-explain while reading with a few example questions such as, "What new information does this paragraph add?" (p. 97). Students who were instructed to self-explain outperformed either the read-once or read-twice condition on assessments administered to all groups after the treatment. Thus, explicitly prompting can help students who may not spontaneously generate self-explanations learn with understanding.

Instead of having a coach or instructional support to prompt self-explanation at specific times, providing students with training on how to self-explain can help students learn. For instance, Renkl, Stark, Gruber, and Mandl (1998) investigated the impact of self-explanation training on understanding of compound interest with a population of vocational students in bank apprenticeships. A researcher modeled self-explanation and then provided coaching while learners practiced self-explaining on a warm-up problem. Learners who received self-explanation training outperformed learners who were trained to think aloud during instruction. Bielaczyc, Pirolli, and Brown (1995) trained university students to self-explain while learning about recursive functions in computer science. Students with self-explanation training explained more and performed significantly better on posttest problems compared to students that discussed, performed simple recall tasks, and wrote essays. McNamara (2004) created a self-explanation reading training program (SERT) that significantly improved undergraduate students' comprehension of science texts. Both the Bielaczyc et al. (1995) and McNamara (2004) training generally involved an introduction of self-explanation, videotaped models of learners self-explaining, and practice self-explaining with feedback from the trainer before instruction.

Eliciting self-explanations may benefit learning not only in the short term (where assessments are given immediately after instruction) but also foster transfer (Chi et al., 1994) and retention of concepts. Wong, Lawson, and Keeves (2002) compared the performance of ninth-grade mathematics students with self-explanation training to a control group of students using their typical studying methods. Self-explaining students outperformed the control group on geometry posttests, especially for far-transfer items that required application of the newly learned material to a novel situation. King (1992) had college students enrolled in a remedial reading course watch a lecture and take notes. One group of students self-explained lecture content, one group generated a summary of the content and another group reviewed lecture notes. On posttests immediately after the session on a retention test one week later, the explainers and summarizers outperformed students who reviewed lecture notes.

Instructional techniques to encourage self-explanation

Various kinds of instructional techniques can help students self-explain. Many studies have asked students for verbal self-explanations while studying texts or listening to lectures. Prompting students for written self-explanations has also been successful (Hausmann & Chi, 2002; Schworm & Renkl, 2007). Other studies have used computer-based systems to prompt explanations (Chiu & Linn, 2013; Wylie & Chi, in press) or have students select the most appropriate explanation for the given concept (Aleven & Koedinger, 2002). All of these techniques of eliciting self-explanations have been shown to benefit learning across many domains and ages.

Computer-based self-explanation programs can greatly increase access to the benefits of self-explanation. Resource-intensive approaches such as one-to-one human prompting or one instructor training a whole class at a time require training of human tutors and may or may not provide individual student attention. Computer-based environments that support self-explanation enable students to get one-on-one benefit without having a human present. Although students do not tend to spontaneously type self-explanations as much as they spontaneously verbalize self-explanations (Hausmann & Chi, 2002), studies demonstrate various successful approaches to computer-supported self-explanations (Recker & Pirolli, 1995; Wylie & Chi, in press). Some learning environments have students generate and type open explanations to specific prompts (Chiu & Linn, 2013; Schworm & Renkl, 2006), whereas other studies find benefit from prompts that involve selecting parts of explanations from drop-down menus (Booth, Lange, Koedinger, & Newton, 2013). For example, Aleven and Koedinger (2002) investigated the impact of drop-down self-explanations with an intelligent computer-based mathematics tutoring program in high school geometry classes. One group of students was prompted to explain each step of the geometry problems, whereas another group entered answers to each step but was not prompted to self-explain. The prompted explanation group significantly outperformed the regular problem-solving group on multiple posttest measures.

Similarly, computer-based training can be an effective alternative to human training for self-explaining. Building upon the success of the human trainers in the SERT program, McNamara and her colleagues developed a computer-based program called the Interactive Strategy Training for Active Reading and Thinking (iSTART) to help students understand complex science texts. iSTART coaches students in the particular successful self-explanation strategies, such as making inferences, anticipative predictions, or elaborations. Animated pedagogical agents provide the training by interacting with each other and the learner. Just like SERT, iSTART involves an introduction, demonstration, and practice of the strategies through teacher and student agents that give feedback on learners' self-explanations. Studies demonstrate that iSTART improves reading comprehension scores for students (McNamara, O'Reilly, Best, & Ozuru, 2006).

Optimal conditions for self-explanation

Although prompting explanations generally enhances learning, the actual explanation may be fragmented, partially correct, or even entirely wrong (Renkl, 2002). Helping students make better self-explanations and finding optimal conditions for self-explaining can make learning through self-explanation more effective. Studies have explored factors that may contribute to the quality or the impact of self-explanations, including the prior knowledge of the learner, instructional material and specific types of explanation prompts.

With regard to prior knowledge, self-explanation can help students with a range of expertise and abilities. Self-explanation has been found to be beneficial for low-knowledge students (O'Reilly, Best, & McNamara, 2004; Renkl et al., 1998) or students with no prior knowledge of the subject (DeBruin, Rikers, & Schmidt, 2007). Some researchers suggest that there may be a greater benefit of self-explanations with more knowledge to draw upon (Wong, Lawson, & Keeves, 2002). Many studies find self-explanation beneficial regardless of prior knowledge (Chi et al., 1994; Griffin, Wiley, & Thiede, 2008). The lack of a clear trend in these studies indicates that self-explanations can benefit students with different abilities in different ways. Learners may use self-explanations to fill in gaps of understanding or to repair existing mental models depending on their prior knowledge and the specific instructional context (e.g. Chi, 2000).

Other studies investigated how the format of material to be learned may influence the self-explanation effect. Findings suggest that if the instructional material contains diagrams, learners tend to generate more self-explanations (Butcher, 2006). For example, Ainsworth and Loizou (2003) compared learners with a text passage about the circulatory system to another group that had diagrams with text. Learners with the diagrams generated significantly more self-explanations and performed significantly better than those with only text.

Learning with multiple diagrams, or multiple forms of media, may benefit from prompting self-explanations (Roy & Chi, 2005; Wylie & Chi, in press). Rau, Aleven and Rummel (2009) found that sixth grade students learned more about fraction concepts using multiple graphical representations when prompted for explanations that helped students integrate information. In the Aleven and Koedinger (2002) study, the prompted explanations specifically helped students integrate verbal and visual information into robust declarative knowledge. Prompting self-explanations may help students use multiple representations in complementary, constraining, or constructing roles to fill in missing pieces or revise their understanding (Ainsworth, 1999; Chi, 2000).

The correctness and coherence of the instructional material can also impact how learners self-explain. Multiple studies show that including incorrect solutions or examples in study materials can be particularly helpful for learning with self-explanation (Booth et al., 2013; Siegler & Chen, 2008). Durkin and Rittle-Johnson (2012) grouped fourth and fifth grade students learning about decimals into two groups, one compared correct with incorrect examples and the other only compared correct examples. Students with both incorrect and correct examples learned more and made more connections back to concepts in their explanations. Self-explaining incorrect examples provides an excellent opportunity for students to either recognize gaps in understanding or repair faulty information, whereas explaining correct examples may be more likely to reaffirm students' existing understanding.

Ainsworth and Burcham (2007) investigated the impact of the coherence of the text on learning with self-explaining. University students with little prior knowledge were either given a text about the circulatory system without elaborations (minimally coherent) or a text with elaborations (maximally coherent). For instance,

"In the lungs, carbon dioxide leaves the circulating blood and oxygen enters it."

would be an example of a minimally coherent text. In contrast,

"In the lungs, carbon dioxide that has been collected from cells as blood has passed around the body, leaves the circulating blood and oxygen enters it."

is an example of a more coherent text because it provides elaborations that help the learner understand the underlying system (Ainworth & Burcham, 2007, p. 288). In some cases, providing a minimally coherent text can be beneficial because learners then increase the amount of self-explanations in order to understand the text (McNamara & Kintsch, 1996). Ainsworth and Burcham (2007) also divided learners into groups that did or did not receive self-explanation training. Learners who self-explained and learners given the maximally coherent text significantly outperformed those without self-explanation training or those with incoherent text. In addition, self-explanation had a greater effect on learning than providing coherent text. Students with the minimally coherent text actually generated more self-explanations, but the explanations were mostly to compensate for the lack of clarity and gaps in the text. Students with the maximally coherent text seemed to use self-explanations to detect and revise incorrect ideas. Results suggest that giving novice students maximally coherent materials with self-explanation training could be most beneficial.

Various levels of explanation support from prompts can influence what students learn. Studies demonstrate that a variety of prompts can be effective, from very open (Chi et al., 1994) to focused on specific concepts (Gadgil, Nokes-Malach, & Chi; 2012), to completely specified through a drop-down menu (Aleven & Koedinger, 2002). Different prompts can elicit different kinds of explanations and differentially impact learning. Berthold, Eysink and Renkl (2009) investigated how different levels of prompt support may affect learning about probability with multiple representations. Participants were either given open prompts ("why" questions about the content), assisting prompts that first required learners to fill-in blanks in an explanation before then completing an open explanation, or no prompts. Learners with both kinds of prompts understood procedural knowledge equally well, but the assisted group outperformed the open group on measures of conceptual knowledge. A closer look at the self-explanations revealed that both kinds of prompts led to equal numbers of connections to underlying principles in explanations, but the assisting prompts led to more high-quality explanations that integrated representations. Providing some sort of scaffolding or support for self-explanations may help foster deeper understanding when students have difficulty answering an open question.

Different kinds of prompts can also impact the kinds of self-explanations generated when studying complex subjects. For example, in the Schworm and Renkl (2007) study, argumentation skills depend on learners understanding both the structure and logic of the argument along with the specific content within the argument. Students were divided into four groups: one group had prompts that specifically targeted the skill of argumentation, one group had prompts that focused on the content, one group had both kinds of prompts, and the last group had no self-explanation prompts. Students with the argumentation prompts made significantly more self-explanations about argumentation, but content prompts made no difference on the number of self-explanations. Learners in the argumentation group also outperformed the other groups on measures of argumentation skills, but not other measures of declarative or content knowledge. Learners tended to make many spontaneous self-explanations about content regardless of explicit prompts. Findings suggest that the nature of the learning domain may interact with the kind of prompts. It could be that in order to engage in argumentation, one first has to understand the content of the argument, so learners tended to naturally self-explain the content. However, to understand the more general skill of argumentation, students needed to be explicitly prompted.

Prompt type can also interact with prior knowledge and learning environment. Novice students trying to learn new material may need certain types of prompts whereas students with more understanding may need different prompts (Renkl, 2002; Yeh, Chen, Hung, & Hwang, 2010). Wylie and Chi (in press) conducted a literature review of studies that prompted self-explanation in multimedia environments. They concluded that providing a more directed approach to prompts was more successful than open prompts in multimedia environments. This finding may be due to the nature of multimedia environments that typically have multiple resources that require active integration and reflection for maximum benefit (Moreno & Mayer, 2007).

Overall, the large body of research on self-explanation demonstrates that it is a powerful learning strategy that can be used across ages, domains, and instructional material. Future research can further distinguish how self-explanations can be used in multimedia and computer-based learning environments, further refine our understanding of the best kinds of prompts for particular learning contexts, and shed light on and provide evidence for how the mechanisms of the self-explanation effect work in various settings.

Instructional Advice

Incorporating self-explanation into instruction can be relatively straightforward and very beneficial for students. The following section provides advice on when and how to incorporate self-explanations into instruction and how to assess the impact of these interventions.

Move along the passive-active-constructive-interactive spectrum

Thinking about instructional interventions as *passive, active, constructive,* or *interactive* provides a general framework with which one can structure classroom activities. Based upon evidence from decades of research, interactive activities are better than constructive activities, constructive activities are generally better than active learning activities, and all are better than passive learning (Chi, 2009). Classifications are based on overt student activities – that is, what one can directly observe students doing. *Passive* activities are when students are not overtly engaged in any activity. Typical examples of passive activities are watching a lecture, video, or reading text without any additional action such as note-taking, asking questions, or discussing with a peer. Of course, students listening to lecture may be very active internally, mentally self-explaining or making deep connections. However, there are no overt behaviors to ensure those processes are occurring, so those activities are classified as passive. *Active* learning activities are where students are physically doing something during learning. Active behaviors include underlining, gesturing, highlighting, navigating a website, or copying down lecture notes. The overt physical activities ensure that learners are paying attention to some minimal degree. *Constructive* activities involve generating some kind of output that goes beyond the given instructional information. Examples of constructive activities include creating a concept map, drawing a diagram, solving a novel problem, and self-explanations. *Interactive* learning activities are when students engage in substantive dialogue with another, where the "other" can be a peer, tutor, teacher, or computer agent. Interactive activities require that the talk between partners contain dialogue indicative of building, sharing, and refining ideas and not just regurgitating or one person dominating the conversation. In general, activities within a category can be thought of as having equivalent effects on learning in comparison to activities in another category. For instance, an interactive activity will be better than a constructive activity. However, activities within a category can vary in their effectiveness. For example, two constructive activities like concept-mapping and self-explanation can have different impacts on learning.

Results from self-explanation studies can be interpreted within this framework. Self-explanation is a constructive activity and thus better than passive or active learning activities (as demonstrated by many of the aforementioned studies). In many cases, self-explanation activities have more benefit than other constructive activities (Fonseca & Chi, in press). Self-explanation can also be used to augment other constructive or interactive activities. For instance, an easy way to make self-explanation interactive is to have student dyads generate joint explanations.

Going beyond a passive lecture

Generally speaking, an easy way to improve instruction is to move along the passive-active-constructive-interactive spectrum. If you typically lecture to your students, you can incorporate something as simple as note-taking, or perhaps having all of your students gesture for some key concept. An example could be to have the class perform "the wave" to demonstrate energy transfer with transverse waves. A quick look across the classroom or lecture hall would ensure that students are actually writing down notes or standing up in time for the wave.

Better instruction would include constructive activities such as self-explaining. The reviewed studies provide insight into how one can implement self-explanation strategies into classes, even into large lecture hall settings. A number of studies demonstrate that explicit training in self-explanation provides great benefit to students (Bielaczyc et al., 1995; McNamara, 2004). Devoting part of an initial class or discussion section to self-explanation training involving an introduction, modeling (could be a video), and then individual or group practice with either text or problems (depending on the domain) could be an invaluable way to start the class. During class, one can incorporate a simple exercise such as having students read a relevant passage and then provide instructions to, for example, "Explain the meaning and relevance of each sentence or paragraph to the overall purpose of the text. Ask yourself questions like: What new information does this add? How does it relate to previous information?" (Griffin et al., 2008). Instructors can also give guidance as to how students can incorporate self-explanation strategies after class. For example, college and high school students who were taught to come up with and answer their own "why" questions following class outperformed students who were not given any direction (King, 1992).

Incorporating interactive activities can be as easy as having students generate joint explanations in dyads instead of explaining to themselves. A similar strategy is to prompt students to generate explanations that will be used to teach another student. Simple tweaks such as these can have a large effect on learning compared to only self-explaining (Coleman, Brown, & Rivkin, 1997; Hausmann, van de Sande, & VanLehn, 2008).

In each of these cases, care should be taken to confirm that students' actions are, in fact, aligned with the categories. For instance, one could set up a jigsaw activity where students build explanations to then share and collaborate to build a larger understanding of a topic (interactive). However, the actual jigsaw could just be students retelling their explanations therefore the activity is really only constructive (Fonseca & Chi, in press).

Use Instructional Materials that Facilitate Self-Explanation

In addition to incorporating self-explanation into instruction, one can also select instructional materials to complement or enhance the self-explanation effect. Based on the research studies, using multiple representations in instructional materials can facilitate self-explanations. Thus, if you have students self-explain a certain concept or procedure, try to include diagrams or technology-enhanced representations like simulations to help students generate self-explanations and learn with understanding.

Giving students incorrect examples can also augment learning through self-explanation. Providing opportunities for students to explain why something is wrong, whether it is an incorrect solution to a problem, an incorrect inference from a section of text, or an incorrect model of a system, enables students to confront and revise their own incorrect ideas.

Another way to augment self-explanation could be to provide more coherent instructional materials. Making sure that information presented to students carefully elaborates on concepts or connections and articulates what may even seem to be redundant or obvious can help students learn through self-explanation. Instructors, as experts, may gloss over or omit pieces of information because they understand the concepts well. Novice students often need these pieces to build understanding (Feldon, 2007). Giving students coherent text can help facilitate mental model revision through self-explanation (Ainsworth & Burcham, 2007).

Adapt to Specific Contexts

Depending on what you are trying to accomplish, the domain and the situation of the learners, instructors need to be able to adapt to use different kinds of strategies or prompts (Gadgil, Nokes-Malach, & Chi, 2012). For example, if you are teaching relatively new material that students may have little to no prior knowledge about, you may want to try self-explanations that encourage gap-filling prompts so that learners can develop a mental model of a situation (Nokes, Hausmann, VanLehn, & Gershman, 2011). Later in instruction when students have more prior knowledge about a subject, one can try approaches that encourage model revision, such as comparing incorrect alongside correct examples.

Assessing learning with self-explanation

An important consideration when learning with self-explanation is to look at the quality of the explanation itself. What are the students saying or writing? Are they just regurgitating bits of text or making connections to underlying principles? Do the explanations contain predictions about what is going to happen, try to go beyond the given instruction or do they just superficially gloss over what is already there? Students who make principle-based, anticipative, or inference-containing explanations benefit the most from self-explaining. If students seem to be failing to make good explanations, one can try to give prompts with more assistance. In practice, this will likely take iteration by the instructor to figure out what combination of content, activity and prompt provides the most benefit to students.

As with any educational intervention, it is vital that the assessment aligns with the instruction and targeted learning outcomes. Self-explanation activities are no exception. Instructors should pay special attention that their assessments cover the targeted concepts. For instance, an instructor could incorporate self-explanation activities while students learn about a particular physics concept but then give an assessment that requires students to solve a problem that is associated with that concept.

Although solving a problem could be a measure of some kind of transfer of the conceptual knowledge, conceptual items should be included if one is interested in the direct impact of self-explaining.

Acknowledgements

The second author is grateful for support from the Institute of Education Sciences (Award #R305A110090).

References

Ainsworth, S. (1999). The functions of multiple representations. *Computers and Education, 33*(2-3), 131-152.

Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self- explanation. *Learning and Instruction, 17*, 286-303.

Ainsworth, S., & Loizou, A. T. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science: A Multidisciplinary Journal, 27*(4), 669-681.

Aleven, V., & Koedinger, K. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*(2), 147-179.

Berthold, K., Eysink, T., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science, 37*(4), 345-363.

Bielaczyc, K., Pirolli, P., & Brown, A. L. (1995). Training in self-explanation and self-regulation

Booth, J., Lange, K., Koedinger, K., & Newton, K. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction, 25*, 24-34.

Bransford, J. D., Brown, A. L., & Cocking, R. (1999). How people learn. Washington, DC: National Academy Press.

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology, 98*(1), 182-197.

Calin-Jageman, R.J., & Ratner, H.H. (2005). The role of encoding in the self-explanation effect. *Cognition and Instruction*, 23(4), 523-543.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In Glaser, R. (Ed). *Advances in Instructional Psychology* (pp. 161-238). Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H. (2009). Active-Constructive-Interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1,* 73-105.

Chi, M. T. H., & VanLehn, K. (1991). The content of physics self-explanations. *The Journal of the Learning Sciences, 1*(1), 69-105.

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18,* 439-477.

Chi, M. T. H., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Chi, M.T.H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change,* (pp. 61-82). Hillsdale, NJ: Erlbaum.

Chiu, J. L. & Linn, M. C. (2013). Supporting knowledge integration in chemistry with a visualization-enhanced inquiry unit. *Journal of Science Education and Technology.* doi:10.1007/s10956-013-9449-5

Coleman, E. B., Brown, A. L., & Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientific texts. *The Journal of the Learning Sciences*, *6*(4), 347-365.

de Bruin, A., Rikers, R., & Schmidt, H. (2007). The effect of self-explanation and prediction on the development of principled understanding of chess in novices. *Contemporary Educational Psychology*, *32*(2), 188-205.

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*(3), 206-214.

Feldon, D. (2007). The implications of research on expertise for curriculum and pedagogy. Educational Psychology Review, 19(2), 91-110.

Ferguson-Hessler, M. G. M., & Jong, T. D. (1990). Studying physics texts: Differences in study processes between good and poor performers. *Cognition and Instruction, 7*(1), 41-54.

Fonseca, B., & Chi, M. T. H. (in press). Instruction based on self-explanation. To appear in R. Mayer & P. Alexander (Eds.), *The handbook of research on learning and instruction.* Routledge Press.

Gadgil, S., Nokes, T.J., and Chi, M.T.H. (2011) Effectiveness of Holistic Mental Model Confrontation in Driving Conceptual Change, *Learning and Instruction, 22, 47-61.*

Goldman, S., Braash, J., Wiley, J., Graesser, A., & Brodowinska, K. (2012). *Comprehending and learning from internet sources: Processing patterns of better and poorer learners. Reading Research Quarterly, 47*(4), 356-381.

Graesser, A. C., Halpern, D. F., & Hakel, M. (2007). 25 principles of learning. Washington, DC: Task Force on Lifelong Learning at Work and at Home. Retrieved from http://www.psyc.memphis.edu/learning/whatweknow/index.shtml

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, *36*(1), 93-103.

Hausmann, R. G. M., & Chi, M. T. H. (2002). Can a computer interface support self-explaining? *Cognitive Technology, 7*(1), 4-14.

Hausmann, R. G., van de Sande, B. & VanLehn, K**.** (2008). Are self-explaining and coached problem solving more effective when done by pairs of students than alone? In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Procedings of the 30th Annual Conference of the Cognitive Science Society*. (pp. 2369-2374). New York, NY: Erlbaum.

Kastens, K. A., & Liben, L. S. (2007). Eliciting self-explanations improves children's performance on a field-based map skills task. *Cognition and Instruction, 25*(1), 45-74.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal*, *29*(2), 303-323.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*(1), 1-30.

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes, 22,* 247-287.

McNamara, D. S., O'Reilly, T. P., Best, R. M., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research, 34*(2), 147-171.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*(3), 309–326.

Nokes, T., Hausmann, R., VanLehn, K., & Gershman, S. (2011). Testing the instructional fit hypothesis: the case of self-explanation prompts. *Instructional Science, 39*(5), 645-666.

O'Reilly, T., Best, R., & McNamara, D. S. (2004). Self-explanation reading training: Effects for low-knowledge readers. In K. Forbus, D. Gentner, & T. Regier (Eds.), Proceedings of the 26th Annual Meeting of the Cognitive Science Society (pp. 1053–1058). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning*. Washington, D.C.

Pillow, B. H., Mash, C., Aloian, S., & Hill, V. (2002). Facilitating children's understanding of misinterpretation: Explanatory efforts and improvements in perspective taking. *The Journal of Genetic Psychology*, *163*(2) 133-148.

Pine, K. J., & Messer, D. J. (2000). The effect of explaining another's actions on children's implicit theories of balance. *Cognition and Instruction*, *18*(1) 35-51.

Pirolli, P., & Recker, M. (1994). Learning strategies and transfer in the domain of programming. *Cognition and Instruction, 12*(3), 235-275.

Rau, M. A., Aleven, V., & Rummel, N. (2009). Intelligent tutoring systems with multiple representations and self-explanation prompts support learning of fractions. *Proceedings of the 14th International Conference on Artifical Intelligence in Education* (pp. 441–448). Amsterdam, the Netherlands: IOS Press.

Recker, M. M., & Pirolli, P. (1995). Modeling individual differences in students' learning strategies. *Journal of the Learning Sciences*, *4*(1), 1–38.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science: A Multidisciplinary Journal, 21*(1), 1-29.

Renkl, A. (2002). Learning from worked-out examples: Instructional explanations supplement self-explanations. *Learning & Instruction*, 12, 149-176.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23*, 90-108.

Roscoe, R.D., & Chi, M.T.H. (2008). Tutor learning: The role of explaining and responding to questions. *Instructional Science*, 36, 321-350.

Roy, M. & Chi, M.T.H. (2005). Self-explanation in a multi-media context. In R. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 271-286). Cambridge Press.

Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education*, *46*(4), 426– 445.

Siegler, R.S., & Chen, Z. (2008). Differentiation and integration: Guiding principles for analyzing cognitive change. *Developmental Science, 11*(4), 433-448.

solving. *Cognition and Instruction, 13*(2), 221-252.

strategies: Investigating the effects of knowledge acquisition activities on problem

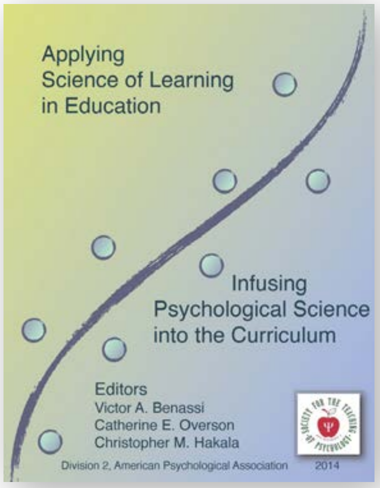
Trabasso, T., & Magliano, J.P. (1996). Conscious understanding during comprehension. *Discourse Processes, 21*, 255-287.

Wolfe, M., & Goldman, S. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction, 23*(4), 467-502.

Wong, R. M. F., Lawson, M. J., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction, 12,* 233-262.

Wylie, R. & Chi, M.T.H. (Submitted). The self-explanation principle in multimedia learning. To appear in R. Mayer (Ed.) *The Cambridge Handbook of Multimedia Learning, 2nd Edition.* Cambridge University Press.

Yeh, Y.F., Chen, M.C., Hung, P.H., & Hwang, G.-J. (2010). Optimal self-explanation prompt design in dynamic multi-representational learning environments. *Computers & Education*, *54*(4), 1089–1100.



Potent Techniques to Improve Learning from Text

Khuyen Nguyen and Mark A. McDaniel
Washington University in St. Louis

In many educational settings today, much of the learning takes place outside of the classroom. Though instructors play an instrumental role in shaping students' educational experiences, a primary source of acquiring knowledge is from textbooks. Because of the limited available class time, instructors typically assign extensive reading, and students spend countless hours trying to learn (or cram) from reading textbooks (though see Sikorski et al., 2002). This practice is potentially problematic because much research suggests that learners are typically poor at comprehending texts due to the complex nature of text comprehension (Gernsbacher, 1990; Kintsch, 1998; McNamara & McDaniel, 2004). To this end, it is imperative that students possess the proper techniques to help them learn from texts. Therefore, a major challenge for researchers and educators has been developing and promoting effective techniques to aid text learning.

A brief review of the educational literature indicates that there exist several techniques capable of improving text learning. Strategies such as Survey-Question-Read-Recite-Review (SQ3R) and concept mapping have been shown to be particularly effective for enhancing text comprehension under certain conditions (Novak, 2010; Robinson, 1941). Although these techniques have been successfully implemented in some classroom settings, a major limitation exists; these techniques require significant time investment (Martin, 1985; Novak, 2010). Not only do these techniques require extensive training in order to yield any benefits, they are also time-intensive for the student. Due to the heavy time demands already placed on students, they are unlikely to implement techniques that require much more time commitment than simply reading through the textbook. Thus, a major goal of the present chapter is to focus on effective techniques for improving text learning that are not so time-intensive, with the hope that students will increase the likelihood of implementing sound techniques for learning from their texts and other assigned readings.

First, we will discuss the merits of rereading and note-taking, which are popular study activities among students. Then, we will introduce three evidence-based techniques that have implications for improving learning from texts. For each technique, we will briefly discuss the empirical support and then suggest how each technique might be successfully implemented in authentic educational settings. Finally, we will discuss how individual differences might moderate the effectiveness of strategies designed to improve text comprehension.

Rereading

Before we discuss strategies to improve text learning, it is important to first consider how students typically study texts when left to their own devices. According to several recent surveys investigating students' study strategies, the widely preferred method of study among most college students is to simply reread the textbook (Hartwig & Dunlosky, 2012; Karpicke, Butler, & Roediger, 2009; Kornell & Bjork, 2007). This overwhelming preference to reread is potentially problematic because empirical evidence suggests that rereading has limited benefits, especially when compared to other study strategies (Callender & McDaniel, 2009; Roediger & Karpicke, 2006b). In four experiments, Callender and McDaniel (2009) had participants read educational texts that included either chapters from a psychology textbook or a *Scientific American* article and assessed learning with a range of tests including multiple-choice questions, short-answer questions, and summaries of the texts. The authors found that immediate rereading did not significantly improve final performance relative to reading the text once. Rawson and Kintsch (2005) found that the merits of rereading are a bit more complicated. In particular, the effectiveness of rereading may depend on when the rereading opportunity takes place and when the final tests are assessed. They found that spacing out the first and second reading opportunities was beneficial for performance on delayed tests relative to when the rereading took place immediately after the first reading. However, students are likely to engage in rereading relatively proximal to the exam. In this situation (immediate tests), Rawson and Kintsch found no benefit of spaced rereading. Although these findings suggest that rereading under some circumstances can be beneficial for performance relative to reading only once, it has been repeatedly shown that rereading has limited benefits when compared to other study strategies (e.g., McDaniel, Howard, & Einstein, 2009; Roediger & Karpicke, 2006b). While rereading might be useful in certain situations, there exist more potent strategies to improve text learning than simply rereading.

Note-taking

Taking notes is another commonly practiced strategy by students (Lahtinen, Lonka, & Lindblom-Ylanne, 1997). But unlike rereading, the benefits of note-taking are much more robust (See Armbruster, 2000 for a review). Slotte and Lonka (1999) investigated the benefits of spontaneous note-taking for high school students. Students were instructed to read a philosophical text and take as much or as little notes as they wanted. Half of the students were allowed to review their notes during the subsequent writing tasks while the other half completed the writing tasks without their notes. They found that quality and quantity of notes were predictive of text comprehension. Furthermore, being able to review notes produced benefits for deep-level text comprehension relative to not being able to review notes. These results suggest that note-taking has two main functions: encoding and storage (Di Vesta & Gray, 1972). The encoding hypothesis posits that the act of taking notes produces effective learning via generative processing (Einstein, Morris, & Smith, 1985; Peper & Mayer, 1978). That is, taking notes requires learners to actively process the information. The other hypothesis, the external storage account, argues that the benefit of note-taking comes from being able to review the notes (Kiewra, 1989). Both of these hypotheses rest on the idea that note-taking is only effective to the extent that students are actually summarizing, paraphrasing, or integrating rather than copying verbatim from the text (Kiewra, Benton, Kim, Risch, & Christensen, 1995). If students are simply transcribing the text, then there is presumably no active processing going on, and reviewing their notes becomes comparable to simply rereading the textbook.

Read-Recite-Review (3R)

Robinson (1941) introduced the Survey-Question-Read-Recite-Review (SQ3R) strategy to help students learn from texts. This strategy has several components that theoretically should be helpful, including organizing (survey), actively anticipating content (Question), practicing retrieval (Recite), and receiving feedback (Review). Accordingly, this strategy held some appeal; indeed, many of our contemporaries were introduced to the SQ3R strategy during their school years and encouraged by teachers to use it. But these contemporaries admitted to never using it, noting that it simply took too much time. This is a likely reason (as highlighted in the introduction) SQ3R has not continued to be promoted by teachers; most of the current students in our classes were not even introduced to the technique in their pre-college classes. Recently, McDaniel, Howard, and Einstein (2009) introduced a truncated version of the SQ3R that would serve to be more portable and time-efficient. Specifically, they dropped the survey and question facets and emphasized only the read-recite-review (3R) portion of the SQ3R strategy. In the 3R strategy, learners are instructed to read through a text passage once and then recite aloud as much of the information as they can without the passage in front of them. Following recitation, learners are given another self-paced opportunity to restudy the passage.

Thus, 3R maintains two components that are theoretically potent. One is the retrieval practice associated with recitation (see Roediger's Chapter in this book for an in-depth review on the benefits of retrieval practice). In one example of retrieval practice effects, Roediger and Karpicke (2006b) found that recall (retrieval) following reading was a more potent strategy than simply rereading. They had participants study a short text for as long as they wanted. Then half of the participants were instructed to free recall (i.e., testing) as much as they could whereas the other half were allowed to study the text for a second time. They found that testing conferred benefits on delayed tests of up to a week. The other crucial component of 3R is the review phase that follows recitation. Rereading following testing is assumed to be beneficial because learners can rely on improved metamemory (testing provides feedback about what the learner does or does not know) to effectively guide their restudy (during review).

To demonstrate the value of 3R, McDaniel et al. (2009) had participants study complex technical passages about brakes and pumps using one of three strategies: 1) 3R, 2) note-taking (plus restudy), or 3) highlighting (plus reread). In line with the theoretical predictions, participants who engaged in the 3R strategy performed much better on tests (free recall, inferential multiple-choice, and problem-solving) that were administered a week later, than those who highlighted. These benefits produced by the 3R strategy have also been found when learners are asked to write down their recitation as opposed to orally reciting the information (Nguyen & McDaniel, 2013a). Importantly, learners in the 3R conditions did not spend more time studying than those in the highlighting or note-taking conditions, suggesting that it is just as efficient as students' typical study strategies (at least for shorter texts).

However, the story became more complicated when the 3R technique was compared to note-taking, a strategy we mentioned above to be effective in promoting learning (See Armbruster, 2000, for a review). Specifically, while the 3R strategy conferred benefits on free recall performance relative to a note-taking condition, it did not improve inferential multiple-choice or problem-solving performance, measures that required integration and/or inferencing. From an educational perspective, these results are likely to be taken with mixed feelings. On the one hand, the 3R strategy seems to improve memory for the text, and does so in a more efficient manner (participants using the 3R strategy spent less time studying than those using a note-taking strategy). However, relative to another effective and more commonly used strategy (i.e., note-taking), the 3R technique does not improve performance on measures that require integration and inferencing. Because a major educational objective is to promote flexible learning beyond just rote memorization, it is important to endorse a strategy that enhances transfer or flexible use of knowledge (Anderson et al., 2001). To this end, Nguyen and McDaniel (2012) examined a potential enhancement to the standard 3R strategy that would augment performance on inference-based tests. Specifically, participants were required to make Judgments of Inferencing (JOIs) following recitation and prior to restudy. These JOIs probed participants' perceived abilities to integrate the information presented in the text and make inferences in order to apply this knowledge. Supplementing the standard 3R technique with JOIs enhanced performance on inference multiple-choice questions relative to a note-taking strategy that also included making JOIs (see Figure 1).

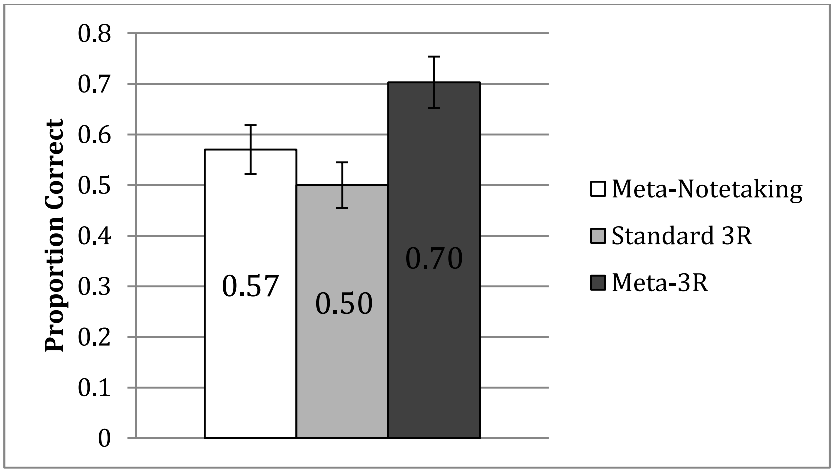


Figure 1. Average inference multiple-choice performance as a function of condition. Nguyen & McDaniel (2012).

Practical Considerations

These findings suggest that the 3R strategy can be an effective and time-efficient technique for enhancing students' text comprehension, and can be adapted for use in virtually any course. The unique value in the 3R strategy is that it is capable of enhancing both memory for the text content and acquisition of flexible knowledge. This makes it a potentially effective strategy for a wide range of classes. The standard 3R strategy can be applied in courses that require mostly memorization (e.g., physiology, foreign language, history). For classes that require using flexibly information obtained from the text (e.g., chemistry, physics), making JOIs following reading and reciting is a recommended addition to the 3R strategy.

It is important to stress that we are not advising students to recite an entire chapter at once. That is obviously not practical. Instead, students should break down their reading assignments into smaller sections and recite immediately following reading a particular section. This will allow them to not only learn that section better, but it should also indirectly help them learn future sections better for several reasons. First, testing (reciting) should help learners develop prior knowledge for the subsequent material, which is an important predictor of text learning (e.g., McNamara & Kintsch, 1996). Furthermore, testing has been shown to reduce proactive interference (Szpunar, McDermott, & Roediger, 2008), the finding of reduced memory for material learned later (e.g., Kalbaugh & Walls, 1973; Keppel & Underwood, 1962). Thus, once students have learned a particular section well, they are in better position to learn future sections. Instructors can further augment the effectiveness of the 3R strategy by having students make JOIs in class. Not only will making JOIs orient students toward a reading approach that fosters deep understanding (cf. Stine-Morrow, Gagne, Morrow, & DeWall, 2004), it can also give instructors feedback about how well their students are comprehending the material and to alter their lectures accordingly.

Embedded Questions

Embedded questions (also referred to as adjunct questions) are questions that are embedded at important parts of texts to probe a reader's understanding of what they have just read. Because they encourage learners to become active readers when studying texts, researchers and educators have promoted embedded questions for some time. Early on, researchers found that embedding questions within texts could improve learning from texts by orienting learners toward the key concepts in a given section (Anderson & Biddle, 1975; Andre, 1979; Hamaker, 1986; Hamilton, 1985; Melton, 1978; Rothkopf, 1966). Much research has confirmed this finding, which is why embedded questions are used in many textbooks.

The value of embedded questions may be augmented with the new trend toward e-textbooks. Using an e-textbook chapter, Nguyen and McDaniel (2013b) found that embedding questions in the text was a useful technique for enhancing comprehension. Specifically, participants read an e-text chapter on social psychology and answered six short-answer questions that were embedded in the text at the end of each section. To simulate realistic educational practices, participants were allowed to look back at the textbook when answering the questions. As seen in Figure 2 (the generated condition is discussed in the next section), answering the embedded questions (while reading) relative to simply reading improved students' performance on a final delayed (one-week) multiple-choice test that consisted of fact-based, inference-based, and applied questions. Additionally, study time did not differ between the two conditions, which suggests that answering embedded questions is actually an efficient study strategy.

While embedded questions are fairly standard in textbooks today, we get the sense from talking to undergraduates that few students actually make use of them. Thus, it is up to instructors and book publishers to encourage readers to actually make efficacious use of the embedded questions. In Nguyen and McDaniel (2013b), the e-textbook allowed participants to move on to the next section only after they submitted an answer to the embedded question. These results suggest that when made mandatory, answering questions embedded within the text can benefit learning from texts.

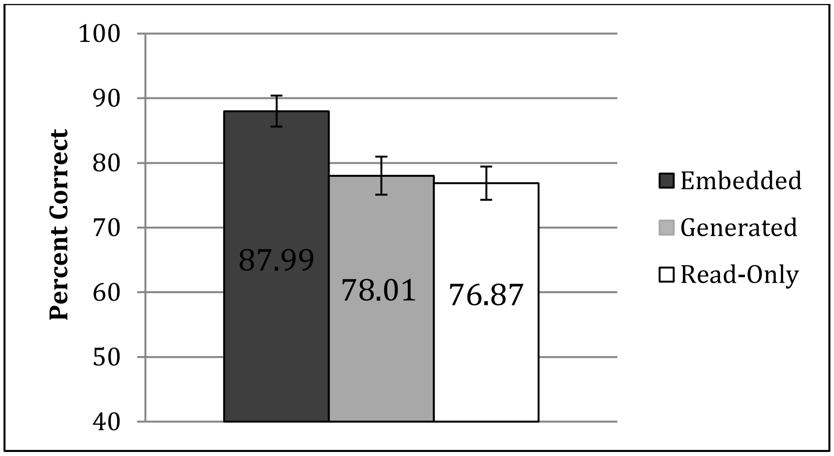


Figure 2. Average final test performance as a function of condition. Nguyen & McDaniel (2013b).

Practical Considerations

Following the theme of the present chapter, answering embedded questions did not prove to be a time-intensive task. Embedded questions are found in textbooks today. The problem is that few students may actually utilize them. Therefore, instructors play an important in encouraging students to answer these embedded questions when learning from textbooks. One way instructors can encourage students to answer these questions is to make it a part of the curriculum. Answering these questions can be part of homework assignments that count towards the final grade. Also, instructors can include a few of the embedded questions on exams to give students additional incentive.

Self-Generated Questions

While there has been plenty of research investigating the benefits of embedded questions on text comprehension, comparatively little research has examined the benefits of self-generated questions. Early on, research in the educational psychology literature viewed self-questioning as a strategy to promote active reading (Davey & McBride, 1986; Rosenshine, Meister, & Chapman 1996; Wong, 1985). Much of the research on self-questioning was motivated by the theoretical view that generating questions while reading would help learners focus and guide their reading. Because of this intuitive appeal, self-questioning became a reading strategy that was advocated by many educators (Wong, 1985). More recently, research studies have refocused an interest on self-questioning as a tool to improve text comprehension because of the current enthusiasm on the retrieval benefits of answering questions. Specifically, the robust benefits of retrieval practice have motivated researchers to examine whether self-generated questions could produce similar enhancements (as answering researcher-provided questions).

Weinstein, McDermott, and Roediger (2010) found that generating and answering one's own questions could be just as beneficial as answering questions provided by the experimenter. They had students study four short passages about Salvador Dali, the KGB, Venice, and the Taj Mahal using one of three strategies: 1) testing, 2) self-generation, or 3) restudy. In the testing condition, participants were given two questions to answer after reading each passage. In the self-generation condition, participants were trained on how to generate questions and then were asked to generate and answer two questions after reading each passage. In the rereading condition, participants simply restudied the passages for a second time. Following a short delay, a 24-item short-answer test was administered to probe comprehension of the passages. They found that both the testing and self-generation conditions performed significantly better on the final test than the restudy condition, and there were no differences between the two experimental conditions. This finding suggests that students do not need to depend on teachers to give them quizzes or tests in order to take advantage of the principle of retrieval practice. Rather, it seems just as effective if students generate and answer their own questions after reading.

With authentic educational materials (textbook chapters), however, the benefits of self-generated questions have been less robust. Nguyen and McDaniel (2013b) found that generating questions was not an effective technique for learning texts relative to answering questions embedded in the text (see Figure 2). This could have been due to the fact that learners did not receive guidance on how to generate questions (i.e., type of question; cf. Weinstein et al., 2010). That is, some learners may have generated detailed questions that could have been answered by copying a phrase verbatim from the text. These questions may not be as effective for promoting learning compared to questions that require learners to integrate information and make inferences.

Bugg and McDaniel (2012) examined the possibility that the benefits of self-generated questions may depend on the type of questions that are being generated. Specifically, they manipulated the types of questions that participants generated by training them to generate two different types of questions. Conceptual questions required participants to integrate information from two or more sentences within a given passage. Detailed questions, on the other hand, only required information from one sentence. As seen in Figure 3, participants who were instructed to generate/answer conceptual questions improved their conceptual test performance, whereas there were no benefits for participants who generated/answered detailed questions. These findings suggest that self-generated questions are only beneficial to the extent that they probe conceptual knowledge rather than detailed verbatim information taken from the text.

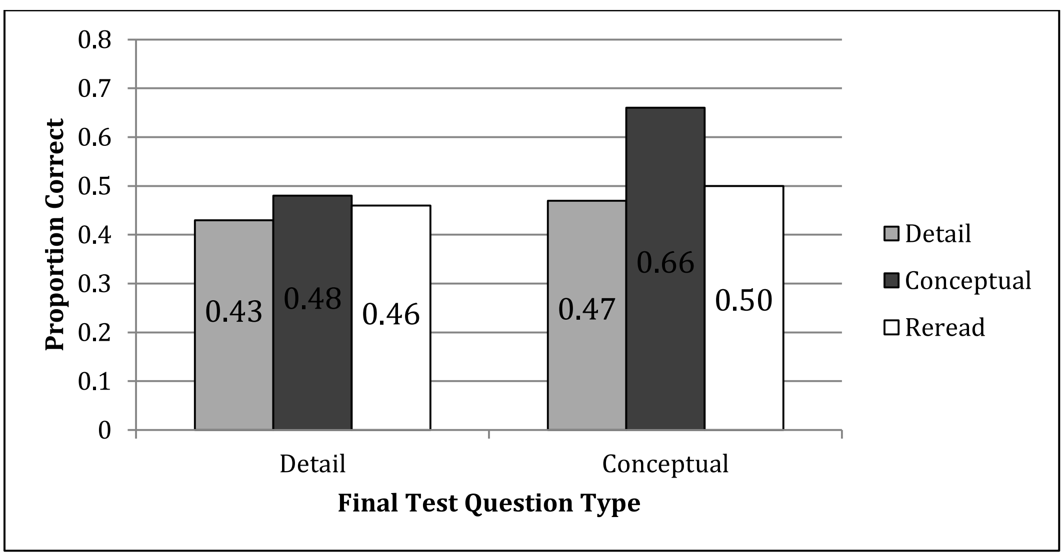


Figure 3. Average proportion correct as a function of condition and question types. Adapted from Bugg & McDaniel (2012).

Practical Considerations

Taken together, these findings suggest that generating and answering one's own questions while reading can be an effective strategy in the absence of questions provided by instructors or textbooks. However, it appears that students should not be left to their own for this strategy to be effective. Instead, instructors can augment this technique with brief guidance for their students on how to generate appropriate types of questions. Of course, the type of question is going to depend on what the class objective is. For classes like history that require knowledge for specific dates and events, then generation of detailed questions may be sufficient to enhance learning. However, for classes like biology that involve learning systems and processes, then generating conceptual questions would likely produce optimal learning.

Individual Differences

An individual difference that we have focused on in our laboratory is derived from Gernsbacher's (1990) Structure Building Framework. The major assertion of this framework is that readers have a working mental model of what they are reading and are constantly updating it whenever they come across new information. This new information could either be incorporated into the existing structure or if it is deemed irrelevant, depending on the reader, the information is inhibited or a new substructure is constructed. It is posited that poor structure builders have a difficult time (a) inhibiting irrelevant information and (b) limiting the number of substructures constructed. Thus, poor structure builders end up shifting and constructing a lot of unnecessary substructures. Good structure builders, on the other hand, do a better job at inhibiting irrelevant information and develop a more coherent, organized mental model of what the text is about. The Structure Building framework originated in considering text comprehension of narratives, and little research has examined whether it is a significant predictor of learning from technical passages.

To this end, Martin, Nguyen, and McDaniel (2013) examined whether differences in structure building abilities influenced the effectiveness of the meta-3R strategy. All participants engaged in the 3R study strategy and also took the Revised Multi-Media Comprehensive Battery (MMCB-R), which is a measure of structure building (Gernsbacher & Varner, 1988). Consistent with the theoretical assumptions of the structure building framework, high structure builders outperformed low structure builders on a wide range of assessments that included free recall, inference multiple-choice, and problem-solving tests. Further, another difference between high and low structure builders was identified: High structure builders were able to guide their restudy more effectively than low structure builders. According to the Discrepancy-Reduction (DR) self-regulated learning framework, learners should spend less time restudying material judged as well-learned than material judged as less well-learned (Thiede & Dunlosky, 1999). Martin et al. found that high structure builders were more likely to follow this self-regulated study policy than low structure builders. This difference in metacognitive strategy played a vital role in helping high structure builders recall significantly more new information (non-recited idea units) on a final free recall test than low structure builders.

More pertinent to the present chapter, structure building has also been found to moderate the effectiveness of study strategies designed to improve text comprehension. Callender and McDaniel (2007) found that placing embedded questions throughout the text improved test performance for low but not high structure builders relative to a rereading group (see Figure 4). The authors interpreted these results as indicating that embedded questions help provide low structure builders with anchors around which to build their mental model of the text's meaning. Because high structure builders are already adept at building an organized and coherent representation, answering embedded questions was superfluous for improving their comprehension.

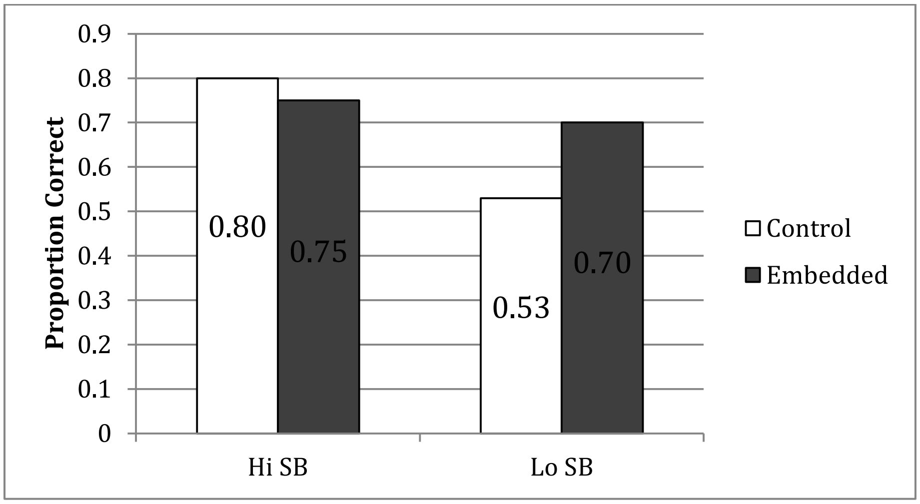


Figure 4. Final test performance as a function of condition and level of structure building ability. Adapted from Callender & McDaniel (2007).

Bui and McDaniel (2013) also found structure building to moderate the effectiveness of outlines on learning. They instructed participants to take notes while listening to an audio lecture. Some learners were provided with an outline of the lecture while others received no learning aid. Following a short delay, learning was assessed with both a free recall test and an inference-based short-answer test. Providing outlines helped all learners on free recall but only improved performance on inference tests for high structure builders. These results suggest that low structure builders do not have the cognitive capability to develop a coherent mental model of the lecture despite the assistance of an outline.

Taken together, these findings suggest that structure building is a particularly important individual difference that should be taken into account when assessing the efficacy of a technique aimed at improving text comprehension. Because of the vital role that structure building plays in moderating the effectiveness of some of these study strategies, we suggest that instructors become mindful of individual differences (especially structure building) when trying to implement them, as benefits may not accrue for all students.

Conclusion

In the present chapter, we have endorsed text learning techniques that are especially useful in classes in which active learning is not an inherent part of the course. For instance, in classes such as chemistry or physics, active problem-solving is an essential part of the class no matter how the course is structured. Accordingly, students may be more likely to actively interrogate the textbook materials in the service of solving problems. However, with classes like introductory psychology or history, it is perhaps natural that students are less able to find active approaches for learning from the textbook. Thus, the techniques introduced in this chapter are particularly important for these classes in which passive learning from textbooks may be the norm. Our hope is that these techniques will help promote more active learning from textbooks than most students would otherwise typically engage.

The techniques introduced in this chapter fall under a broader umbrella called desirable difficulties, the idea that learning perceived to be difficult and challenging in the moment tend to yield greater benefits later on (Bjork, 1994; McDaniel & Butler, 2011). While researchers and instructors may understand the value of these techniques, it is important to remember that students may perceive them to be less worth their while than reading (and rereading) through the text. Even though our data suggest that these techniques are not any more time-intensive than simply rereading or taking notes, it is essential to consider that students' levels of motivation may decrease in the beginning when they do not see immediate benefits. Consider the experience of Young, a medical student who was having a difficult time learning the material in his classes before he came across retrieval practice as a learning technique. Below is an excerpt taken from Brown, Roediger, and McDaniel (in press) that exemplifies the difficulty that Young experienced at first when trying to implement retrieval practice:

> *"It makes you uncomfortable at first. If you stop and rehearse what you're reading and quiz yourself on it, it just takes a lot longer. If you have a test coming up in a week and so much to cover, slowing down makes you pretty nervous…You just have to trust the process, that was really the biggest hurdle for me, was to get myself to trust it. And it ended up working out really well for me."* (pp. 213-214)

Indeed, implementing retrieval practice worked out well for Young. He went from the bottom of his class to the top 10-15%. This is a real-life example of a student who persevered and kept implementing retrieval practice despite the frustration he encountered when first implementing the strategy. However, the average student's knowledge of effective study strategies is already poor so continued practice of a technique that is perceived as being difficult is less likely to occur (McCabe, 2011). Thus, it is important for instructors to reinforce students' motivation to continue using the technique by demonstrating the benefits of the strategy. Einstein, Mullet, and Harrison (2012) had students study one text using a study-study strategy and another text using a study-test strategy. Not surprisingly, they found the standard testing effect with students performing better on the text in which they used a study-test strategy. More importantly, they had students analyze their results to see first-hand how they performed on the respective tests. Following this demonstration, students reported that they were more likely to engage in a testing strategy during future study. These results suggest that students can be motivated to use a particular technique if they experience the benefits first-hand.

Attempts to apply cognitive principles to improve educational practices have picked up steam lately. One area that would benefit is text comprehension, given the considerable amount of time students spend learning from textbooks. In the present chapter, we have discussed several techniques that show great promise in producing effective learning from texts. In addition, we have also attempted to tackle the important question of translating laboratory findings to the classroom. Although we have made some practical recommendations, our hope is that instructors will use these recommendations as a starting point and adapt them to fit the constraints of their classes. Because of the easy to learn and time-efficient nature of these techniques, we are hopeful that they would be well-received by students and thus implemented routinely while studying texts. Nevertheless, instructors still play a vital role in augmenting these techniques by constructing an environment that encourages students to apply these techniques.

Acknowledgment

This research was supported by a Collaborative Activity Grant from the James S. McDonnell Foundation (No. 22002004I). The first author was supported by the National Science Foundation Graduate Fellowship

Disclosure

The Nguyen and McDaniel (2013b) study used materials related to an e-textbook project with Worth Publishers on which McDaniel has been consulting.

References

Anderson, R.C., & Biddle, W.B. (1975). On asking people questions about what they read. In G. Bower (Ed.), *Psychology of learning and motivation*. New York, NY: Academic Press.

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A taxonomy of learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York, NY: Addison Wesley Longman, Inc.

Andre, T. (1979). Does answering higher level questions while reading facilitate productive learning? *Review of Educational Research, 49*, 280-319.

Armbruster, B.B. (2000). Taking notes from lectures. In R.F. Flippo & D.C. Caverly (Eds.). *Handbook of college reading and study strategy research* (pp. 175-199). Hillsdale, NJ: Erlbaum.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Brown, P., Roediger, H.L., III, & Mcdaniel, M.A. (in press). *Making it stick: The science of successful learning*. Cambridge, MA: Harvard University Press..

Bugg, J.M. & McDaniel, M.A. (2012). Selective benefits of question self-generation and answering for remembering expository text. *Journal of Educational Psychology, 104*, 922-931.

Bui, D.C., & McDaniel, M.A. (2013). Structure building and note-taking. Unpublished manuscript.

Callender, A.A., & McDaniel, M.A. (2007). The benefits of embedded question adjuncts for low and high structure builders. *Journal of Educational Psychology, 99*, 339-348.

Callender, A.A., & McDaniel, M.A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology*, 34, 30-41.

Davey, B., & McBride, S. (1986). Effects of question-generation training on reading comprehension. *Journal of Educational Psychology, 78*, 256-262.

Di Vesta, F.J., & Gray, G.S. (1972) Listening and note taking. *Journal of Educational Psychology, 63*, 8-14.

Einstein, G.O., Morris, J., & Smith, S. (1985). Note-taking, individual differences, and memory for lecture information. *Journal of Educational Psychology, 77*, 522-532.

Einstein, G.O., Mullet, H.G., & Harrison, T.L. (2012). The testing effect: Illustrating a fundamental concept and changing study strategies. *Teaching of Psychology, 39*, 190-193.

Gernsbacher, M.A., Varner, K.R., & Faust, M.E. (1990). Investigating differences in general comprehension skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 430-445.

Gernsbacher, M.A., & Varner, K.R. (1988). *The multi-media comprehension battery*. Eugene, OR: University of Oregon, Institute of Cognitive and Decision Sciences.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56*, 212-242.

Hamilton, R. (1985). A framework for the evaluation of the effectiveness of adjunct questions and objectives. *Review of Educational Research, 55*, 47-85.

Hartwig, M.K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review, 19*, 126-134.

Kalbaugh, G.L., & Walls, R.T. (1973). Retroactive and proactive interference in prose learning of biographical and science materials. *Journal of Educational Psychology, 65*, 244-251.

Karpicke, J.D., Butler, A.C., & Roediger, H.L., III. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory, 17*, 471-479.

Keppel, G., & Underwood, B.J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1*, 153-161.

Kiewra, K.A. (1989). A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review, 1*, 147-172.

Kiewra, K.A., Benton, S.L., Kim, S., Risch, N., Christensen, M. (1995). Effects of note-taking format and study technique on recall and relational performance. *Contemporary Educational Psychology, 20*, 172-187.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.

Kornell, N. & Bjork, R.A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*, 219-224.

Lahtinen, V., Lonka, K., & Lindblom-Ylanne, S. (1997). Spontaneous study strategies and the quality of knowledge construction. *British Journal of Educational Psychology, 67*, 13-24.

Martin, M.A. (1985). Students' applications of self-questioning study techniques: An investigation of their efficacy. *Reading Psychology, 6*, 69-83.

Martin, N., Nguyen, K., & McDaniel, M.A. (2013). The effects of structure building on text learning. Unpublished manuscript.

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition, 39*, 462-476.

McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A Benjamin (Ed.), *Successful Remembering and Successful Forgetting: A Festschrift in Honor of Robert A. Bjork* (pp. 175-199). Taylor & Francis.

McDaniel, M.A., Howard, D.C., & Einstein, G.O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science, 20*, 516-522.

McNamara, D.S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22*, 247-288.

McNamara, D.S., & McDaniel, M.A. (2004). Suppressing irrelevant information: Knowledge activation or inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 465-482.

Melton, R.F. (1978). Resolution of conflicting claims concerning the effects of behavioral objectives on student learning. *Review of Educational Research, 48*, 291-302.

Nguyen, K., & McDaniel, M.A. (2013a). 3R study strategy: Oral versus Written recitation. Unpublished manuscript.

Nguyen, K., & McDaniel, M.A. (2013b). The relative benefits of embedded questions versus self generated questions. Unpublished manuscript.

Nguyen, K., McDaniel, M.A., Martin, N. (2012, November). *Meta-3R: Unique contributions of recitation and metacomprehension judgments*. Poster presented at the annual meeting of the Psychonomic Society, Minneapolis, MN.

Novak, J.D. (2010) *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations*. New York, NY: Erlbaum.

Peper, R.J., & Mayer, R.E. (1978). Note taking as a generative activity. *Journal of Educational Psychology, 70*, 514-522.

Rawson, K.A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology, 97*, 70-80.

Robinson, F.P. (1941). *Effective Study*. New York, NY: Harper and Brothers. Roediger, H.L., III, & Karpicke, J.D. (2006b). Test-enhanced learning: Taking tests improves long-term retention. *Psychological Science, 17*, 249-255.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66*, 181-221.

Rothkopf, E.Z. (1966). Learning of written instructive material: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal, 3*, 241-249.

Sikorski, J. F., Rich, K., Saville, B. K., Buskist, W., Drogan, O., & Davis, S. F. (2002). Student use of introductory texts: Comparative survey findings from two universities. *Teaching of Psychology, 29,* 312-313.

Slotte, V., & Lonka, K. (1999). Review and process effects of spontaneous note-taking on text comprehension. *Contemporary Educational Psychology, 24*, 1-20.

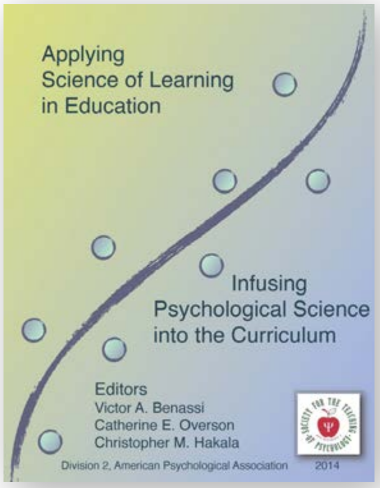
Stine-Morrow, E. A. L., Gagne, D. D., Morrow, D. G., & DeWall, B. H. (2004). Age differences in rereading. *Memory & Cognition, 32,* 696-710.

Szpunar, K.K., McDermott, K.B., & Roediger, H.L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1392-1399.

Thiede, K.W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1024-1037.

Weinstein, Y., McDermott, K.B., & Roediger, H.L., III. (2010). A comparison of study strategies for passages: Rereading, answering questions, and generating questions. *Journal of Experimental Psychology: Applied, 16*, 308-316.

Wong, B.Y.L. (1985). Self-questioning instructional research: A review. *Review of Educational Research, 55*, 227-268.



Learning From Worked Examples: How to Prepare Students for Meaningful Problem Solving

Alexander Renkl
University of Freiburg, Germany

If you teach mathematics, statistics, or science you probably know the following situation very well. You have introduced a theorem or physics law in your classroom. When the students then work on corresponding problems, many of them mainly try to find the correct numerical solution by applying the corresponding formulae; they do not further reflect on the rationale of the solution or on the underlying mathematical or physical structures. In this way, many students become rather proficient in applying formulae. This is at least true as long as it is clear from the classroom context that the problems to be solved are related to the relevant formulae and as long as the formulae can be directly applied without modifications. However, when the students encounter related problems in an unfamiliar context or embedded in a new type of cover story, many of them are not able to identify the underlying mathematical or physical structures and to apply the correct theorem or physics law. Many students also fail when they have to slightly modify a known solution procedure. Such phenomena show that mechanically applying formulae that were most recently introduced in the classroom does not deepen understanding and does not enable the students to apply their knowledge in new contexts or for new, but related problem types.

One reason why many students mechanically apply formulae when solving problems is that they have not yet developed a sufficient understanding of a theorem or of a physics law before solving problems. Hence, they simply are not able to solve the problems in a meaningful way, that is, by considering mathematical or physical structures and corresponding theorems or laws. One tried-and-tested method to address this problem is to use learning from worked examples for the initial acquisition of cognitive skills (Sweller, Ayres & Kalyuga, 2011; Paas & van Gog, 2006; Renkl, in press-b). In this method, the learners study several worked examples after some principles (e.g., theorems or laws) have been introduced. The students are best guided to explain the rationale of the worked solutions to themselves. Only after the learners have developed the understanding that is necessary for meaningful problem solving, they work on problems.

This short characterization of learning from worked examples illustrates the basic idea of this learning method. When I introduce this rationale to teachers I frequently hear three objections. Although these objections all have a "true core," none of them are sufficient for an instructor to refrain from using this technique.

Objection 1: It's just for maths and related stuff. Actually, much on the initial research on learning from worked examples was done with mathematics and science contents. However, it has been shown that learning from worked examples can be applied in many different domains, for example, learning to argue (Schworm & Renkl, 2007); learning about designer styles (Rourke & Sweller, 2009); learning to apply learning strategies (Hilbert & Renkl, 2009).

Objection 2: Worked examples are typical of old-fashioned pedagogy focusing on algorithms and "recipes". Yes, worked examples are often used in a way that does not deepen understanding, that is, mainly as a tool to teach algorithms or simple recipes. In many of these cases, just one worked example is presented. Note, however, that learning from worked examples, as conceptualized by their proponents, means that a series of examples are presented with the main goal to deepen understanding before problem solving (Renkl, in press-b).

Objection 3: Students usually read worked examples just superficially without striving for real understanding. This is true (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Renkl, 1997). For this reason, a central guideline when implementing learning from worked examples is to support the students in explaining the rationale of the worked solutions to themselves (e.g., Renkl, in press-b).

How Do Worked Examples Look?

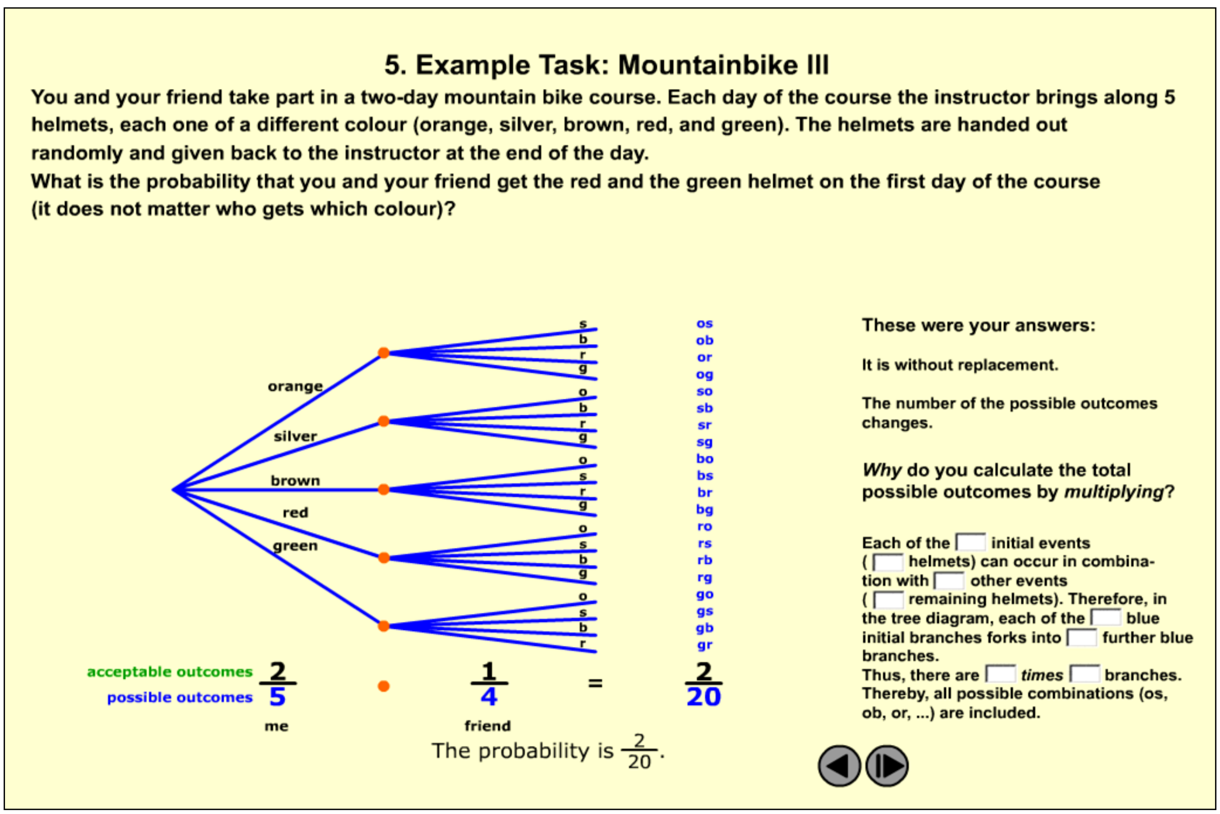


Figure 1. A worked example from probability theory (from Berthold & Renkl, 2009; Copyright © 2009 by American Psychological Association).

Worked examples consist of a problem formulation - which may be a mathematics word problem, a request to find a mathematical proof, or a request for proper argumentation - and the (final) problem solution. Very often such an example additionally provides a series of steps that lead to the final solution (see Figure 1). Mathematics and science examples with algorithmic solution procedures are only some of the many possible varieties of solution steps. For example, Hilbert, Renkl, Kessler, and Reiss (2008) provided the heuristic steps that students should apply when trying to find a mathematical proof (note that such worked examples may comprise several pages); Rummel, Spada, and Hauser (2009) have employed worked examples in the form of videos displaying different steps in productive cooperation; Nievelstein, van Gog, van Dijck, and Boshuizen (2013) successfully employed worked examples detailing steps to analyze legal cases. As Figure 2 shows, not all worked examples provide solution steps. Hilbert, Renkl, Schworm, Kessler, and Reiss (2008) taught (future) teachers how to design worked examples for their mathematics lessons by providing a well-designed and a sub-optimal worked example. The variety of possible applications of worked examples shows that this learning method is not only feasible when students need to learn algorithmic solutions, but also when heuristic methods (that do not necessarily lead to an optimal solution) should be learned (e.g., Renkl, in press-b)

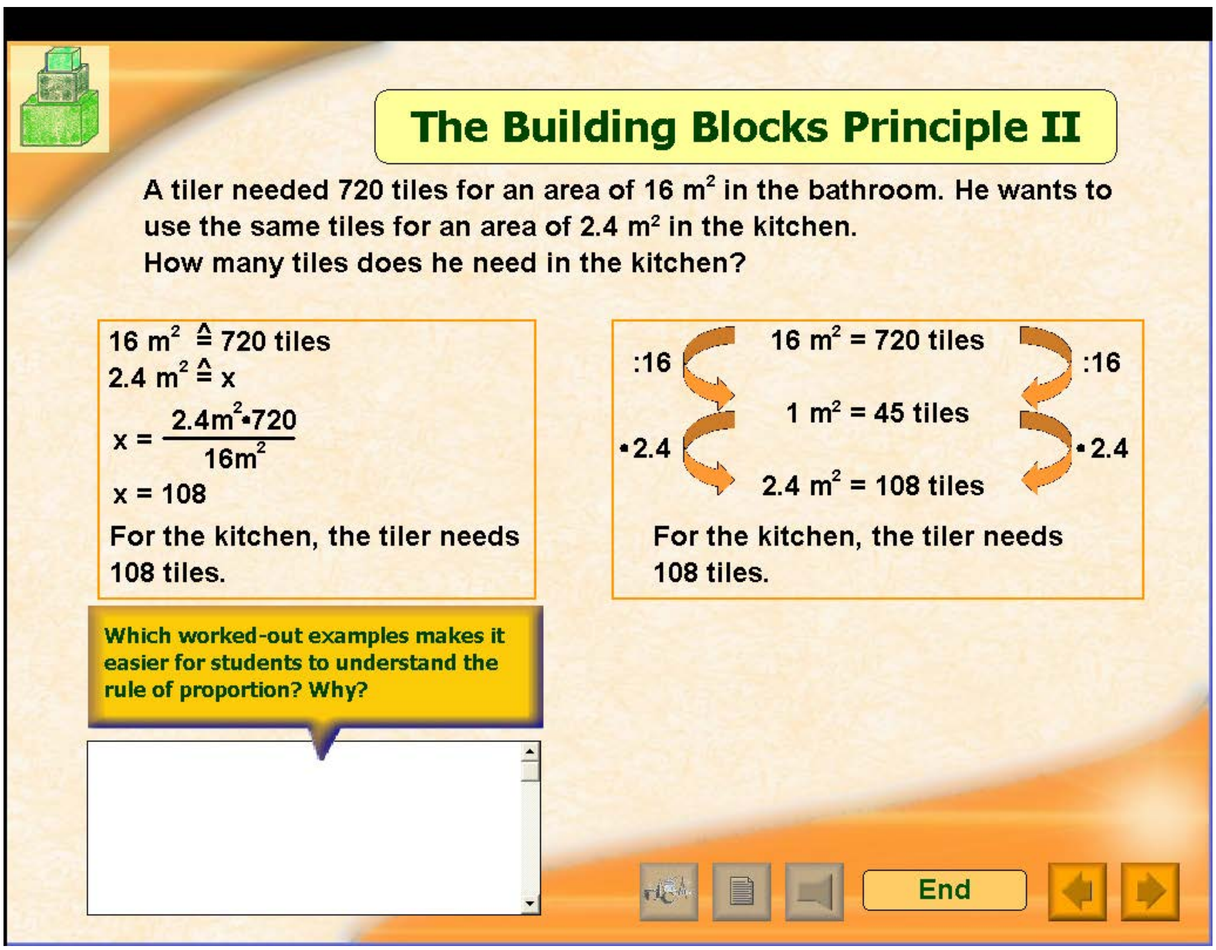


Figure 2. A worked example in a "double sense": Screenshot from an example-based learning environment for teachers on how to design effective worked examples for high-school students (from Hilbert, Renkl, Schworm et al., 2008; Copyright © 2008 by John Wiley & Sons, Inc.).

As argued in the beginning of this chapter, students are best prompted to self-explain the solution (steps) (see Chiu & Chi, this volume). The questions on the right side in Figure 1 and in the lower-left corner of Figure 2 are self-explanation prompts. In the learning environment, from which Figure 1 is taken, the students are guided to understanding not only when the multiplication rule in probability should be applied but also why one has to multiply. As this issue is difficult to understand, self-explaining is initially supported by a partially worked instructional explanation to be completed. In later worked examples this help is faded out and the students have to provide the complete explanation by themselves.

Another important characteristic of worked examples, illustrated by the Figures 1 and 2, is that they often include special features to deepen students' understanding. As you can see in Figure 1, the solution is provided in two ways and types of representations: tree diagrams and equations. If the students are encouraged to use the tree diagrams as a "bridge" in order to see how the problem formulations translate into the equations, their understanding and transfer performance is greatly enhanced (Schwonke, Berthold, & Renkl, 2009). You might have also noted that color coding (corresponding elements are displayed in the same color) makes it easier for students to figure out which elements in the tree diagrams correspond to which elements in the equations and vice versa. In addition, the self-explanation prompts (right side in Figure 1) encourage interrelating the different solution procedures. In Figure 2, two contrasting examples are provided to make an important point stand out, clearly marking the single solution steps and their "sense".

Effectiveness of Worked Examples

How Effective are Worked Examples?

Hattie (2009) reported in his synthesis of research on learning and teaching that conventional worked examples, when compared to learning by problem solving, have medium to strong effects (see also the classic studies by Sweller & Cooper, 1985). Note that conventional worked examples do not include prompts or other means to foster students' self-explanation. When self-explanations are fostered the positive effect roughly doubles to a strong effect (Hattie, 2009; Renkl, in press-a). If other favorable features – some of them have already been discussed in reference to the Figures 1 and 2 – are added, the effectiveness can be further advanced (a variety of other favorable features are discussed as instructional principles later in this chapter). Learning from worked examples, if implemented well, is even superior to very well-supported learning by problem solving (Salden, Koedinger, Renkl, Aleven, & McLaren, 2010). In many cases, worked examples not only foster learning outcomes, including transfer performance, but also reduce the necessary learning time (Salden et al.). Zhu and Simon (1987) found that a three-year mathematics curriculum could be taught in two years by learning from worked examples without any performance losses. Against the background of these convincing results, Pashler et al. (2007), in their evidence-based practice guide, count worked examples as one of seven main recommendations for teaching and learning.

The positive effects of worked examples can be fully canceled out if important design principles are neglected. For example, if you used worked examples such as in Figure 1 but without color-coding and without self-explanation prompts and support, learning would very likely not be effective; the learners can be overwhelmed with how to use and integrate the different representations (Berthold & Renkl, 2009; Tarmizi & Sweller, 1988). In a nutshell, learning from worked examples is highly effective but only when implemented properly. Instructional principles that guide the implementation of this learning method will be extensively discussed in a later section.

Why are Worked Examples Effective?

As already argued, students typically lack a profound understanding of the domain principles and their application in the beginning of skill cognitive acquisition (Renkl, in press-b). Hence, they are not able to rely on domain strategies for problem solving that are based on the principles to-be-learned (VanLehn et al., 2005). Their "back-ups" are shallow strategies such as a key word strategy (i.e., selecting a procedure by a key word in the cover story of a problem), a copy-and-adapt strategy (i.e., copying the solution from a presumably similar problem and adapting the numbers), or a means-ends analysis focusing on superficial problem features (Renkl, in press-b). Not only do these shallow strategies fail to deepen domain understanding, they bind attentional capacities for activities that are extraneous to gaining understanding (i.e., extraneous cognitive load). Worked examples free learners from such extraneous activities so that there are cognitive resources for self-explanation available. By self-explaining, the students become aware of a solution's rationale, ideally in reference to the underlying domain principles. After the learners have understood the domain principles and their application, they should be encouraged to solve problems requiring the application of these principles. When taking the reasons for the effectiveness of worked examples into account, it is becomes clear that this learning method is only effective for the initial acquisition of cognitive skills (for a more thorough discussion see Renkl, in press-b)

Which Instructional Principles are Important for Effective Learning From Worked Examples?

The effects of learning from worked examples heavily depend on how well this learning method is implemented. Following are ten instructional principles that should be taken into account when implementing this learning method.

Self-Explanation Principle

The importance of self-explanations has already been stressed in this chapter; learners' self-explanations are crucial for fully exploiting the potential of learning from worked examples ("self-explanation effect"; Chi et al., 1989). Two main types of productive self-explanations can be differentiated: (a) elaborating on individual examples in order to foster their understanding and (b) comparing examples, which typically helps to form or differentiate abstract problem categories (Gerjets, Scheiter, & Schuh, 2008; Nokes-Malach, VanLehn, Belenky, Lichtenstein, & Cox, 2013).

Elaborating on individual examples: The most important self-explanation type when elaborating on individual examples are principle-based explanations. They relate solutions to abstract principles, such as mathematics theorems, physics laws, and guidelines for proper argumentation (e.g., Chi et al., 1989; Renkl, 1997; Schworm & Renkl, 2007). For example, Atkinson, Renkl, and Merrill (2003) encouraged students to justify worked solution steps in terms of the underlying probability principle. The learners could select a probability principle from a menu of potentially relevant principles. Such self-explaining helps students solve not only similar problems but also novel (transfer) problems later on. Schworm and Renkl (2007) had their learners identify argumentative elements of proper scientific argumentation when they watched video-based worked examples, which enabled them to argue about another topic in a differentiated way.

Comparing examples: Figure 2 illustrates self-explaining when comparing examples. The (future) teachers might refer in their self-explanations to the *meaningful building block principle* (i.e., the solution steps should be labeled as meaningful building blocks of a procedure) when they explain why the right example is superior to the left example for high-school students (Schworm & Renkl, 2006). Rittle-Johnson, Star, and Durkin (2009) guided their learners to explain the difference between the two possible worked solution methods and the conditions that must be met so that the more parsimonious method can be applied. Flexible problem solving was taught by this comparison procedure.

The two main possibilities of fostering self-explanations are training and prompting. Many tried-and-tested prompting procedures request that the learners type their self-explanations into text boxes. Sometimes such self-explanations are supported by menus providing a list of potential principles (e.g., Conati & VanLehn, 2000). Self-explanations can also be trained (e.g., Renkl, Solymosi, Erdmann, & Aleven, 2013). Renkl, Stark, Gruber, and Mandl (1998) found the following elements of training useful: (a) information on the importance of self-explanations, (b) modeling self-explanations, that is, a teacher demonstrates this activity by articulating productive self-explanations, and (c) coached practice.

A *qualification* of this principle is that self-explanation demands can overwhelm the learners when working memory load is already high due to complex learning tasks in relation to learners' prior knowledge and/or to suboptimal instructional design (e.g., Kalyuga, 2010; Sweller, 2006). In addition, poor prior knowledge can be a barrier to provide correct self-explanations (see Ambrose & Lovett, this volume).

Explanation-Help Principle

As just argued, learners are sometimes unable to self-explain correctly and productively. Help in the form of instructional explanations should then be provided. In addition, as instructional explanations are often just superficially processed (Wittwer & Renkl, 2008), prompts to further process the instructional explanations enhance their effects (Berthold & Renkl, 2010). A *qualification* is that too easily available instructional explanations (e.g., the learners can demand them without having first tried to provide substantial self-explanations) can reduce learners' self-explanation efforts and learning outcomes (Schworm & Renkl, 2006; see also Clark & Bjork on desirable difficulties, this volume). For this reason, learners working on the screen displayed in Figure 2 could just demand an explanation when they have first typed in some self-explanations into the box.

Example-Set Principle

This principle goes beyond using multiple examples which is already inherent in the present learning method. Sets of examples are typically used to direct the learners' attention to crucial aspects (e.g., Bransford & Schwartz, 1999). The specific goals of example sets can be quite varied. We have already mentioned the case of providing different solution procedures in order to teach the conditions for applying the simpler solution method and, thus, to foster flexible problem solving (e.g., Rittle-Johnson et al., 2009). "Structure-emphasizing example sets” (Quilici & Mayer, 1996) combine examples in a way that (a) each problem category (e.g., probability problems with and without order relevant) is exemplified by a set of different cover stories (i.e., surface) and (b), the same set of cover stories is used across the problem categories. Learners observe that cover stories and solution-relevant mathematical structures do not necessarily co-vary, and that relying on surface features can mislead when trying to find the correct solution. A *qualification* is that example-set effects are not reliable if the learners are not explicitly prompted to compare the examples (Scheiter, Gerjets, & Schuh, 2003). Only prompted or guided example comparison assures the desired learning outcomes (e.g., Gentner, Loewenstein, & Thompson, 2003; Richland & McDonough, 2010).

Easy-Mapping Principle

Worked examples lose their effectiveness when learners have difficulty relating different information sources to each other (e.g., text, picture, or equations; see also Figure 1; e.g., Tarmizi & Sweller, 1988). Visual search processes may require so much cognitive capacity that self-explanations are blocked. Facilitating the mapping between information sources frees cognitive resources for self-explaining, and worked examples are again effective. In Figure 1, color coding was used to facilitate mapping between tree diagram and equation (Berthold & Renkl, 2009). Another possibility is to rely on the modality effect, which means that spoken (instead of written) text is combined with a picture (e.g., Ginns, 2005, see also Mayer, this volume). In addition, a cue can signal to which pictorial element the current narrated text is referring (for more options on how to facilitate mapping see, Renkl, in press-b). A *qualification* is that it is open when to best use which procedure to facilitate mapping (e.g., color coding or modality arrangement with signaling).

Meaningful-Building Blocks Principle

Frequently, students learn a solution procedure just as a "fixed chain" of steps (Catrambone, 1998). When students work on transfer problems for which a modified solution procedure applies, they can fail because the chain cannot be broken into meaningful building blocks that can be flexibly reassembled. To enable students for such transfer, individual steps should be presented as meaningful building blocks of a solution procedure (for an example see Figure 2, right side). Hence, examples should be designed in way that the sub-goals that are achieved by the individual solution steps are easily identifiable. Such salience of single steps can be attained by visually isolating the steps (e.g., by circles) and by assigning labels (see also Spanjers, van Gog, & van Merriënboer, 2012). Salient subgoals lead to self-explanations about what these steps accomplished. Another way to make (sub-) goals salient is to use a step-by-step presentation of worked solutions (Atkinson & Derry, 2000; Schmidt-Weigand, Hänze, & Wodzinsksi, 2009; Renkl, in press-b, for a more thorough discussion). A *qualification* of the meaningful-building blocks principle is that up to now the really convincing evidence has originated just from studies on mathematics.

Studying Errors Principle

Self-explaining correct and incorrect worked solutions is usually more beneficial than self-explaining correct solutions only (Siegler & Chen, 2008); explaining incorrect solutions helps to avoid these errors in later problem solving (Durkin & Rittle-Johnson, 2012). Note that in Figure 2, the (future) teachers compared well-designed and suboptimal versions of a worked example for high-school students. When worked examples are presented as video models, learners usually profit more from coping models that initially also commit errors and show how they can be overcome, as compared to mastery models showing just smooth performance (Kitsantas, Zimmerman, & Cleary, 2000; Zimmerman & Kitsantas, 2002). A *qualification* is that only learners with sufficient prior knowledge might profit from studying errors in examples (Große & Renkl, 2007). Providing errors too early in the learning process overwhelms learners. Weaker learners need additional support by explicitly marking errors (Große & Renkl) or by expert explanations as to why certain moves were correct or not (Stark, Kopp, & Fischer, 2011).

Model-Observer Similarity Principle

Recent research has increasingly employed worked examples in the form of videos displaying model persons that show the respective solutions (e.g., Rummel et al., 2009; Schworm & Renkl, 2007). The model-observer similarity is one of the classic factors influencing learning, as determined by observational learning research (e.g., Bandura, 1986; Schunk & Zimmerman, 2007). When models are too dissimilar, observers do not identify and, thus, do not imitate. Furthermore, when models are dissimilar in the sense that they are too advanced, observers do not believe that they can show the appropriate behavior on their own, that is, they lack self-efficacy. Braaksma, Rijlaarsdam, and van den Bergh (2002) displayed both competent and non-competent models in their worked examples to learn argumentative writing. In one condition, the participants were instructed to focus on the competent model and, in another condition, on the non-competent model. Weak students profited more from a focus on the non-competent model, stronger students learned best when focusing on the competent model. This pattern can be interpreted as similarity effect. A *qualification* of this principle is that it is not fully clear which similarity features are most crucial.

Focus-On-Learning Domain Principle

Worked examples for complex learning contents such as scientific argumentation (Schworm &Renkl, 2007) or mathematical proof (Hilbert et al., 2008) run the risk of overwhelming learners' cognitive capacities. For example, Schworm and Renkl (2007) used stem-cell research as content domain to exemplify proper scientific argumentation; Hilbert et al. (2008) used some geometry contents to teach heuristic strategies in proof finding. If the students have difficulty understanding these exemplifying domain contents (i.e., stem-cell research or geometry) or if their attention is sub-optimally directed by instructional design deficits (Schworm & Renkl, 2007), the learners' attention is largely bound to the exemplifying domain. Hence, achieving the central learning goals is hindered (Renkl, Hilbert, & Schworm, 2009). Instructional design should use easy-to-process exemplifying domains and processing prompts that focus on the learning domain. A *qualification* is that there might also be learning situations in which deeply processing and remembering superficial features from the exemplifying domain may facilitate transfer to related problems. For example, when argumentation skills are exemplified in the area of one medical treatment problem (e.g., surgery or conservative treatment) the transfer of these argumentation skills to other medical topics can be facilitated by overlapping exemplifying domain contents.

Imagery Principle

In the imagery principle, students first read a worked solution, then turn away from the screen or work sheet and imagine performing the solution procedure (e.g., Cooper, Tindall-Ford, Chandler, & Sweller, 2001; Ginns, Chandler, & Sweller, 2003). This effect is consistent with the broader research on learning by mental imagery (e.g., Hall, Munroe-Chandler, Cumming, Law, Ramsey, & Murphy, 2009) whereby mental imagery can have effects similar to actually performing the task. A *qualification* is that imagining is not effective when working memory load is very high (e.g., due to two information sources that must be integrated and mentally manipulated; Tindall-Ford & Sweller, 2006). In addition, learners with low prior knowledge simply cannot imagine the solution when looking away from the example. Hence, it is a sensible procedure to first provide an example for study and second an example for imagery (Ginns et al., 2003).

Fading Principle

Worked examples are effective in initial stages of cognitive skill acquisition. In later stages, when automation is one of the main goals, problem solving is superior (Renkl, in press-b). A tried-and-tested method for structuring the transition from studying worked steps to problem solving is to gradually fade worked steps (e.g., Renkl & Atkinson, 2003; Atkinson et al., 2003; Kissane, Kalyuga, Chandler, & Sweller, 2008). In such a procedure, a complete example is presented first. Second, an isomorphic example is presented in which a single step has been omitted. After trying to supplement the faded step, the learner receives feedback about the correct solution. Then, in the following examples, the number of blanks increases until only the problem formulation is left, that is, a problem to be solved. Such a fading procedure has proven to be effective. Ideally, the fading procedure is adapted to the individual learner's progress. Salden, Aleven, Renkl, and Schwonke (2009) faded a specific worked step when the learner provided correct self-explanations on a preceding isomorphic step, thereby indicating understanding of the respective knowledge component. A *qualification* is that the fading principle has just been well established with worked examples providing mathematical solution procedures. It is open how to fade effectively in less well-structured skills domains.

Wrapping up the Recommendations

Learning from worked examples is a very well established method when students should acquire cognitive skills. This method supports students in the beginning of skill acquisition when the major goal is to achieve an understanding of the underlying domain principles and their application. Once students understand the domain principles, worked examples can help them solve problems in a meaningful way and to fine-tune and optimize their skills.

As worked examples lose their effectiveness in the course of cognitive skill acquisition, the worked steps should be gradually faded out (*fading principle*). Another strategy to help students move from studying worked steps to problem-solving is to instruct the learners to first read and then imagine a solution (*imagery principle*). Because many students do not deeply process worked examples, self-explanations should be fostered (*self-explanation principle*). In this context, it is important that students are supported to integrate different information sources, which are often provided in worked examples, to lower working memory demands; otherwise self-explaining is blocked by cognitive overload (*easy-mapping principle*). When students, particularly those with low prior knowledge, have difficulties in providing good self-explanations, instructional explanations can be used as a back-up help (*explanation-help principles*).

When later applying the skills learned from worked examples, typical errors can be reduced by providing examples illustrating these "traps" (*studying errors principle*). Sometimes sets of examples to be compared can also help to avoid errors in later problem solving, in particular when the example sets are designed in ways that the crucial point to be attended stands out (*example-set principle*). Many students tend to learn worked solution procedures as fixed chains of steps that can only be applied to the very same problem type in the future. To prevent this lack of flexibility and to foster transfer, the single solution steps should be marked as meaningful building blocks (*meaningful-building blocks principle*).

When worked examples have the form of models, it is recommended to have a model person similar to the learners (*model-observer similarity principle*). Such models often use complex exemplifying domains to demonstrate the skill (e.g., stem-cell research for argumentation skills). In these cases, learners' attention should be primarily directed - for example, by corresponding prompts - on the learning domain (e.g., argumentation) (*focus-on-learning domain principle*).

Some of these principles for learning from worked examples "add up" in their effects. For example, self-explanation prompts and fading independently contribute to better learning (e.g., Atkinson et al., 2003). However, there are not only the-more-the-better relations between the instructional principles. Some principles depend inherently on each other (e.g., example sets and self-explanations with respect to example comparisons): Learners cannot compare a good and a poor writing model without providing two such models (Braaksma et al., 2002). In other cases, neglecting a principle such as "easy mapping" might countermand the effect of self-explanation prompts, because learners are cognitively overloaded (e.g., Sweller, 2006).

When implementing worked examples in their own classrooms, teachers should adopt a careful and theoretically informed (hopefully by this chapter) view as to whether this method might work in a specific context. Such a caveat is necessary, although the effectiveness of well-designed learning from worked examples is empirically very well established. Note that also a doctor when prescribing some tried-and-tested medication has to carefully check whether the expected effects emerge or whether there are small effects or even severe negative side effects in a specific case (see Renkl, in 2013). I hope, dear reader, that this chapter has not only motivated you to use learning from worked examples but also provided you with enough basic knowledge to check its effects and to further optimize it for your specific context. Good luck if you try out learning from worked examples!

References

Atkinson, R. K., & Derry, S. J. (2000). Computer-based examples designed to encourage optimal example processing: A study examining the impact of sequentially presented, subgoal-oriented worked examples. In B. Fishman & S. F. O'Connor-Divelbiss (Eds.), *Proceedings of the 4th International Conference of Learning Sciences* (pp. 132-133). Hillsdale, NJ: Erlbaum.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Combining fading with prompting fosters learning. *Journal of Educational Psychology, 95,* 774-783.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive.* Englewood Cliffs, NJ: Prentice Hall.

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology, 101*, 70-87.

Berthold, K., & Renkl, A. (2010). How to foster active processing of explanations in instructional communication. *Educational Psychology Review, 22*, 25-40.

Braaksma, M. A. H., Rijlaarsdam, G., & van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology, 94*, 405-415.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*, 61-100.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127,* 355-376.

Chi, M. T., H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145-182.

Conati, C., & VanLehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education*, *11*, 398-415.

Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagining procedures and concepts. *Journal of Experimental Psychology: Applied*, *7,* 68-82.

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using in correct examples to support learning about decimal magnitude. *Learning & Instruction, 22*, 206-214.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*, 393-408.

Gerjets, P., Scheiter, K., & Schuh, J. (2008). Information comparisons in example-based hypermedia environments: Supporting learners with processing prompts and an interactive comparison tool. *Educational Technology Research and Development, 56,* 73-92.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning & Instruction, 15*, 313-332.

Ginns, P., Chandler, P., & Sweller, J. (2003). When imagining information is effective. *Contemporary Educational Psychology, 28,* 229-251.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning & Instruction, 17*, 612-634.

Hall, C. R., Munroe-Chandler, K. J., Cumming, J., Law, B., Ramsey, R., & Murphy, L. (2009). Imagery and observational learning use and their relationship to sport confidence**.** *Journal of Sports Sciences, 27*, 327-337

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New, York, NY: Routledge.

Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Computers in Human Behavior, 25,* 267-274.

Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning & Instruction, 18,* 54-65.

Hilbert, T. S., Renkl, A., Schworm, S., Kessler, S., & Reiss, K. (2008). Learning to teach with worked-out examples: A computer-based learning environment for teachers. *Journal of Computer-Assisted Learning, 24,* 316–332.

Kalyuga, S, (2010). Schema acquisition and sources of cognitive load. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 48-64). New York, NY: Cambridge University Press.

Kissane, M., Kalyuga, S., Chandler, P., & Sweller, J. (2008). The consequences of fading instructional guidance on delayed performance: The case of financial services training. *Educational Psychology, 28*, 809–822.

Kitsantas, A., Zimmerman, B. J., & Cleary, T. (2000). The role of observation and emulation in the development of athletic self-regulation. *Journal of Educational Psychology, 92*, 811-817.

Nievelstein, F., van Gog, T., van Dijck, G., & Boshuizen, H. P. A. (2013). The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemporary Educational Psychology, 38*, 118–125.

Nokes-Malach, T. J., VanLehn, K., Belenky, D. M., Lichtenstein, M., & Cox, G. (2013). Coordinating principles and examples through analogy and self-explanation. *European Journal of Psychology of Education, 28*, 1237-1263.

Paas, F., & Van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning & Instruction, 16*, 87-91

Pashler, H., Bain, P. M., Botge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning. IES practice guide.* Washington, DC: US Department of Education.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88,* 144-161.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*, 1-29.

Renkl, A. (in press-a). The worked-out-examples principle in multimedia learning. In R. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (2nd revised ed.). Cambridge, UK: Cambridge University Press.

Renkl, A. (in press-b). Towards an instructionally-oriented theory of example-based learning. *Cognitive Science.*

Renkl (2013). Why practice recommendations are important in use-inspired basic research and why too much caution is dysfunctional. *Educational Psychology Review, 25,* 317-324.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skills acquisition: A cognitive load perspective. *Educational Psychologist, 38,* 15-22.

Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-based learning in heuristic domains: A cognitive load theory account. *Educational Psychology Review, 21*, 67-78.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology, 23*, 90-108.

Renkl, A., Solymosi, J., Erdmann, M., & Aleven, V. (2013). Training principle-based self-explanations: Transfer to new learning contents. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp.1205-1210). Austin, TX: Cognitive Science Society.

Richland, L. E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology, 35*, 28-43.

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology, 101*, 836-852.

Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognize designers' styles. *Learning & Instruction, 19*, 185-199.

Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning, 4*, 69–92.

Salden, R., Aleven, V., Renkl, A., & Schwonke, R. (2009). Worked examples and tutored problem solving: Redundant or synergistic forms of support? *Topics in Cognitive Science, 1*, 203-213.

Salden, R., Koedinger, K. R., Renkl, A., Aleven, V., & McLaren, B. M. (2010). Accounting for beneficial effects of worked examples in tutored problem solving. *Educational Psychology Review, 22*, 379-392.

Scheiter, K., Gerjets, P., & Schuh, J. (2003). Are multiple examples necessary for schema induction? In F. Schmalhofer, R. Young, & G. Katz (Eds.), *Proceedings of EuroCogSci 03. The European Cognitive Science Conference 2003* (pp. 283-288). Mahwah, NJ: Erlbaum.

Schmidt-Weigand, F., Hänze, M., & Wodzinski, R. (2009). Complex problem solving and worked examples: The role of prompting strategic behavior and fading-in solution steps. *German Journal of Educational Psychology, 23*, 129-138.

Schwonke, R., Berthold, K., & Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology, 23*, 1227-1243.

Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education, 46*, 426-445.

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology, 99*, 285-296.

Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading & Writing Quarterly, 23*, 7–25.

Siegler, R. S., & Chen, Z. (2008). Differentiation and integration: Guiding principles for analyzing cognitive change. *Developmental Science, 11*, 433-448.

Spanjers, I. A. E., van Gog, T., & van Merriënboer, J. J. G. (2012). Segmentation of worked examples: Effects on cognitive load and learning. *Applied Cognitive Psychology, 26*, 352–358.

Stark, R., Kopp, V., & Fischer, M. R. (2011). Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. *Learning & Instruction, 21,* 22-33.

Sweller, J. (2006). The worked example effect and human cognition. *Learning & Instruction, 16,* 165–169.

Sweller, J., Ayres, P. & Kalyuga, S. (2011). *Cognitive load theory.* New York: Springer.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition & Instruction, 2,* 59-89.

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80,* 424-436.

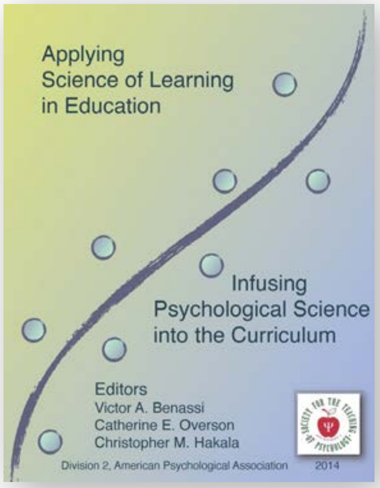
Tindall-Ford, S., & Sweller, J. (2006). Altering the modality of instructions to facilitate imagination: Interactions between the modality and imagination effects. *Instructional Science, 34,* 343–365.

VanLehn, K., Lynch, C., Schulze, K. Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., & Wintersgill, M. (2005). The Andes physics tutoring system: Lessons Learned. *International Journal of Artificial Intelligence and Education*, 15, 1-47.

Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist, 43*, 49-64.

Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition & Instruction, 4*, 137-166.

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94,* 660–668.



Spacing and Interleaving of Study and Practice

Shana K. Carpenter
Iowa State University

Overview and Scope of the Chapter

Learning requires repetition. Whether conjugating verbs in a foreign language, applying a mathematical formula, or learning to pitch a softball, a particular concept or skill must be practiced multiple times before it is fully mastered. It seems rather intuitive that learning benefits from repetition, however, what is less intuitive is how these repetitions should be scheduled. When learning a new concept, should students study that concept again immediately? Or should they wait a while and study it again later? A related question concerns how to schedule repetitions of similar concepts within a subject. When students learn about fractions, for example, should they practice one type of fractions problem over and over (e.g., how to add fractions) before moving on to a different type of problem (e.g., how to multiply fractions)? Or, would they learn the information better by practicing both types of problems together?

Students and instructors are faced with these decisions on a daily basis. Research on human cognition has shown that learning can be significantly affected by the way in which repetitions are scheduled. This research can help inform the decisions that students must make concerning when to study information in order to maximize learning. This chapter describes relevant research and offers pedagogical recommendations regarding two specific learning principles—the spacing effect and the interleaving effect.

The Spacing Effect

A number of studies have manipulated the timing of repeated exposures to information and then measured the resulting effects on memory retention. These studies find that learning is better when two or more exposures to information are separated in time (i.e., spaced apart) than when the same number of exposures occurs back-to-back in immediate succession. For example, Dempster (1987) found that students retained a greater number of vocabulary definitions when a given term and definition were repeated approximately every 5 minutes, rather than when the same term and definition were repeated consecutively. The learning advantage for information that is repeated in a "spaced" fashion is commonly referred to as the ***spacing effect***.

The design of a typical study on the spacing effect is illustrated in Figure 1. This design includes: (1) At least two study sessions that involve exposure to the same information, (2) A final test over the information, (3) a period of time, referred to here as the ***spacing gap***, that separates the two study sessions, and (4) Another period of time, referred to here as the ***test delay***, that separates the last study session from the final test. In the most basic experimental designs, the test delay is usually fixed (e.g., 20 minutes), and the spacing gap is manipulated. When the spacing gap is set at zero such that two exposures of the same information occur immediately (i.e., the same vocabulary term is repeated back-to-back), the exposures are said to be ***massed***. When the spacing gap is greater than zero such that the two exposures are separated by some amount of time, the exposures are said to be ***spaced***. The duration of spaced exposures varies across studies and has ranged anywhere from a few seconds (e.g., Carpenter & DeLosh, 2005) to several weeks (Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008). Regardless of the exact value of the spacing gap, spaced repetitions typically yield better learning than massed repetitions.

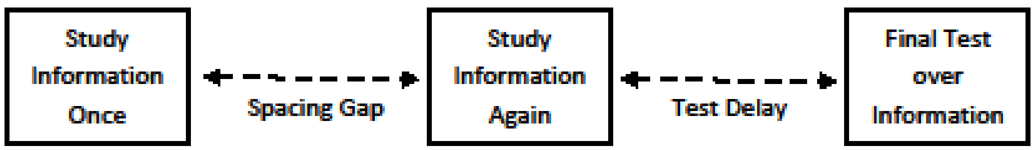


Figure 1. Illustration of a typical experiment on the spacing effect. Students are exposed to information at least twice, with each exposure separated by a spacing gap that can range anywhere from zero (i.e., the same information is repeated back-to-back), all the way to several weeks. Retention on the final test is typically better following a spacing gap greater than zero (i.e., spaced) compared to a spacing gap of zero (i.e., massed).

Sometimes the spacing gap is manipulated across individual items, such that a given term or concept is repeated in massed or spaced fashion (as in the study by Dempster, 1987). Other times, the spacing gap is manipulated across study sessions. For example, students may use flashcards to study each of 30 vocabulary definitions once. After going through the list, students may go back through the list immediately (massed), or they may wait for a period of time (e.g., 30 minutes, one day, etc.) before going back through the list (spaced). This way, there is no immediate repetition of a given item. Rather, the second exposure to a given item will occur after a relatively brief time interval (when the list repetitions are massed) vs. a relatively longer time interval (when the list repetitions are spaced). Under these conditions, spaced study sessions result in better learning than massed study sessions (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Somewhat related to this, some studies have found that longer spacing gaps produce better learning than shorter spacing gaps (e.g., Kahana & Howard, 2005). This finding is sometimes referred to as the *lag effect*. In this chapter, the term ***spacing effect*** is used to refer to both the advantage of spaced vs. massed exposures, and the advantage of longer vs. shorter spacing gaps.

Research on the Spacing Effect

The spacing effect has been documented in hundreds of published studies over the last century (for review articles, see Cepeda et al., 2006; Delaney, Verkoeijen, & Spirgel, 2010). The majority of these studies have been conducted in the laboratory involving adult participants who demonstrate benefits of spacing over relatively brief time intervals of less than one day. However, other studies demonstrate benefits of spacing in more diverse populations such as young children (e.g., Rea & Modigliani, 1985; Toppino, 1991; Toppino & DiGeorge, 1984) and older adults (e.g., Balota, Duchek, & Paullin, 1989). Of particular interest here are studies that have demonstrated benefits of spacing on long-term learning in authentic educational contexts.

Recent studies conducted in classrooms have revealed promising results for the potential of spacing to enhance student learning in real educational settings. In one recent study, Sobel, Cepeda, and Kapler (2011) taught fifth grade students the definitions for several uncommon English words (e.g., abscond). Students learned these terms and definitions by means of an interactive tutorial that included oral practice, paper-and-pencil activities, and instructor-led demonstrations. Students completed the tutorial twice, with both sessions either occurring on the same day (massed group) or on two separate days separated by one week (spaced group). Five weeks after completing the second tutorial, all students completed a final test over the vocabulary terms. On this test, retention of vocabulary definitions was significantly higher for students who repeated the tutorial after one week (20.8%) than students who repeated the tutorial on the same day (7.5%).

In another study, Carpenter, Pashler, and Cepeda (2009) examined eighth-grade students' retention of information they had learned in their U. S. history course as a function of when they received a review over the information. After completing the course, all students participated in a review session that involved answering questions that were drawn from the last unit of their course (e.g., Who assassinated president Abraham Lincoln?). Students provided their answers on a sheet of paper, and then were given a sheet of paper containing the correct answers to check their accuracy. One group completed this review activity one week after the course ended, and another group completed the review 16 weeks after the course ended. Nine months after completing the review, all students completed a final test assessing their knowledge of the same information. On this test, retention of course content was significantly higher for students who completed the review after 16 weeks (12.2%) compared to one week (8%). After such a lengthy test delay, it is not surprising that substantial forgetting occurred in both of these classroom-based studies. The key finding from these studies is that spacing appeared to reduce the amount of information forgotten over time, resulting in a significantly greater proportion of information retained, even after several months.

In another classroom-based study, Seabrook , Brown, and Solity (2005) demonstrated significant benefits of spacing on first graders' learning of phonics reading skills. During their regularly scheduled classes, all children received a total of six minutes of phonics instruction per day for two weeks. One group (the massed group) received a single session of instruction that lasted six minutes, whereas another group (the spaced group) received three separate instructional sessions on the same day, each lasting two minutes. At the end of two weeks, students who completed the spaced two-minute sessions showed greater improvement in these skills than students who completed the massed six-minute sessions.

Other studies involving adult learners have demonstrated significant benefits of spacing on the learning of educationally-relevant materials. For example, Bird (2010) found that the learning of English grammatical rules, as assessed by a test given 60 days after learning, was enhanced by practicing these rules with a 14-day spacing gap as compared to a three-day spacing gap. Rohrer and Taylor (2006) taught participants to solve a mathematics permutation problem by working through 10 practice problems on the same day (zero-day spacing gap) vs. working on the same 10 problems divided over a spacing gap of seven days (i.e., five problems on one day, and five problems a week later). A final test covering the same type of problems given four weeks later demonstrated significantly greater retention for the group that worked through the problems with the seven-day spacing gap (64%) vs. the zero-day spacing gap (32%).

In another demonstration of the long-term benefits of spacing, Bahrick (1979) taught participants English-Spanish vocabulary through six learning sessions that were administered after spacing gaps of zero days (i.e., all six sessions took place on the same day), one day (i.e., one session per day for six days), or 30 days. A final test given one month after the last learning session revealed the best retention for participants who experienced the 30-day spacing gap. In a follow-up study, the advantage for those who practiced with the 30-day spacing gap was still apparent after eight years (Bahrick & Phelps, 1987).

The studies by Bahrick raise an important question: Are longer spacing gaps always better? A recent study by Cepeda et al. (2008) revealed important information about the optimal spacing gap. In this study, adult participants learned a set of obscure trivia facts via a flashcard-like tutorial that involved answering questions (e.g., What country's flag consists of a single solid color?) and receiving feedback (e.g., Libya). Participants completed two of these learning sessions followed by a final test. Each participant experienced a different combination of spacing gap and test delay, with spacing gaps ranging between zero and 105 days, and test delays ranging between seven and 350 days. Cepeda et al. found that the optimal spacing gap depended on the test delay (see Figure 2), such that shorter spacing gaps tended to produce better retention after relatively short delays, whereas longer spacing gaps tended to produce better retention after longer delays. More specifically, memory retention was best when the spacing gap was about 10-20% of the test delay. Thus, in order to optimize the benefits of spacing, learners should be aware of how long they wish to retain the information, with longer spacing gaps ideal for long-term retention.

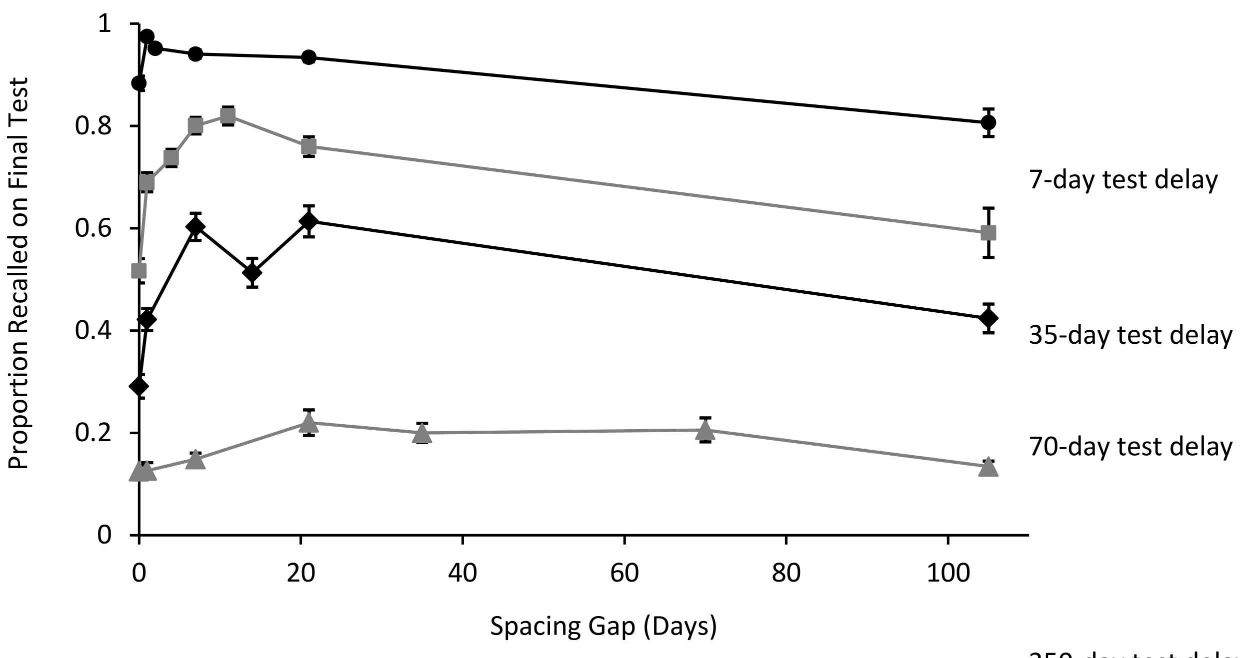


Figure 2. Memory retention of obscure facts from Cepeda et al. (2008). Participants le through various combinations of spacing gaps (0, 1, 2, 4, 7, 11, 14, 21, 35, 70, or 105 days) and test delays (7, 35, 70, or 350 days). Shorter spacing gaps (e.g., 1 day) were more beneficial for relatively short-term retention (e.g., 7-day test delay), whereas longer spacing gaps (e.g., 21 days) were more beneficial for longer-term retention (e.g., 70 days).

The Interleaving Effect

In any academic subject, students must learn a variety of different subareas. For example, in learning to conjugate verbs in a foreign language, students need to learn present-tense conjugations, as well as past-tense and future-tense conjugations. In geometry, students must learn different formulas for calculating the volume of different types of objects. Just like the decision of how to schedule repetitions of the same information (i.e., spacing vs. massing), students and instructors are also faced with the decision of how to schedule study repetitions of different types of material within a particular subject.

An intuitive approach is to study the same type of material over and over again before moving on to a different type of material. For example, in foreign language instruction, students typically receive extensive practice at conjugating verbs in the present tense before moving on to past-tense or future-tense conjugations. In mathematics learning, students often receive repeated practice at solving one type of problem (e.g., addition of fractions) before moving on to a different type of problem (e.g., subtraction or multiplication of fractions). In this sense, the scheduling of repetitions is said to be ***blocked*** by type of problem, such that one type of problem is repeatedly practiced before moving on to a different type.

An alternative approach is to practice all of the problems in an order that is more random and less predictable. For example, after learning about fractions, students may encounter one problem requiring addition of fractions, followed by a problem requiring subtraction, then a problem requiring multiplication, then another problem requiring subtraction, and so on. In this case, the order of problems is said to be ***interleaved*** because problems of a different type are mixed together rather than separated.

Figure 3 illustrates the design of a typical study on interleaving vs. blocking. Here, students receive the same number of exposures to the material (e.g., three examples each of concepts A, B, and C). The only difference is that blocking involves repeated exposure to one concept at a time, whereas interleaving involves a mixed presentation of all three concepts.

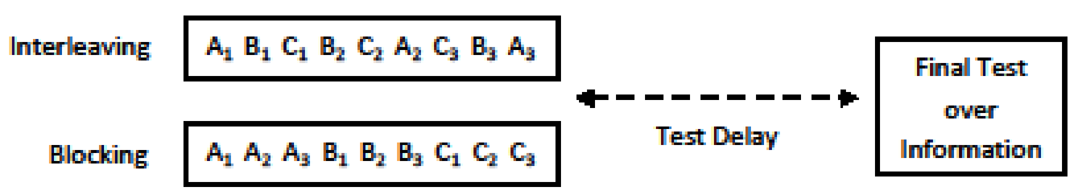


Figure 3. Illustration of a typical experiment on interleaving. Students encounter multiple concepts (A, B, and C) in an order that is either blocked by concept, or interleaved such that a mixture of all of the concepts is presented together. Retention on a final test is typically better following interleaving than blocking.

Studies on this topic typically find that interleaving produces better learning than blocking, a finding referred to here as the ***interleaving effect***. For example, in a study by Rohrer and Taylor (2007), students learned the formulas for calculating the volume of four different types of solid figures—wedges, spherical cones, spheroids, and half cones. All students worked through 16 practice problems and were given feedback on the answer after completing each problem. Problems were either blocked by type of figure (e.g., four problems on wedges, followed by four problems on spherical cones, followed by four problems on spheroids, and four problems on half cones), or interleaved (e.g., the same 16 problems occurred in mixed fashion). One week later, a test given over new problems that required the same formulas revealed significantly greater performance for those students who learned through interleaving (63% accuracy) compared to blocking (20% accuracy).

Research on Interleaving

A number of other studies have demonstrated benefits of interleaving in mathematics learning. In a study by Taylor and Rohrer (2010), fourth grade students were taught a two-step formula for calculating each of four properties of a prism—the number of faces, corners, edges, and angles—based on information provided about the number of base sides. For each problem, students were presented with a question (e.g., A prism with 17 base sides has how many edges?), wrote down their answers in a booklet, and were then shown the correct answer. Due to the complexity of the task, students first completed a partial practice phase (requiring them to solve one step of each problem) consisting of 32 problems (eight of each type). Students in the Blocked Group completed all of the problems of one type before moving on to problems of a different type (e.g., eight face problems, followed by eight corner problems, etc.), whereas students in the Interleaved Group completed the same 32 problems in an order that was intermixed and pseudo-randomized such that each type of problem was fairly evenly balanced across the 32 presentations and no two problems of the same type occurred consecutively. Students then completed a full practice phase (requiring them to solve both steps of each problem) consisting of 12 additional problems (three of each type) arranged in the same blocked vs. interleaved fashion as before. On a final test one day later, students were given four new problems (one of each type), and were required to provide the full solution without receiving feedback. Performance on this test revealed significantly higher accuracy for students who learned the problems through interleaving (77%) compared to blocking (38%).

In another study on mathematics learning, Mayfield and Chase (2002) observed benefits of interleaving on the learning of algebraic rules. Over the course of several sessions, college students with poor mathematics skills learned five algebraic rules through completing worksheets that provided an explanation of the rule, examples, and practice problems. One group of students learned the rules through a procedure akin to blocking, in which each rule was learned and then practiced extensively before moving on to the next rule. Another group learned the same rules through a procedure akin to interleaving, which involved continuous practice of previously-learned skills. For example, after learning two types of skills (e.g., order of operations and multiplying exponents), students practiced a mixture of problems tapping these two skills. A third skill was then introduced (e.g., finding the roots of exponents) and this skill was then added to the practice set along with the two previously-learned skills, such that participants practiced a mixture of problems tapping all three skills. A fourth skill was then added, followed by a mixture of problems tapping all four skills, and so on, until all five skills had been learned and practiced. Although a number of logistical constraints prevented achieving precise control over factors such as the exact timing and amount of exposure to practice problems, the results of this study suggest that an approach based on interleaving appears to be more effective than one based on blocking. An exam given at the end of training that assessed knowledge of all five skills revealed a significant benefit of interleaving (97% correct) over blocking (85% correct).

A different line of research has shown that interleaving can also benefit the learning of categorical concepts. In a study by Kornell and Bjork (2008), adult participants learned to classify paintings by particular artists (e.g., Georges Braque, Judy Hawkins, Bruno Pessani, etc.). Participants were shown several example paintings by each artist, in an order that was either blocked by artist (i.e., several Braque paintings, followed by several Hawkins paintings, followed by several Pessani paintings, etc.), or interleaved such that no two paintings by the same artist occurred consecutively (e.g., Braque, Hawkins, Pessani, etc.). Later on, participants were shown new paintings by the same artists, and their job was to identify which artist had painted each painting. Participants who learned the paintings through interleaving, as compared with blocking, performed significantly better at this task. Similar results were observed in a study by Kang and Pashler (2012) who used similar types of paintings, and interleaving has also been shown to benefit participants' abilities to classify different bird species (e.g., Wahlheim, Dunlosky, & Jacoby, 2011).

A logical question concerns whether the benefits of interleaving are simply due to spacing. As discussed in the previous section, two presentations that are spaced apart in time are learned better than two presentations that occur consecutively. Because of the presence of other items in-between each presentation of a given concept, interleaving naturally involves spacing, so the benefits of interleaving may simply reflect the spacing effect. The benefits of interleaving do not appear to be due solely to spacing, however. Some studies have inserted time delays in-between successive presentations of blocked items to match the time delays that occur during interleaved presentations, and have still observed significant benefits of interleaving over blocking (e.g., Kang & Pashler, 2012; Taylor & Rohrer, 2010).

What might account for the benefits of interleaving, then? It has been proposed that interleaving benefits students' ability to make important discriminations between concepts that are easily confused (e.g., Rohrer, 2012). For example, students may confuse the formula for calculating the volume of one type of geometric figure with the formula for another type of figure. Or, based on similarities in painting styles, they may confuse the artist of one particular painting with the artist of another painting. For materials that are prone to such confusions, interleaving may be beneficial because it allows students the opportunity to notice the key differences between two similar concepts, helping them to make that important distinction later on. Supporting this notion, Rohrer and Taylor (2010) found that interleaving was particularly beneficial for reducing errors in discrimination on the final test. That is, students who learned the prism problems through interleaving (compared to blocking) were much less likely to mistakenly apply the solution for one type of problem (e.g., edges) to another type of problem (e.g., corners). Interleaving thus enhanced the abilities of students to apply the appropriate solution to the problem.

There are some exceptions to the benefits of interleaving. In a recent study by Carpenter and Mueller (2013), native English speakers learned several pronunciation rules in French by seeing and hearing example words that represented these rules. The words were either blocked by rule (e.g., bateau, fardeau, tonneau... mouton, genou, verrou... tandis, verglas, admis) or interleaved such that no two words representing the same rule ever occurred consecutively (e.g., bateau, mouton, tandis… fardeau, genou, verglas… tonneau, verrou, admis). On a later test, participants demonstrated greater proficiency for pronunciation rules that were learned through blocking compared to interleaving. The reason why the usual advantage of interleaving was not observed in this study could be due to the possibility that this type of learning did not require discriminative contrast. Instead of noticing key differences between mathematical formulas or artists' painting styles that are easily confused, learning a series of fairly distinct pronunciation rules may benefit more by noticing the similarities among words that share a given rule (e.g., the sound made by "eau" after listening to bateau, fardea, tonneau, etc.), rather than noticing the differences between words that represent different rules. Thus, it is likely that the utility of interleaving depends on the degree to which the learning task requires noticing of similarities within a category, vs. noticing of differences between categories.

Some research also suggests that blocking may be more effective than interleaving when the learning task is particularly difficult. For example, de Croock and van Merriënboer (2007) presented participants with different types of problems that could occur in a simulated complex distiller system (e.g., pipe leakage, sensory malfunction, etc.), and participants were required to troubleshoot these problems. Participants performed better when problems were presented to them in a way that was blocked by type of malfunction, rather than interleaved such that the type of malfunction was different on each problem.

Thus, although some studies have revealed promising effects of interleaving on the learning of mathematical formulas and visual discriminations, much research remains to be done before we fully understand the boundary conditions of the interleaving effect, particularly with respect to how it may be affected by the complexity of the task and the type of processing required—noticing of similarities vs. noticing of differences. For difficult concepts, one approach that may prove useful is to implement a mixture of blocking and interleaving. For example, it may be better to use blocking during early stages of learning, and then once the concept becomes familiar and students begin to grasp it better, they may transition to interleaving. This hybrid approach has been discussed (e.g., Rohrer, 2012), but has not yet been thoroughly explored in any of the known research on interleaving.

Implementing Spacing and Interleaving in Educational Curricula

There are several practical ways that students and instructors might implement spacing and interleaving in educational practice. Research on the spacing effect suggests that students' long-term retention of information is enhanced by reviewing previously-learned information. In particular, if instructors wish for students to retain the information over the long-term, it may be best to review previously-learned information at periodic time intervals of several weeks (e.g., Cepeda et al., 2008).

One way to accomplish this might be to incorporate into current lectures and class activities brief reviews of previously-learned information. These reviews could also be implemented as homework assignments, which may be particularly advantageous when class time is limited. For example, while learning about dependent-samples t-tests in a statistics course, students could receive a number of practice problems dealing with dependent-samples t-tests, in addition to some problems that cover the earlier-learned concepts of independent-samples t-tests and one-sample t-tests. Such an approach capitalizes on the benefits of spacing by requiring students to revisit previously-learned material. Including different types of (potentially confusable) problems also capitalizes on the benefits of interleaving by providing students with the opportunity to notice key differences in the computational steps required for each of these procedures.

Another way to review previously-learned information is to implement cumulative exams and quizzes. By testing students over information that was learned relatively recently as well as in the more distant past, the exams themselves will serve to re-expose students to previously-learned information. In addition, cumulative assessments will likely have an indirect benefit on student learning by providing students with a good incentive to review the information during their own study time.

Research on interleaving shows that students' retention of mathematics knowledge is significantly enhanced by simply re-arranging the order of practice problems. Thus, instructors may wish to take a number of practice problems dealing with a particular concept (e.g., fractions) and give these to students in an order that is unpredictable (e.g., some problems dealing with addition of fractions, some with multiplication, some with division, etc.), as opposed to an order that is blocked by type of problem (e.g., several addition problems, followed by several subtraction problems, etc.).

These implementations do not come without challenges. One obstacle is that students may be reluctant to use spacing and interleaving because these methods increase the perceived difficulty of learning. Students work more slowly, and make more errors, when working a set of practice problems that is interleaved rather than blocked. For example, in Rohrer and Taylor's (2010) study, students who worked practice problems in blocked fashion performed much better during the practice problems (89% accuracy) than students who worked the same problems in interleaved fashion (60% accuracy). Even though performance on the final test—a more reliable indicator of long-term learning—revealed significant advantages for interleaving over blocking, students may nonetheless be disinclined to use interleaving because of the fact that it makes learning *seem* harder. Some studies have surveyed students on the perceived effectiveness of both methods, and found that students consider blocking to be a more effective strategy, even if their own memory retention favors interleaving (e.g., Kornell & Bjork, 2008). The fact that blocking makes learning easier initially may lead students to favor this approach and be reluctant to implement the more difficult, yet more effective, method of interleaving.

Spacing too may be difficult to implement for the same reason. When reviewing information after several days or weeks, it is natural for students to forget much of what they have learned previously. Experiencing difficulty in remembering previously-learned concepts, and perhaps needing to consult previous chapters or notes, may result in discouragement and the sense that one's previous efforts to learn the information were ineffective. Importantly however, the process of forgetting occurs rapidly, even for well-learned material (e.g., Carpenter, Pashler, Wixted, & Vul, 2008), and should not be taken as a sign that previous learning efforts were in vain. Although it may at first seem that students need to re-learn a great deal of what they have learned before, students can re-learn information faster than they originally learned it (e.g., Berger, Hall, & Bahrick, 1999), and re-visiting this knowledge at periodic time intervals will help inoculate it against further forgetting.

A final obstacle to implementing the principles of spacing and interleaving is the fact that educational materials do not always encourage these approaches. Although spacing improves foreign language vocabulary learning (e.g., Bahrick, Bahrick, Bahrick, & Bahrick, 1993), the vocabulary items in many foreign language textbooks are contained within chapters that are devoted to specific topics (e.g., food, parts of the body), and these items typically do not re-appear in later chapters. Similarly, mathematics textbooks contain chapters that are typically devoted to specific topics, and practice problems usually pertain only to the most recently-learned topic. Given that educational materials do not typically provide spacing and interleaving, instructors may need to modify or supplement the information from these materials in order to ensure that students have the opportunity to use these methods.

References

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108,* 296-308. DOI: 10.1037/0096-3445.108.3.296.

Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 13,* 344-349. DOI: 10.1037/0278-7393.13.2.344.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science, 4,* 316-321. DOI: 10.1111/j.1467-9280.1993.tb00571.x.

Balota, D. A., Duchek, J. M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology & Aging, 4,* 3-9. DOI: 10.1037/0882-7974.4.1.3.

Berger, S. A., Hall, L. K., & Bahrick, H. P. (1999). Stabilizing access to marginal and submarginal knowledge. *Journal of Experimental Psychology: Applied, 5,* 438-447. DOI: 10.1037/1076-898X.5.4.438.

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied PsychoLinguistics, 31,* 635-650. DOI: 10.1017/S0142716410000172.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name-learning. *Applied Cognitive Psychology, 19,* 619-636. DOI: 10.1002/acp.1101.

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition, 41,* 671-682. DOI: 10.3758/s13421-012-0291-4.

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U. S. history facts. *Applied Cognitive Psychology, 23,* 760-771. DOI: 10.1002/acp.1507.

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition, 36,* 438-448. DOI: 10.3758/MC.36.2.438.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354-380. DOI: 10.1037/0033-2909.132.3.354.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science, 19,* 1095-1102. DOI : 10.1111/j.1467-9280.2008.02209.x.

de Croock, M. B. M., & van Merriënboer, J. J. G. (2007). Paradoxical effects of information presentation formats and contextual interference on transfer of a complex cognitive skill. *Computers in Human Behavior, 23,* 1740-1761. DOI: 10.1016/j.chb.2005.10.003.

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *The Psychology of Learning & Motivation: Advances in Research & Theory, 53,* 63-147. DOI: 10.1016/S0079-7421(10)53003-2.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79,* 162-170. DOI: 10.1037/0022-0663.79.2.162.

Kahana, M. J., & Howard, M. W. (2005). Spacing and lag effects in free recall of pure lists. *Psychonomic Bulletin & Review, 12,* 159-164. DOI: 10.3758/BF03196362.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26,* 97-103. DOI: 10.1002/acp.1801.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction?" *Psychological Science, 19,* 585-592. DOI: 10.1111/j.1467-9280.2008.02127.x.

Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis, 35,* 105-123. DOI: 10.1901/jaba.2002.35-105.

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning: Journal of Practical Research & Applications, 4,* 11-18. DOI: Unavailable.

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review, 24,* 355-367. DOI: 10.1007/s10648-012-9201-3.

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20,* 1209-1224. DOI: 10.1002/acp.1266.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science, 35,* 481-498. DOI: 10.1007/s11251-007-9015-8.

Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19,* 107-122. DOI: 10.1002/acp.1066.

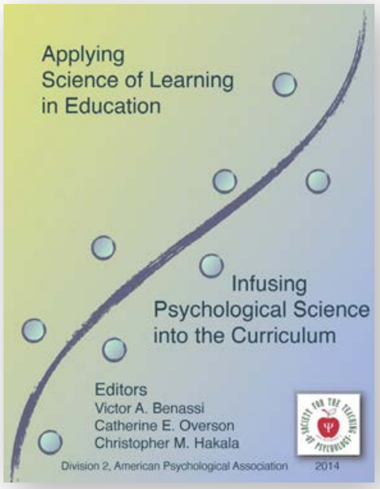
Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology, 25,* 763-767. DOI: 10.1002/acp.1747.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24,* 837-848. DOI: 10.1002/acp.1598.

Toppino, T. C. (1991). The spacing effect in young children’s free recall: Support for automatic-process explanations. *Memory & Cognition, 19,* 159-167. DOI: 10.3758/BF03197112.

Toppino, T. C., & DiGeorge, W. (1984). The spacing effect in free recall emerges with development. *Memory & Cognition, 12,* 118-122. DOI: 10.3758/BF03198425.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition, 39,* 750-763. DOI: 10.3758/s13421-010-0063-y.



How Accuracy in Students' Self Perceptions Relates to Success in Learning

Joyce Ehrlinger and E. Ashley Shain
Washington State University

Most teachers are all too familiar with the challenge of helping students who remain baffled about why they are not doing well in class. "But I studied so hard!" students might say. "I don't understand how I could have gotten such a low test score!" The frequency of such concerns among students highlights the difficulty that students face when trying to understand whether they have truly mastered learning a topic, when they prepared themselves well for exams, and when their learning might fall short.

Effective learning requires metacognition — the process of reflecting upon which learning concepts one has already mastered, what still needs to be learned, and how to best approach the task of learning. Within educational contexts, metacognitive perceptions are defined as students' beliefs about their learning process, including the degree to which they have mastered learning tasks. Students' metacognitive perceptions reflect the degree to which they believe that they understand lesson material and have mastered learning different concepts or tasks. In this chapter, we review research suggesting that students' metacognitive perceptions are often error-prone. We discuss factors that introduce error into students' perceptions and offer concrete suggestions designed to help encourage greater accuracy in students' perceptions and, in turn, to facilitate student learning.

Accuracy of Metacognitive Perceptions

To the degree that students hold accurate metacognitive perceptions during learning, they can make wise decisions about how best to allocate their time studying and which strategies to use in order to learn course material. However, it can be very challenging for students to have true insight into the quality of their understanding and learning. Laboratory-based research suggests that there are some learning tasks about which students make quite accurate predictions about their level of learning and about the degree to which they will perform well on tests of their knowledge (e.g., King, Zechmeister, & Shaughnessy, 1980; Leonesio & Nelson, 1990; Nelson & Dunlosky, 1991). However, for many learning tasks, students' metacognitive perceptions tend to be error-prone. The degree to which students' perceptions are accurate depends on the topic being studied and the students' level of ability, as well as a number of other factors. On average, however, research suggests that students' metacognitive perceptions tend to be plagued by error across a variety of academic subjects (for review, see Dunning, 2005). Often students are wildly inaccurate about a host of abilities and personal qualities (e.g., Fischhoff, Slovic, & Lichtenstein, 1977). In the classroom, many students tend to hold overly positive impressions of their current level of knowledge (e.g., Ehrlinger & Dunning, 2003), the quality of their performance on academic tests (e.g., Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008), and the speed with which they can master and complete future academic tasks (e.g., Buehler, Griffin, & Ross, 1994).

One might wonder whether there is any reason to be concerned about overconfidence in students. After all, confidence is often considered a positive trait. Indeed, research suggests that students who are confident about their ability to succeed in school tend to perform better on academic tests than those with less confidence (e.g., Zimmerman, Bandura, & Martinez-Pons, 1992). That said, negative consequences also stem from being too confident in the classroom. Students who are overconfident about their ability to succeed in college end up feeling more disconnected and disillusioned than those with more modest expectations (Robins & Pals, 2001). Overconfidence can also leave students with mistaken impressions that they are fully prepared for tests and no longer need to study (Metcalfe & Finn, 2008; Metcalfe & Kornell, 2005). People who have relatively accurate perceptions regarding their progress in learning tend to use more effective study habits and perform better on tests than do those with more error-prone views of their knowledge (Thiede, Anderson, & Therriault, 2003).

Overconfidence can be particularly troubling when it exists in those who are truly performing poorly in the classroom. Top students do often hold overly positive views of their learning but this seems far less troubling than overconfidence among those with failing test scores. Tragically, it tends to be the poor students, however, who demonstrate the most overconfidence (Dunning, Johnson, Ehrlinger, Kruger, 2003; Ehrlinger et al., 2008). Consider, for example, one study in which college students who had just completed a test in their psychology course were then asked to estimate how well they had performed. Overconfidence was common in students' performance estimates. However, those students scoring in the bottom quartile were, by far, the most overconfident students in the class. The actual performance of these students averaged in the 12^{th} percentile in the class. However, these students estimated that they had performed in the 61st percentile on average (Ehrlinger et al., 2008). They were vastly overconfident about the quality of their test performance. This seems almost unfair. Certainly poor students have the most room for improvement and, thus, the greatest need for understanding when improvement is necessary. However, these students lack knowledge necessary to know when they have offered answers that are incorrect. Thus, bottom performers in the class struggle both with their poor grades and with particularly error-prone, overconfident metacognitive perceptions about their progress in learning.

Although overconfidence creates troubling roadblocks to learning, the opposite issue — underconfidence — comes with its own costs in the classroom. Perhaps the most noteworthy cases in which underconfidence impedes learning are in the fields of science, technology, engineering, and math (STEM). A wealth of research suggests that women in general and, in particular, women who are Black, Hispanic, or American Indian are vastly underrepresented in STEM fields (e.g., NSB, 2006; NSF, 2009). For example, women account for less than 20% of undergraduate computer science and engineering degrees and less than 25% of the doctoral degrees awarded in these fields (NSF, 2009). Multiple factors combine to create and perpetuate the lack of diversity in STEM fields. However, underconfidence is among the factors that dissuade women and underrepresented minorities from participating in these fields. Women and underrepresented minority students tend to report lower confidence in their abilities for STEM subjects compared to white men (Eccles, Barber, Jozefowicz, Malenchuk, & Vida, 1999; Ehrlinger & Dunning, 2003). This lack of confidence then leads these students to perceive that they are performing poorly in their classes and, ultimately, to avoid future opportunities related to STEM. For example, Ehrlinger and Dunning (2003) demonstrated that a tendency for women to hold less confidence in their scientific ability, compared to men, led women to be less confident about their performance on a science test. This difference in confidence emerged even though men and women performed equally well on the test! Further, women's comparative lack of confidence led them to show less interest in a future science opportunity than men in the study (see also Ehrlinger, 2008; Goodman et al., 2002). Thus, even talented students can end up withdrawing from important fields because they lack the confidence necessary to understand how much they could accomplish.

How Metacognitive Monitoring Relates to Learning

One important reason that students might hold overly positive or, in the case of STEM fields sometimes negative, metacognitive perceptions is that they do not always strive to hold accurate views of their progress in the classroom. Students differ in the degree to which they actively engage in metacognitive monitoring, or monitoring whether and how well they have mastered learning tasks. Individuals who engage in metacognitive monitoring are more likely to discover gaps in their understanding than are those who do not monitor. Metacognitive monitoring allows students to identify which concepts they have mastered, which they want to work on more, and which concepts might be so difficult that it isn't worth the time and effort necessary to truly learn them (Metcalfe & Kornell, 2005; Kornell & Metcalfe, 2006). Past research suggests that students can draw upon this metacognitive monitoring information when deciding whether and when to engage in learning tasks (Ehrlinger & Dunning, 2003), how to focus their study time (e.g., Kornell & Bjork, 2007; Metcalfe & Kornell, 2005), and which strategies to use (e.g, Kornell & Metcalfe, 2006; Thiede & Dunlosky, 1999).

Metacognitive monitoring has important benefits for student learning. Those who engage in monitoring perform better on measures of learning than those less likely to monitor. Students who do not monitor are likely able to maintain confidence in their skill level even when it is not merited. By engaging in metacognitive monitoring, students take an active role in the process of learning (Chen, 2002; Schraw & Dennison, 1994; Vrugt & Oort, 2008; Yukselturk & Bulut, 2007). In addition, study strategies that entail metacognitive monitoring are particularly effective in advancing learning. For example, quizzing oneself on course material (Roediger & Karpicke, 2006; see Pyc, Agarwal, & Roediger, this volume, for a review) and summarizing key points from text passages (Thiede, Anderson, & Therriault, 2003) lead to improvements in learning. Further, students who are encouraged to monitor their understanding and knowledge show better student outcomes than those who are not. Students who received instruction in metacognitive monitoring (i.e., were asked to make confidence judgments) engaged more fully in the classroom, remembered more course material, and performed better on subject-matter tests than control participants (Georghiades, 2000). In addition, students required to make judgments of learning in conjunction with readings in their physics class demonstrated significantly higher text comprehension than control participants (Koch, 2001). It is worth noting, however, that the degree to which metacognitive monitoring leads to improvements in student learning depends heavily on the degree to which students are able to accurately evaluate their level of knowledge and understanding (Dunlosky, Hertzog, Kennedy, & Theide, 2005). To the degree that students are able to accurately identify what they have not fully mastered, they can effectively direct their efforts toward improving mastery (e.g., Kornell & Bjork, 2007; Metcalfe & Finn, 2008; Thiede et al., 2003) and perform better on tests (Dunlosky & Rawson, In Press).

Recommendations for Educators

We have described research suggesting that students who develop accurate metacognitive perceptions and who monitor their metacognition reap benefits in terms of success in learning. What, then, can teachers do to encourage accurate metacognition and metacognitive monitoring among their students? There are at least two broad strategies that educators can utilize in order to encourage improved metacognition in the classroom. First, we will discuss how teachers can encourage students to utilize learning strategies that have been demonstrated to help them gain more accurate perceptions of their learning and, in turn, to perform better on tests. Finally, we will discuss the ways in which teachers can encourage positive mindsets that minimize the degree to which students feel threatened by difficulty in the classroom and, as such, are more open to understanding how they can improve their success in learning.

Encouraging metacognitive monitoring through study strategies

We began this chapter by referencing the challenges that teachers face in trying to help students who fail to understand why they have performed poorly. As mentioned earlier, these students serve as salient reminders that they do not always hold accurate perceptions of their learning. Further conversation with these frustrated students also tends to reveal other ways in which they lack insight in the classroom. Often, they fail to understand the best strategies for improving their level of learning. It is especially important that students learn more effective strategies so that they can better direct their own learning and have more accurate insight into their learning processes.

Students do not always have good insight into what specific strategies they need to use to improve their learning (e.g., Pyc & Dunlosky, 2010), so it is important for educators to explicitly encourage these strategies or use them when appropriate as part of a classroom setting. For example, students may feel like cramming all of their studying into a long session just before the test is an effective strategy for learning information. Indeed, this type of cramming before an exam can leave students feeling very confident of their knowledge (Baddeley & Longman, 1978; Bjork, 1999; Simon & Bjork, 2001). However, actual sustained learning is much more likely to result from spacing out one's study (Bahrick, 1979; Schmidt & Bjork, 1992; see Carpenter, this volume, for a review). Students are particularly likely to mass their study together into single study sessions when working on difficult material (e.g., Pyc & Dunlosky, 2010). However, spacing out one's study over time can hold particularly strong benefits for learning difficult concepts (Cuddy & Jacoby, 1982).

Another common learning strategy that can impede accurate metacognition is the use of flashcards to learn or study new material. Students who use flashcards may become overconfident that they have learned association because they have immediate access to an answer (Koriat & Bjork, 2006). This overconfidence may lead students to neglect practice on items they do not understand well. A more helpful strategy that gives students better insight into their own learning is for students to quiz themselves on the study material, which has been show to encourage long-term learning of material (e.g., Roediger & Karpicke, 2006). When students self-quiz, they are required to think of questions to ask of themselves (i.e., judge what is important about the topic), and if used properly, this strategy will help students simulate the testing environment by answering their own questions without immediate access to answers to the questions. Other effective learning strategies include summarizing main points of a text (Thiede, Anderson, & Therriault, 2003), and asking students to make judgments of how well they understood reading or textbook passages for their classes (Koch, 2001). To summarize a text, students need a basic understanding of what the text means and how the main points in the text relate to one another; this process helps provide students with feedback on what they have learned or not learned according to how well they can write about the text. Further, making judgments of understanding for a text overtly asks students to think about what they need to read over again or spend more time understanding. Overall, more effective learning strategies help give students more feedback and insight into how they learn and how they learn well. For students, having better insight into their own learning is associated with improved long-term retention of material and improved performance in college (Chen, 2002; Schraw & Dennison, 1994; Vrugt & Oort, 2008; Yukselturk & Bulut, 2007). Teaching these strategies to students and reinforcing them through classroom use (e.g., asking students to reflect on how well they understood various parts of a class reading) will give them the tools to gain better insight into their own learning.

Encouraging metacognitive monitoring through motivation

We know that metacognitive monitoring helps students to see the "big picture" of their learning habits and current knowledge, which then can provide an opportunity for students to improve learning habits and increase knowledge when appropriate. Given the opportunity to gain knowledge and improve the learning experience, why do some students not engage in metacognitive monitoring? Even when students are taught effective study strategies for learning, they might not be willing to expend the extra effort inherent in some of the most effective learning strategies. Students are especially vulnerable to ineffective habits when facing difficult material. Even though spacing study out over time instead of massing it together creates "desirable difficulties" that promote sustained learning of material (e.g., Schmidt & Bjork, 1992; Bjork, 1994; see Clark & Bjork, this volume, for a review), students often choose to mass study rather than space it over time (Pyc & Dunlosky, 2010). When students separate study sessions, they must come back to the material and reevaluate what they remember and need to study harder since the last study session. Students who do not space practice may have a hard time realizing how much they have left to learn, contributing to overconfidence in their learning.

One strategy that teachers can use to help students use effective learning strategies is to foster mastery goals in the classroom. Students who hold mastery goals focus mainly on trying to learn new material, while students holding performance goals tend to focus on appearing intelligent and avoiding failure (e.g., Elliot & Harackiewicz, 1996; Elliot & McGregor, 2001). Educators can help students by creating a "culture of learning" in the classroom, which would entail encouraging students to adopt mastery goals by focusing on acquiring new knowledge and skills, even when doing so may mean that students look unintelligent or failing sometimes. Students who hold mastery goals tend to process material more deeply (Elliot, McGregor, & Gable, 1999), seek out help more often when appropriate (Linnenbrink, 2005) and ultimately have greater learning success than students holding performance goals (Elliot & McGregor, 1999; Robbins et al., 2004).

Similarly, teachers can encourage learning by teaching students an incremental view of intelligence. An incremental view of intelligence is characterized by the belief that intelligence is malleable and can be developed. In contrast, one might hold an entity view of intelligence, which is characterized by the belief that intelligence is fixed (for review, see Dweck, 1999). Students' theories of intelligence influence the goals that students adopt (Dweck & Leggett, 1988; Elliot & McGregor, 2001). For example, a view of intelligence as malleable encourages students to adopt learning goals over performance goals (Blackwell et al., 2007; Robins & Pals, 2002). In contrast, an entity view also predicts a tendency to make helpless attributions when facing difficulty on learning tasks (Blackwell et al., 2007; Hong et al., 1999; Nussbaum & Dweck, 2008; Robins & Pals, 2002). Students with an incremental theory, in comparison, make more beneficial attributions in the face of difficulty and report more positive beliefs about effort (Blackwell et al., 2007). These differences in attributions and perceptions of effort carry over to help students perform well on learning tasks. For example, Blackwell, Trzesniewski, and Dweck (2007) found that holding an incremental view of intelligence in the 6th grade predicted higher grades in students' math classes two years later. Other research has shown benefits of teaching students an incremental view of intelligence for math grades (Aronson et al., 2002) and standardized test scores (Good, Aronson, & Inzlicht, 2003). Collectively, this research provides intriguing evidence that teaching students an incremental view of intelligence can improve student achievement.

Research suggests that an incremental theory of intelligence is also associated with more effective study practices (e.g., Cury, Da Fonseca, Zahn, & Elliot, 2008; Kovach, Fleming, & Wilgosh, 2001) and more accurate metacognitive perceptions (Ehrlinger & Dweck, 2013). For example, Blackwell, Trzesniewski, and Dweck (2007) found that students with an incremental theory of intelligence reported a greater willingness to engage in positive study strategies than entity theorist students (e.g., willingness to "work harder in this class from now on" and "spend more time studying for tests"). In addition, incremental theorists engaged in reading strategies that aid comprehension more often than did entity theorists (Law, 2009) and were more likely to report using positive learning strategies (Braten & Olaussen, 1998; Stipek & Gralinski, 1996), including elaborating and organizing concepts when studying (Lachman and Andreoletti, 2006; Ommundsen, Haugen & Lund, 2005; Smith, 2005). Past research has also found that high endorsement of performance goals correlates with lower self-reported use of self-regulated learning strategies (Bartels & Magun-Jackson, 2009; Vrugt & Oot, 2008). Ehrlinger and Ward (2012) found that high school students with stronger incremental beliefs more often reported using deep learning strategies than students with stronger entity views, as measured by responses to questions like "I reviewed material by trying to see how new material related to things I knew" and "When I studied I put important ideas into my own words." We found that stronger incremental views also predicted greater reported use of metacognitive monitoring, as measured by agreement with questions like "I asked myself questions to make sure I knew the material I had been studying" and "I thought about what I already knew and what I still had to learn." Further, students with an incremental view offered less overconfident, more accurate self-assessments than those with a stronger entity view. These findings suggest that incremental theorists might choose self-regulated strategies and show greater metacognitive accuracy more often than do entity theorists.

Conclusion

Many students' metacognitive perceptions are error-prone. In particular, those students who are the poorest performers in the class are often the most overconfident. By encouraging students to monitor their level of understanding and progress in learning, teachers can encourage more accurate self-perceptions among their students. Teachers can accomplish this important goal through several strategies outlined in this chapter. By encouraging effective study strategies that increase students' metacognitive insight, teachers can improve both the accuracy in students' metacognitive perceptions and students' success in learning. Similarly, by fostering a culture of learning in the classroom in which intelligence is viewed as a malleable quality and students are encouraged to pursue mastery-oriented learning goals, students are likely to be more open to facing difficulty in the classroom. We view these strategies as important tools that teachers can use to help their students reach their true potential.

References

Aronson, J., Fried, C., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113–125. doi:10.1006/jesp.2001.1491

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics, 21,* 627-635. doi:10.1080/00140137808931764

Bahrick, H. P (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108,* 296-308. doi: 10.1037/0096-3445.108.3.296

Bartels, J.M., & Magun-Jackson, S. (2009). Approach-avoidance motivation and metacognitive self-regulation: The role of need for achievement and fear of failure. *Learning and Individual Differences, 19,* 459-463. http://dx.doi.org/10.1016/j.lindif.2009.03.008

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher and A. Koriat (Eds.), Attention and Performance XVII. *Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). Cambridge, MA: MIT Press.

Blackwell, L., Trzesniewski, K.H., & Dweck, C.S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78,* 246-263. doi: 10.1111/j.1467-8624.2007.00995.x

Braten, I., & Olaussen, B. S. (1998). The relationship between motivational beliefs and learning strategy use among Norwegian college students. *Contemporary Educational Psychology, 23,* 182–194. http://dx.doi.org/10.1006/ceps.1997.0963

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. Journal of Personality and Social Psychology**,** 67, 366-381. doi: 10.1037/0022-3514.67.3.366

Chen, C. S. (2002). Self-regulated learning strategies and achievement in an Introduction to Information Systems course. *Information Technology, Learning, and Performance Journal, 20,* 11-25.

Cuddy, L.J., & Jacoby, L.L. (1982) When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning & Memory, 21,* 451-467. http://dx.doi.org/10.1016/S0022-5371(82)90727-7

Cury, F., Da Fonseca, D., Zahn, I. & Elliot, A. (2008). Implicit theories and IQ test performance: A sequential mediational analysis. *Journal of Experimental Social Psychology, 44,* 783-791. ***doi***: 10.1016/j.jesp.2007.07.003

Dunlosky, J., Hertzog, C., Kennedy, M., & Thiede, K. (2005). The self-monitoring approach for effective learning. *Cognitive Technology, 10,* 4-11.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22,* 271-280. doi: 10.1016/j.learninstruc.2011.08.003.

Dunning, D., Johnson, K., Ehrlinger, J., Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12(3),* 83-87. doi: 10.1111/1467-8721.01235

Dunning, D. (2005). *Self-Insight: Roadblocks and Detours on the Path to Knowing Thyself.* New York, NY: Psychology Press.

Dweck, C. S. (1999). *Self theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press/Taylor and Francis.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95,* 256-273. doi: 10.1037/0033-295X.95.2.256

Ehrlinger, J., & Dweck, C. S. (2013) Understanding overconfidence: implicit theories, preferential attention, and distorted self assessment. *Manuscript in Preparation, Washington State University.*

Ehrlinger J., & Dunning, D. (2003). How chronic self-views influence (and mislead) estimates of performance*. Journal of Personality and Social Psychology. 84,* 5-17. doi: 10.1037/0022-3514.84.1.5

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105, 98-121. doi*:10.1016/j.obhdp.2007.05.002

Ehrlinger, J., & Ward, P.W. (2012). The effect of intelligence beliefs on use of self-regulated learning. *Unpublished data, Florida State University*.

Elliot, A.J., & McGregor, H.A. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 76*, 628-644. *doi.apa.org/psycinfo/1999-10970-009*

Elliot, A.J. & Harackiewicz, J.M. (1996) Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology, 51,* 1058-1068. doi: 10.1177/0146167298247001

Elliot, A., & McGregor, H. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology, 80(3),* 501-519. doi: 10.1037/0022-0663.94.2.381

Fischhoff, B., P. Slovic, and S. Lichtenstein (1977), "Knowing with certainty: The appropriateness of extreme confidence," Journal of Experimental Psychology: *Human Perception and Performance, 3,* 552-564. doi: 10.1037/0096-1523.3.4.552

Georghiades, P. (2000). Beyond conceptual change learning in science education: Focusing on transfer, durability, and metacognition. *Educational Research, 42*, 119-139. doi: 10.1080/001318800363773

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology, 24,* 645-662. doi:10.1016/j.appdev.2003.09.002

Hong, Y., Chiu, C., Dweck, C.S., Lin, D. M.S., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77,* 588-599. doi: *10.1037/0022-3514.77.3.588*

King, J. F., Zechmeister, E. B., & Shaughnessy, J.J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology, 93,* 329-343.

Koch, A. (2001). Training in metacognition and comprehension of physics texts. *Science Education, 85,* 758-768. doi: 10.1002/sce.1037

Koriat, A. & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32,* 1133-1145. doi: 10.1037/0278-7393.32.5.1133

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*, 219-224. ***doi***: 10.3758/BF03194055.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 609-622. doi: 10.1037/0278-7393.32.3.609

Kovach, K., Fleming, D. & Wilgosh, L. (2001). The relationship between study skills and conceptions of intelligence for high school students. *The Korean Journal of Thinking and Problem Solving, 11,* 39-49.

Kruger, J. & Dunning, D. (1999) Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology, 77,* 1121-1134. doi: 10.1037/0022-3514.77.6.1121

Lachman, M. E., & Andreoletti, C. (2006). Strategy use mediates the relationship between control beliefs and memory performance for middle-aged and older adults. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 61B,* 88-94.

Law, Y. (2009). The role of attribution beliefs, motivation and strategy use in Chinese fifth-graders' reading comprehension. *Educational Research, 51,* 77-95. *doi:*10.1080/00131880802704764

Leonesio, R.J., & Nelson, T.O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 464-470. doi: 10.1037/0278-7393.16.3.464

Linnenbrink, E.A. (2005). The dilemma of performance-approach goals: The use of multiple goal contexts to promote students' motivation and learning. *Journal of Educational Psychology*, *97,* 197-213. *doi. 10.1037/0022-0663.97.2.197*

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15,* 174-179. doi: 10.3758/PBR.15.1.174

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52,* 463-477. doi:10.1016/j.jml.2004.12.001

Nelson, T.O., & Dunlosky, J. (1991). The delayed-JOL effect: When delaying your judgements of learning can improve the accuracy of your metacognitive monitoring. *Psychological Science, 2,* 267-270.

Nussbaum, A. D., & Dweck, C.S. (2008) Defensiveness versus remediation: Self-theories and modes of self-esteem maintenance. *Personality and Social Psychology Bulletin, 34,* 599-612. doi: 10.1177/0146167207312960

Ommundsen, Y., Haugen, R., & Lund, T. (2005). Academic self-concept, implicit theories of ability, and self-regulation strategies. *Scandinavian Journal of Educational Research, 49,* 461-474. doi: 10.1080/00313830500267838

Pyc, M.A., & Dunlosky, J. (2010). Toward an understanding of students' allocation of study time: Why do they decide to mass or space their practice? *Memory & Cognition, 38,* 431-440. ***doi***: 10.3758/MC.38.4.431

Robbins, S.B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do Psychosocial and study skill factors predict college outcomes: A meta-analysis. *Psychological Bulletin, 130,* 261-288. doi: *10.1037/0033-2909.130.2.261*

Robins. R.W., & Pals, J.L. (2002) Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self and Identity, 1,* 313-336. doi: 10.1080/15298860290106805.

Roediger, H., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1(3)*, 181-210. doi:10.1111/j.1745-6916.2006.00012.x

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice—common principles in 3 paradigms suggest new concepts for training. *Psychological Science, 3,* 207-214.

Schraw, G., & Dennison, R. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460-475.

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 907-912. doi:10.1037//0278-7393.27.4.907

Smith, B.P., (2005). Goal orientation, implicit theory of ability, and collegiate instrumental music practice. *Psychology of Music, 33,* 36-57. doi: 10.1177/0305735605048013

Stipek, D., & Gralinski, J. H. (1996). Children's beliefs about intelligence and school performance. *Journal of Educational Psychology, 88,* 397–407. doi: *10.1037/0022-0663.88.3.397*

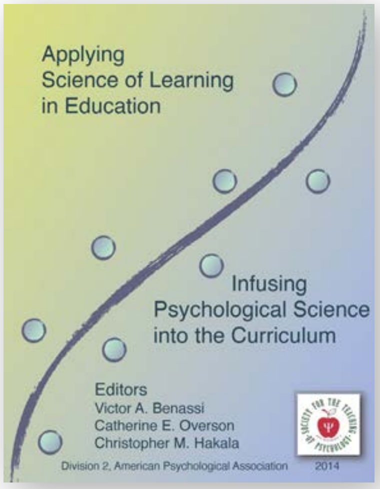
Thiede, K.W., Anderson, C.M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95,* 66-73. doi: 10.1037/0022-0663.95.1.66

Thiede, K. T., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1024-1037. doi: 10.1037/0278-7393.25.4.1024

Vrugt, A., & Oort, F. J. (2008). Metacognition, achievement goals, study strategies and academic achievement: Pathways to achievement. *Metacognition and Learning, 30,* 123–146. Doi: 10.1007/s11409-008-9022-4

Yukselturk, E., & Bulut, S. (2007). Predictors for student success in an online course*. Educational Technology & Society, 10,* 71-83.

Zimmerman, B.J., Bandura, A., & Martinez-Pons, M. (1992). Self-Motivation for Academic Attainment: The role of self-efficacy beliefs and personal goal setting, *American Educational Research Journal, 29,* 663-676. doi: 10.3102/00028312029003663



Metacognition and Instruction

John Girash
Derek Bok Center for Teaching and Learning, Harvard University

Metacognition refers to thinking about, and planning and control of, one's own thinking. While clearly linking both "meta" and "cognition", the dual aspects of *awareness* and *regulation* in this definition can lead to considerable ambiguity in the literature regarding where metacognition fits in relation to other cognitive functions. That metacognition's status as a "fuzzy concept" makes it perhaps better characterized through example than by operational definition (Wellman, 1985; Paris & Winograd, 1990) is consistent in the literature. Despite this fuzziness, metacognitive skills interventions and instruction have been found to be especially effective for improving academic performance of low-performing students, in terms of providing strategies for reading comprehension specifically (Pressly & Gaskins, 2006) as well as across varied subjects in general (Zohar & Ben-David, 2008). Quoting *How People Learn,* "Metacognitive approaches to instruction have been shown to increase the degree to which students will transfer to new situations without the need for explicit prompting" (National Research Council [NRC], 2000, p.67).

History and Overview of Metacognition

Flavell (1976) described metacognition as both the "active monitoring and consequent regulation and orchestration of these [cognitive transaction] processes... usually in the service of some concrete goal or objective" (p. 232). In the 1970s there was still some question among psychologists about the need for metacognition to be considered a distinct phenomenon and area of study — either as one of the then-current "meta's" of *metamemory*, *metalearning*, *meta-attention*, *metacomprehension*, and *metalinguistics*, or as the cognitive process underlying all of the above — or alternatively whether it was an incidental "epiphenomenon" and not fundamental to cognition and learning (Brown, 1978). By the 1980s, metacognition was well accepted as an area of cognitive psychology (Lawson, 1984) especially in the study of reading and language acquisition, but there remained uncertainty regarding the dual nature of metacognition as both awareness and regulation of cognition and how the two aspects relate in both normal and abnormal cognitive development. Metacognition was assumed to exist in adults and to have developed during childhood; research tended to focus as much on young children's learning of a conception of "self" (Wellman, 1985) as on more material or academic learning. However, some pedagogical implications of metacognition were already being considered, both for teaching in general and for the need to explicitly train teachers about discipline-specific metacognitive skills, and this research was being linked to other results from cognitive research such as novice-expert differences (Rohwer & Thomas, 1989).

The 2-fold distinction of metacognition is typically expressed as knowledge (and sometimes personal beliefs) about cognition and understanding, and the regulation or control of cognition and thinking. Knowledge in turn can be parsed into thinking about person, task, or strategy (Vrugt & Oort, 2008) or into declarative, procedural, and conditional thinking (McCormick, 2003). Flavell (1976) included "person variables" in metacognition, but Pintrich, Wolters, and Baxter (2000) identified such self-knowledge as "motivational constructs" distinct from metacognition. The theoretical assumption is that knowledge about cognition and the ability to regulate it are positively correlated, but Vrugt and Oort (2008) noted that personal experience is inter-correlated with both and so this assumption does not apply when the knowledge is "incorrect". This was demonstrated by Barenberg and Dutke (2013) in their study of students anticipating either a pass/fail or a graded test on the accuracy of their metacognitive monitoring of their own learning. As discussed below, Kung and Linder (2007) also found situations in which learners may exhibit metacognitive knowledge and awareness in a group setting, but there is no resulting change in the planning behavior of the group.

Current practice is to consider metacognitive knowledge as consisting of declarative (knowledge of one's abilities and about cognition in general), procedural (how to think and learn efficiently), and conditional (context-specific) aspects (McCormick, 2003; Schraw & Moshman, 1995). Another way to think about metacognitive knowledge is in terms of stability and levels of consciousness. McCormick (2003) considered metacognitive knowledge to be usually explicitly "statable." Zull (2011) defined it to be specifically conscious, however, Schraw and Moshman (1995) equally included knowledge affecting behavior that is not conscious or statable, and Veenman, Van Hout-Wolters and Afflerbach (2006) also consider "automatic" or non-conscious metacognition. Metacognitive control is typically said to involve planning (i.e., choice of approach for a task), monitoring or evaluation of progress, and regulation or revision (e.g., Vrugt & Oort, 2008). McCormick (2003) identified metacognitive control with the concepts of "executive control" or "executive function" from general cognitive models, and indicates there is conflicting evidence in the research regarding whether metacognitive skills can cross disciplinary domains, even if they do seem to be first developed in the context of specific domains (see also Veenman, Van Hout-Wolters, & Afflerbach, 2006).

Empirically, a person's monitoring of predicted learning and future knowledge-recall performance is the most readily observed aspect of metacognition. The types of such monitoring are usually listed as ease of learning ("EoL"), judgment of learning ("JoL"), and feeling of knowing ("FoK"), and these bear directly on other areas of learning research such as comprehension error-detection and -correction and the "illusion of knowing" (Kornell & Bjork, 2008). Typical findings include that better-performing students also have more accurate metacognitive predicting abilities (McCormick 2003). Much of the research is on comparing self-predictions to actual performance in learning tasks (see Ehrlinger, this volume for a review), i.e., metacognition in and of itself, and not as much on how metacognition relates to improvements in learning performance[1]. Two main areas of attention in the study of metacognition and

---

[1] For the instructor interested in measuring students' metacognition in itself, Jacobse and Harskamp (2012) provide a methodology combining predictive judgments, problem visualizations, and postdictive judgments. It features higher validity than student self-reports but without the labor intensiveness of the transcribing and coding of "think-aloud" activities that appear in much of the primary literature, albeit at the expense of not covering as full a range of metacognitive functions.

learning are reading and writing, and mathematics and problem solving. McCormick's (2003) excellent chapter in *Handbook of Psychology* extensively covers writing and reading skills; this chapter is intended to be complementary to it, especially along the problem-solving direction and specifically for the postsecondary-level instructor.

Metacognition's Relationship to Other Aspects of Cognitive Pedagogy

Metacognition was part of the increased emphasis on "Teaching Thinking" (Jones & Idol, 1990, pp. 2-6), which also included self-regulated learning [SRL], and critical thinking. The following sections describe how metacognition relates to these and other aspects of cognition-focused pedagogy.

Novice/Expert Differences

Explicit, flexible metacognitive functioning appears to be a strong component of expert thinking. Experts monitor their own understanding as they work so that they are explicitly aware when their own knowledge limits are reached, and additionally verify and analogize new information within their existing knowledge set in order to adaptively grow their understanding (NRC, 2000). For example, models of writing expertise contrast the "knowledge telling" writing style of novices vs. the "knowledge transforming" mode of experts, each being largely influenced by metacognitive functioning; additionally, in problem solving, appropriate metacognitive functioning can provide the executive control required for "expert" modes of thought and analysis (McCormick 2003).

Critical Thinking

There are striking parallels between metacognition and critical thinking as topics for instruction; metacognitive regulation could even be conceived to be *critical* thinking about one's own thinking. Using a think-aloud protocol, Ku & Ho (2010) found a strong correlation between university students' use of metacognitive planning, monitoring, and evaluating strategies and their performance on critical thinking tasks. This is probably not surprising. If critical thinking is centered around being aware of the larger context of a given piece of knowledge or content and how to take into account that context when thinking analytically about it, then metacognition is similarly about being aware of the context of one's own thinking about the content and how to take *that* into account. Both are fuzzy concepts usually measured indirectly by performance, and both have fostered debates regarding whether to teach them generally or within a content-driven disciplinary context. In terms of teaching, the latter view seems to have prevailed; Willingham (2007) argued that, from an assessment perspective, it is nonsensical to teach critical thinking devoid of content, and Redish (2003) pointed out that in order to learn metacognitive thinking skills, students must think metacognitively about something *other* than metacognitive thinking skills, which may as well be course content.

Executive Function and Self-Regulated Learning

An unsettled question is how metacognitive regulation of cognition (as it is framed by developmental research) relates to one's "executive function" that oversees mental information processing (from the information-processing research perspective): are they identical aspects of the mind (McCormick, 2003)[2], or are they merely related in that it requires a conscious decision to initiate the "meta" way of

---

[2]Note that the dual awareness-and-control nature of metacognition is consistent with information-processing's analogous framework of "declarative" and "procedural" knowledge (Paris & Winograd, 1990).

thinking and the resulting metacognitive thinking can't be "forced" (i.e., controlled) without potentially being counter-productive (Zull, 2011)? To paraphrase a colleague, even if for most advanced physics problems one can define an ideal solution procedure, for some specific problems the best advice to give a student is “put down your pencil and go for a walk” (J. Barandes, personal communication, June 2013).

Pintrich et al. (2000) identified self-regulated learning (SRL) as a way of describing metacognitive control and executive function in the context of learning. However, again considering the information-processing perspective, Gitomer and Glaser (1987) considered SRL to be analogous to metacognitive awareness, not control, and Schraw, Crippen and Hartley (2006) presented SRL as a conceptual umbrella over cognition, metacognition, and motivation. This presents another case of fuzzy concepts overlapping without clear definitional or functional boundaries; for cognitive research it is clear that better linking of the two areas of study is needed (Slavin, Hurley & Chamberlain 2003). For the practicing instructor, however, the salient point may be as stated by Zimmerman and Martinez-Pons (1992):

> Although all learners use regulatory processes to some degree, self-regulated learners are distinguished *metacognitively* by their awareness of strategic relations between regulatory processes and learning outcomes and their use of specific strategies to achieve their academic goals. Thus, self-regulated learning involves the awareness and use of learning strategies, perceptions of *self-efficacy*, and a commitment to academic *goals*. (p. 187).

Given the conflated nature of the research, it is fortunate that *executive function* vs. *metacognitive control* vs. *self-regulated learning* may be a distinction without much of a difference for the instructor, at least at the level of gleaning implications from current research. And unless we are willing to consider “meta-metacognition”, knowing when to put down that pencil and go for a walk is definitively a metacognitive function.

Metacognition and Learning: Empirical Results

A distinction with more relevance may be that metacognitive *knowledge* is not equivalent to metacognitive *awareness*, as the latter refers specifically to skills that a student applies to other learning situations and so is applicable to actual teaching (Pintrich, Wolters & Baxter, 2000). In other words, a student may be able to state metacognitive principles and yet might not employ them in his or her own thinking while learning other content. This can be especially prevalent in a group setting with varied overlapping levels of activity from the purely logistical to high-level conceptual (e.g., a teaching lab), in which it has been observed that students can express numerous metacognitive statements, however, few of them lead directly to changes of behavior (Kung & Linder, 2007). It is important that the metacognitive thoughts not get "lost in the noise" of a hectic learning environment, whether that environment is external or internal to the students' minds.

Metacognitive Awareness and Student Affect

Paulsen and Feldman (2007) surveyed college students across disciplines and found that those holding more-naive epistemological beliefs about learning — e.g., whether learning itself is a learnable skill — exhibited less metacognitive functioning while performing a learning task. In their literature review on student affect and self-efficacy, Pintrich and De Groot (1990, and references therein) concluded that students who see themselves as more capable exhibit higher metacognitive functioning, and, in addition, are more persistent in a learning task than students with less confidence in their ability to perform. They noted that metacognitive strategies cannot be reduced to mere formulaic instructions without losing their effectiveness: research on metacognition and self-regulation "suggests that students must be able to understand not only the 'what' of cognitive strategies, but also how and when to use strategies appropriately" (p. 38). Regarding affect of individual students more specifically, Strain, Azevedob, and D'Mello (2013) found that undergraduates' affective feelings regarding a given task can have a large effect on their learning as well as their metacognitive functioning. Students' prior affect — their expectations and biases when being initially exposed to a new concept or idea — also can greatly influence their later affect, thinking, and metacognitive functioning when performing a learning task around that concept (Redish, 2003); metacognitive instruction can make students less anxious about making mistakes (McInerney, McInerney, & Marsh 1997) and so more able to tolerate errors and confusion as a constructive part of the learning process.

Research on Self-Prediction of Learning and Performance

Adults, including college-age students, are consistently found to be robustly bad predictors of their own learning as later measured by performance on tests (see, for example, Hacker, Bol, Horgan, & Rakow, 2000), but there are interesting subtleties to the phenomenon. Pressley and Ghatala (1990) noted that adults self-assess their own learning reasonably well for reading and recalling of simple texts, but not for difficult ones. For example, university students tended to overestimate their own learning and to study inefficiently, re-reading an entire text rather than focusing on crucial or difficult segments of it, and (similarly to Barenberg and Dutke's 2013 findings) accuracy of self-monitoring was found to depend on the anticipated test: a free recall task induced less overestimation of learning than a fill-in-the-blank task. Miller and Geraci (2011) found that college students almost always anticipate a higher grade than they actually earn on exams, but with lower-performing students having even worse predictive accuracy than high-performing students — consistent with the Dunning-Kruger effect seen in undergraduate exam-takers even after statistical effects such as regression to the mean are taken into account (Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008) — and that providing extra credit for predictive accuracy was found to have no effect unless it was coupled with concrete feedback to students such as strategic tips on how to improve accuracy.

That low-performing students may lack metacognitive awareness skills is a common theme in the predictive-abilities research. For undergraduates in an introductory psychology course, Hacker, Bol, Horgan, and Rakow (2000) found that low-performing students were significantly overconfident in their learning, whereas high-performing students were slightly under-confident. They noted that repeated practice and feedback have been found to increase meta-memory judgments (i.e., accuracy of predicting one's recall ability) in general. However, while the high-performing students increased both their predictive and postdictive accuracy regarding test performance through the course, the low-performing students continued to exhibit "very poor self-assessment and self-evaluation of their knowledge" (p. 168).

In another introductory psychology course, Thiede and Anderson (2003) found that having students write a summary of a text improved their metacomprehension (i.e., analytical, not just recall) accuracy. Delaying the writing assignment helped even more. Much of this research bears directly on pedagogical concepts such as desirable difficulties and the benefits of frequent and delayed testing (Logan, Castel, Haber & Viehman (2008); Rawson & Dunlosky, 2011). In a survey of undergraduates that included the standardized 52-item Metacognitive Awareness Inventory (MAI; Schraw & Dennison, 1994), Coutinho (2007) found that students' metacognition strongly correlated with their having mastery-directed goals (i.e. motivated by learning), and not with performance-centric goals (motivated by grades), and also that while metacognition and mastery goals modestly correlated with academic success (as measured by GPA), performance goals did not. Coutinho concluded that "mastery goals influence performance directly as well as indirectly through metacognition... performance goals do not influence performance directly or through metacognition" (p. 43).

Research on Metacognitive Prompting

The idea of explicitly prompting students regarding metacognitive questions to improve learning and performance has been especially prominent in mathematics. In his pioneering work, Schoenfeld (1985) was interested in helping students avoid wasting excessive amounts of time on unproductive approaches. He had students in a group-problem-solving class be prepared to explicitly question their strategy at any time by posting the following questions in the classroom:

1. *What (exactly) are you doing? (Can you describe it precisely?)*
2. *Why are you doing it? (How does it fit into the solution?)*
3. *How does it help you? (What will you do with the outcome when you get it?)*

Schoenfeld asked students the questions regularly and also modeled use of them in verbalizing his decision-making process while teaching. By the middle of the term, students were asking themselves the questions as well. Schoenfeld transcribed and coded students' conversations into six levels of problem-solving behavior: *reading* and *analyzing* the problem statement, *exploring* possible solution methods, *planning* a solution process, *implementing* it, and *verifying* its correctness. Example plots are shown in Figure 1. By the end of the class, students were spending much more time on the higher-level behaviors (planning and above), and the times in which they expressed metacognitive awareness in the transcripts were more likely to lead to changes in behavior — i.e., metacognitive control — such as additional planning or verification, compared to when the class started.

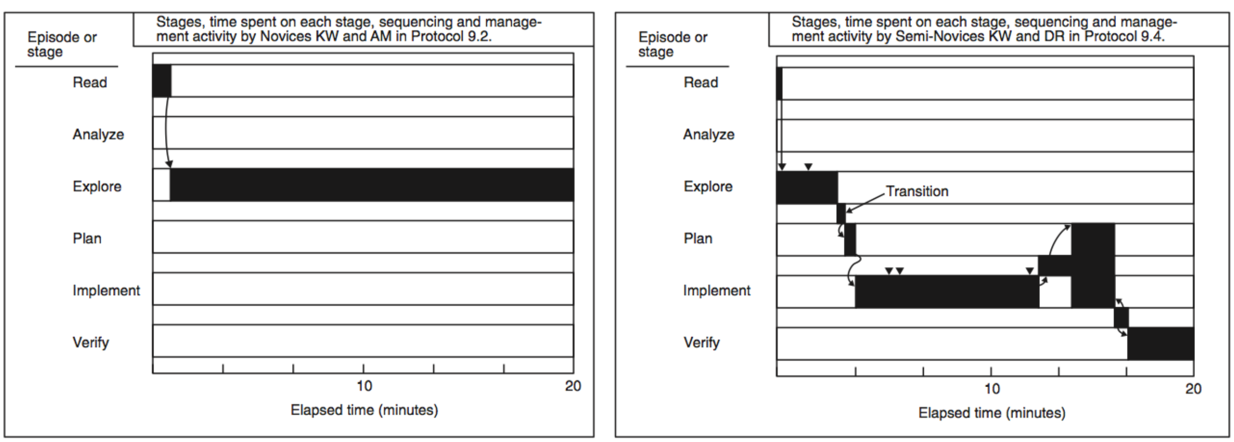


Figure 1. Examples of student-behavior plots from Schoenfeld (1985). At left, a pair of students at the beginning of the term spend the full solution time exploring methods; at right, students later in the term exhibit more episodes of metacognitive awareness (indicated by triangle symbols) and more complex metacognitive behavior. (Plots reproduced with permission of Elsevier/Academic Press).

Although the intervention in Schoenfeld's initial work demonstrated improvement in students' higher-level behaviors, it lacked empirical controls. In an experimental setting, Hoffmana and Spatariu (2008) studied the relationship between metacognitive prompting, problem-solving efficiency and accuracy, and self-efficacy of undergraduates performing the task of mental multiplication of 2-digit numbers. Their prompts were more explicitly metacognitive and fine-grained than Schoenfeld's — intentionally so, so that when used in a real classroom setting students could better be given feedback on their resulting metacognitive functioning — while somewhat embedding Schoenfeld's coding scheme:

1. *Have you solved similar problems before?*
2. *What strategy can you use to solve these problems?*
3. *What steps are you using to solve the problem?*
4. *Can your answer be checked for accuracy?*
5. *Are you sure your answers are correct?*
6. *Can the problem be solved in steps?*
7. *What strategy are you using to solve the problems?*
8. *Is there a faster method to solve the problem?*
9. *Are these problems similar to addition in any way?*

The metacognitive prompting improved students' accuracy and efficiency specifically for more-complex problems (with no effect found for easier problems), and induced more awareness of "unmindful" (unproductive) and alternate (more productive) solution strategies, compared to a control group.

In a classroom setting, Mevarech and Fridkin (2006) performed a controlled experiment in which one group of pre-college math students received traditional instruction while another was taught using the *IMPROVE* method and found that students in the IMPROVE group outperformed those receiving traditional instruction both in mathematical knowledge gained and in reasoning during problem solving. IMPROVE (see Kramarski, 2008 for an extensive description) stands for Introducing new concepts, Metacognitive questioning, Practicing, Reviewing, Obtaining mastery, Verification, and Enrichment and remediation. Kramarski (2008) compared its role in math to that of the Reciprocal Teaching method of Palincsar and Brown (1984) in reading comprehension.[3] IMPROVE had already been found to have a positive effect on various aspects of mathematical reasoning and problem solving in secondary school students; it extends Schoenfeld's efforts by providing a framework for how the instructor should structure and model his or her own metacognitive behavior as well as that of the students.

Research on Teachers' Metacognition

Metacognitive prompting also has beneficial impacts on *teachers'* skills and behavior. Kramarski (2008) studied the effects of IMPROVE training on pre-service elementary math teachers in a professional development program and found that it improved their algebraic reasoning, problem-solving skills and conceptual explanations compared to a group receiving the program's traditional instruction. Teachers in the IMPROVE group also developed more self-monitoring and evaluation skills in algebraic problem solving (although both groups showed a similar improvement in overall cognitive skills, which were already at a high level coming into the program).

---

[3] Specifically, both IMPROVE and Reciprocal Teaching involve self-monitoring of learning. However, IMPROVE is expressly designed to induce metacognitive activity, whereas Reciprocal Teaching is largely cognitive in nature (Rosenshine & Meister, 1994) and focuses on improving students' comprehension while reading via explication and elaboration (NRC, 2000).

Implications for Instruction

Much metacognition research has focused on childhood development. Even so, metacognitive skills, such as evaluating one's own knowledge in a learning task and monitoring one's approach in order to alter it as needed, are equally important in the higher-education setting. Teaching metacognitive skills in higher education, however, poses a two-fold challenge: not being "content," metacognitive skills are not likely to be part of the curriculum, and even the initial metacognitive step of task assessment may not be easy or natural for college-level students (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010). One approach to confront this challenge on both fronts is to engage in teaching methods in which metacognitive strategies are first modeled by the teacher and are then practiced by the students. After discussion and feedback, this method should foster eventual metacognitive self-prompting without teacher intervention (NRC, 2000). Even though for experts in a given academic field metacognitive skills can transfer outside of their disciplinary domain, teaching these skills to non-expert students is more effective when done in the context of specific domain knowledge, especially when students appreciate there are different ways to attain new knowledge and apply it to problem solving (e.g., Fredricks, Blumenfeld, & Paris 2004). Context-specific instruction also leads to better transfer of metacognitive skills than when taught generically, in isolation (NRC, 2000). Further, when *not* taking students' metacognition into account, implementation of other research-based pedagogies, such as active learning strategies, can create environments that are expressly "active" but neglect "learning" (Tanner, 2012).

Veenman, Van Hout-Wolters, and Afflerbach (2006) identified three research-based principles for effective metacognitive instruction: motivating learners regarding the usefulness of metacognitive learning activities so they exert the perceived required initial "extra effort"; teaching metacognition in the context of disciplinary content matter; and ongoing instruction to maintain students' metacognitive practices. As with any pedagogical intervention, teachers must take care not to overload students with complexity that might inhibit implementation of higher-level metacognitive functions, nor to under-load them with simple tasks that are successfully completable without engaging in metacognition at the start. Good pedagogy indicates providing tasks that range in complexity, paced and scaffolded appropriately for the students (McCormick 2003, and references therein).

Explicit Instruction: Metacognitive Prompting

Explicit prompting methods such as those of Schoenfeld (1985) and Hoffmana and Spatariu (2008) mentioned above provide a solid, research-driven place for instructors to start. Although framed specifically for problem solving, these methods are easily generalized to most types of learning task. For example, writing can be seen as essentially a problem-solving task in monitoring and revision so as to create a final text that corresponds to the "intended text" (McCormick, 2003). Explicit metacognitive interventions have been shown to robustly improve students' comprehension of written material as well as their general academic performance, particularly when deliberate practice across different texts and contexts is included (McNamara, 2011). Metacognition can also be effective in process-oriented instructional methods such as inquiry-based learning (Schraw, Crippen, & Hartley 2006). Finally, case-based teaching is inherently problem-focused, and fosters metacognition by putting students vicariously in the subjects' position; cases are authentic, and provide the "next best thing to being there" (Gallucci, 2006).

Schoenfeld's prompting technique in particular brought out the usefulness of students exploring and evaluating different approaches, the effect of students self-assessing their progress, and the importance of modeling, scaffolding, and repetitive and reflective feedback in small- and full-group settings (NRC, 2000). Hoffmana and Spatariu's prompting method, although developed outside the classroom, provides instructors with even more fine-grained tools in all of these areas.

The IMPROVE method provides an even more highly structured framework for instructors, with some indication that its explicit metacognitive guidance can help students with the transfer of reasoning skills outside a specific context (Kramarski, 2008). The instructor models metacognitive self-questioning prior to attempting any direct analysis or solution when introducing new concepts. The four categories of questions include:

1. Comprehension (What is the task about?)
2. Connection (What are the similarities/differences between this task and ones you have performed in the past?)
3. Strategic (What are the strategies/tactics/principles needed to solve the task, and why?)
4. Reflection (What am I doing right now? Why am I doing it? Does the result make sense? Can I solve it differently?)

Note that the reflection questions are intentionally phrased in the first person, it order to set the expectation of ongoing self-assessment. Redish (2003) identified self-reflective questions such as those that check the plausibility of an intermediate or final result and the thinking behind it as being especially key.

Going beyond explicit modeling, instructional practices that foster metacognition and increase students' transfer of learning include focusing on sense-making, defining learning goals, self-assessment, and reflection about what approaches were more or less optimal for a given task (NRC, 2000). Tanner (2012) suggests several metacognitive learning activities:

- *Preassessments* that prompt students to examine their initial thinking — e.g., an initial "What do I already know about this topic that could guide my learning?" self-question along with explicit metacognitive prompts to guide students' planning.
- *Muddiest Point* responses (a specific type of minute paper) to questions such as "What was most confusing to me about the material being explored in class today?" This not only engenders reflection on understanding, it sets the norm that confusion is the beginning of learning and is to be expected.
- *Reflective Journals* for self-monitoring, for example, low-stakes writing assignments on "What about my exam preparation [or other learning task] worked well that I should remember to do next time? What did not work so well that I should not do next time or that I should change?"
- *Integrating Reflection into Credited Course Work* with questions such as "What was most challenging to you about this assignment?" or "What questions arose during your work on this assignment that you had not considered before?" Diagramming or concept mapping (see, for example, Novak & Canas, 2008) can be assigned to similar effect.

Coil, Wenderoth, Cunningham, and Dirks (2010) described the value of simply having undergraduates complete a low-stakes pre-test combining several of the above ideas. When given explicit metacognitive instruction, a pre-test should help students both recognize any lacking skills as well as address the tendency to over-estimate performance.

Metacognitive Instruction for Teachers

The metacognition of teachers *themselves* can be considered from two perspectives: their ability to teach metacognition ("meta-instruction", to coin a term), and their own metacognitive behavior regarding their teaching practice. In general, teachers come into the profession with a "rich pedagogical understanding of metacognition", but this awareness may not be exhibited in their practice, perhaps due to the conflicting pressures of good pedagogy on one hand and mandated curricula or content coverage on the other (Wilson & Bai, 2010). In a science teaching-skills workshop, Zohar (1999) found that in-service teachers were able to *teach* higher-order thinking skills intuitively and procedurally, but were not able to *verbalize* specifically metacognitive aspects of those skills (e.g., declarative metacognitive knowledge) unless they themselves were also explicitly instructed in such aspects. And Tanner (2012) concluded that "cultivating a metacognitive lens toward one's teaching does not appear to automatically or easily transfer" (p. 118) from the deep metacognition that university faculty apply to their own research. He suggested that teachers might pose self-analytical questions such as: "What assumptions do I hold about students? To what extent do I have evidence for those assumptions? Why do I make the instructional decisions that I make? What do I know about teaching? What would I like to learn? What am I confused about?"

It may be that teaching, inherently a social activity in addition to an academic one, calls for unique metacognitive "moves" regarding the practice of being an instructor. Lin, Schwartz, and Hatano (2005) argued that classroom teaching is manifestly fluid, long-timescale and context-dependent, yet successful metacognitive interventions have tended to target short-term problems that have well-defined optimal solution procedures, and thus have focused on intra-problem skills and not on longer-term, inter-task reflection. In addition, individual teachers tend to dismiss examples of novel approaches regardless of evidence of effectiveness if they do not identify with the example's situational context, unless they have been primed with a metacognitive intervention to foster receptiveness to new ways of teaching. To this end, Lin et al. (2005) tested an "adaptive metacognition" technique with pre-service teachers, calling on their different backgrounds to provide multiple perspectives regarding an apparently familiar situation in order to expand their thinking about alternatives and so avoid dismissing a possible teaching moment as "typical" and thus uninteresting. They found that participants in the group exposed to the adaptive technique ended up asking metacognitive questions (e.g., "How/Why" or "If/Then") in their own teaching twice as often as the control group, who more frequently asked only "What" or ""Yes or No"-type questions.

Metacognitive Processes Made Explicit: Instructional Interventions

In many classroom studies metacognitive training coupled with peer- or cooperative-learning techniques has been shown to be especially effective for increasing younger students' flexible use of mathematical and writing knowledge (for examples see Kramarski & Mevarech, 2003, or Yarrow & Topping, 2001).

At the university level, two computer science courses provided examples of successful metacognitive interventions in cooperative or collaborative environments. McInerney, McInerney, and Marsh (1997) ran a controlled study in their mandatory *Introduction to Computers* course, dividing students randomly into traditional direct-instruction (control group) and metacognitive/cooperative (metacognitive group) sections. In addition to the structured class work, the metacognitive group was prompted with general self-regulatory questions including:

1. What did I learn this week in my computing class?
2. With what did I have difficulty this week?

3. What types of things can I do to deal with this difficulty?
4. What specific action(s) am I going to take this week to solve any difficulties?

Although the two groups performed equally well on a practical test, the metacognitive/cooperative group significantly outperformed the traditional group in more conceptual assessments such as a research report — but only when the metacognitive intervention was introduced from the *beginning* of instruction and not only towards the end. The gains were largest for those students in the cooperative group who had the lowest initial self-ratings of computer competency.[4]

A more holistic course-based intervention study was Volet's (1991) introductory programming course, which featured two independently-chosen sections. Students in the control section had the option of making use of tutoring sessions and opportunities for group work, and were encouraged to write algorithms by hand before starting to code (even though few did so), but students in the experimental section were expressly required to do all those things and also participated in a peer support network that included coaching in making problem-solving processes explicit. Again, students in the control section learned as much factual knowledge as did those in the experimental section (as measured on an exam), but the experimental group did significantly better in applying knowledge to new problems. Even more tellingly, a greater proportion of students from the experimental section passed the *next* advanced course compared to the control students, and with better grades.

Metacognitive Processes Made Explicit: Constructing Student-learning Groups

Even without a specifically metacognitive intervention on the instructor's part, small-group work has many advantages in making metacognitive functioning explicit in the context of interpersonal group interactions. Group (and peer) work manifestly involves students comparing different strategies and solution methods — even bringing to light the *possibility* of alternative strategies — and requires students to continuously verbalize their thinking and thus to subject it to more explicit checks of comprehension (Garfield, 1993). Following are two specific instructional techniques illustrating the functionality of metacognition in a group setting.

In an overview of best practices for setting up task-based student groups, Sarkisian (1997) provided several prompts that essentially build metacognitive awareness and regulation of group process and progress into the procedural structure, such as:

- *Can discussions be managed differently so all can participate? Are people listening to each other and allowing for different kinds of contributions?*
- *Are all members accomplishing the work expected of them? Is there anything group members can do to help those experiencing difficulty?*
- *Is outside help needed to solve any problems?*
- *Who has dealt with this kind of problem before?*

---

[4] McInerney et al. (1997) also measured students' anxiety levels towards computer learning and their own compentence with computers and found that for students with the initially highest anxiety, levels dropped more for students in the direct-instruction group than for those in the cooperative one despite the greater learning gains exhibited by the cooperative group. For students with low levels of initial anxiety, levels in the cooperative group remained lower than those in the direct-instruction group.

- *What are the pluses of that approach? The minuses?*
- *We have two basic choices. Let's brainstorm. First let's look at the advantages of the first choice, then the disadvantages.*
- *Let's try ranking these ideas in priority order.*

Sarkisian's list addresses some of the more common issues regarding group-metacognitive functioning: *floundering, digressions and tangents, getting stuck*, and *rushing to work* before sufficient planning has been accomplished. Giving students experience in dealing with issues of awareness and regulation in the group setting, and linking it to their own personal cognitive learning habits is one way to build robust metacognitive training into students' experiences without taking “extra time” away from other instruction.

In their specific instructional technique “Guided Reciprocal Peer Questioning", the National Institute for Science Education (NISE, 1997) recommended giving students several minutes to compose content-specific *questions* for their group and provided many prompts to guide thinking, several of which again manifestly lead to metacognitive thinking, including:

- How does _______ relate to what I've learned before?
- What is another way to look at _______ ?
- What is a new example of _______ ?
- What would happen if _______ ?
- What are the implications of _______ ?
- Why is _______ important?

Later in the group-work process, NISE (1997) suggested “Think-Pair-Square”, a variation on the traditional think-pair-share method in which after first discussing strategies as a pair, two pairs of students combine so that they can compare the different approaches they came up with.

Looking beyond prompts, elements that make for effective cooperative learning include: requiring groups to report to the full class about their confusions and differing opinions; requiring all group members to participate and to be open to, and *constructively* critical of, differing approaches and views; teaching students how to ask each other clarification questions; and providing specific places for groups to discuss the group’s functioning, what is working well and what can be improved (Tanner, Chatman, & Allen, 2003). The literature on assigning roles within groups is in broad agreement regarding assigning members with useful functions (Garfield, 1993; Sarkisian, 1997; Tanner et al., 2003; Center for Faculty Excellence, U. of North Carolina, 2006), all of which can be seen to map onto an individual’s metacognitive functioning:

- a facilitator to guide discussion;
- a timekeeper/moderator to stay on task;
- a coordinator to seek out needed information;
- a seeker to challenge the group to explore other approaches;

- a reporter to check that all individual and group tasks are being done and to record/report on them;
- a checker/clarifier to verify group members' understanding of the task, intermediate concepts, and each other's contributions;
- an elaborator to connect with prior learning;
- a summarizer to bring together the group's progress towards the task at critical junctures; and
- an observer/questioner to examine what can be learned from the group's functioning.

Some of the roles may be combined or rotated depending on group size, task, and timeframe — several of them are only relevant at specific points in the group process — but it is crucial to note that the roles are *procedural* in order to make for effective group functioning, and are not based on individual student skill or intellect.

Conclusion

Although this chapter deals with metacognitive awareness and functioning on a technical and skill-based level, as a concept, metacognition gets to the very core of our — and our students' — mindful existence as learners and actors in a tangible world. Further, Zull (2011) stated that personal discovery, including the learning of new knowledge and skills, directly leads to a feeling of joy, and it is expressly metacognitive to be aware of having learned and thus to experience the joy. Whether our fundamental instructional goals are based in our students' personal discovery or in a more applied aspect of their newly-gained knowledge, being part of a student's joy of learning will always be a primary (and highly metacognitive) motivation for the teacher as well.

References

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). How do students become self-directed learners? *How learning works: Seven research-based principles for smart teaching* (pp. 188-214). San Francisco: Jossey-Bass.

Barenberg, J., & Dutke, S. (2013). Metacognitive monitoring in university classes: anticipating a graded vs. a pass-fail test affects monitoring accuracy. *Metacognition and Learning, 8*, 121-143. doi:10.1007/s11409-013-9098-3

Brown, A. L. (1978). Knowing when, where, and how to remember: a problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1, pp. 77-165). Hillsdale, NJ: Erlbaum.

Center for Faculty Excellence, University of North Carolina (2006). *Student Learning Groups.* Retrieved from http://cfe.unc.edu/publications/fyc23.html

Coil, D., Wenderoth, M. P., Cunningham, M., & Dirks, C. (2010). Teaching the process of science: faculty perceptions and an effective methodology. *CBE Life Sciences Education, 9*, 524-535. Retrieved from http://ww.lifescied.org/content/9/4/524.full doi:10.1187/cbe.10-01-0005

Coutinho, S. A. (2007). The relationship between goals, metacognition, and academic success. *Educate, 7*(1), 39-47. Retrieved from http://educatejournal.org/index.php/educate/article/view/116/134 doi:10.1348/000709901158514

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes 105*, 98-121. doi:10.1016/j.obhdp.2007.05.002

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-235). Oxford: Erlbaum

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59-109. doi:10.3102/00346543074001059

Gallucci, K. (2006). Learning concepts with cases. *Journal of College Science Teaching, 36*, 16-20.

Garfield, J. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education, 1*(1), 1-9. Retrieved from http://amstat.org/publications/jse/v1n1/garfield.html

Gitomer, D. H., & Glaser, R. (1987). If you don't know it works on it: Knowledge, self-regulation and instruction. In R. E. Snow & M. Farr (Eds.), *Aptitude, learning and instruction: Cognitive and affective process analyses* (pp. 301-325). Hillsdale, NJ: Erlbaum.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*, 160-170. doi:10.1037/0022-0663.92.1.160

Hoffmana, B., & Spatariu, A. (2008). The influence of self-efficacy and metacognitive prompting on math problem-solving efficiency. *Contemporary Educational Psychology, 33*, 875-893. doi:10.1016/j.cedpsych.2007.07.002

Jacobse, A. E., & Harskamp, E. G. (2012). Towards efficient measurement of metacognition in mathematical problem solving. *Metacognition and Learning, 7*, pp. 133-149. doi:10.1007/s11409-012-9088-x

Jones, B.F., & Idol, L. (1990). *Dimensions of thinking and cognitive instruction*. Erlbaum: Hillsdale, NJ.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: is spacing the "enemy of induction?" *Psychological Science, 19*(6), 585-592. doi:10.1111/j.1467-9280.2008.02127.x

Kramarski, B. (2008). Promoting teachers' algebraic reasoning and self-regulation with metacognitive guidance. *Metacognition and Learning, 3*(2), 83-99. doi:10.1007/s11409-008-9020-6

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: the effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40*(1), 281-310. doi:10.3102/00028312040001281

Ku, K. Y. L., & Ho, I. T. (2010). Metacognitive strategies that enhance critical thinking. *Metacognition and Learning, 5*(3), 251-267. doi:10.1007/s11409-010-9060-6

Kung, R. L., & Linder, C. (2007). Metacognitive activity in the physics student laboratory: is increased metacognition necessarily better? *Metacognition and Learning, 2*, 41-56. doi:10.1007/s11409-007-9006-9

Lawson, M. J. (1984). Being executive about metacognition. In J. R. Kirby (Ed.), *Cognitive strategies and educational performance* (pp. 89-109). Orlando: Academic Press.

Lin, X., Schwartz, D. L., & Hatano, G. (2005). Toward teachers' adaptive metacognition. *Educational Psychologist, 40*, 245-255. doi:10.1207/s15326985ep4004_6

Logan, J. A., Castel, A. D., Haber, S., & Viehman, E. J. (2012). Metacognition and the spacing effect: the role of repetition, feedback, and instruction on judgments of learning for massed and spaced rehearsal. *Metacognition and Learning, 7*, 175-195. doi:10.1007/s11409-012-9090-3

McCormick, C. B. (2003). Metacognition and learning. In I. B. Weiner (Ed.), *Handbook of psychology* (Two, pp. 79-102). Wiley Online. doi:10.1002/0471264385.wei0705

McInerney, V., McInerney, D. M., & Marsh, H. W. (1997). Effects of metacognitive strategy training within a cooperative group learning context on computer achievement and anxiety: an aptitude-treatment interaction study. *Journal of Educational Psychology, 89*, 686-695. doi:10.1037/0022-0663.89.4.686

McNamara, D. S. (2011). Measuring deep, reflective comprehension and learning strategies: challenges and successes. *Metacognition and Learning, 6*, 195-203. doi:10.1007/s11409-011-9082-8

Mevarech, Z., & Fridkin, S. (2006). The effects of IMPROVE on mathematical knowledge, mathematical reasoning and meta-cognition. *Metacognition and Learning, 1*, 85-97. doi:10.1007/s11409-006-6584-x

Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: the influence of incentives and feedback on exam predictions. *Metacognition and Learning, 6*, 303-314. doi:10.1007/s11409-011-9083-7

National Institute for Science Education (1997). *Doing CL.* Retrieved from the U. of Wisconsin-Madison, Wisconsin Center for Education Research website: http://wcer.wisc.edu/archive/cl1/cl/doingcl/DCL1.asp

National Research Council, Committee on Developments in the Science of Learning with additional material from the Committee on Learning Research and Educational Practice (2000). J. D. Bransford, A. L. Brown, & R. R. Cocking (Eds.), *How people learn: Brain, mind, experience, and school: Expanded edition.* Washington, DC: National Academies. Retrieved from http://nap.edu/openbook.php?record_id=9853

Novak, J. D., & Canas, A. J. (2008). *The theory underlying concept maps and how to construct them*, (IHMC Technical Report CmapTools 2006-01 Rev 01-2008). Pensacola, FL: Florida Institute for Human and Machine Cognition. Retrieved from http://cmap.ihmc.us/publications/researchpapers/theorycmaps/theoryunderlyingconceptmaps.htm

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*, 117-175. doi: 10.1207/s1532690xci0102_1

Paris, S. G., & Winograd, P. (1990). How metacognition can promote academic learning and instruction. In B. F. Jones, & L. Idol, (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 15-51). Hillsdale, NJ: Erlbaum.

Paulsen, M. B., & Feldman, K. A. (2007). The conditional and interaction effects of epistemological beliefs on the self-regulated learning of college students: cognitive and behavioral strategies. *Research in Higher Education, 48*, 353-401. doi:10.1007/s11162-006-9029-0

Pintrich, P. R., & De Groot, E. V. (1990). Motivated and self-regulated learning components of academic performance. *Journal of Educational Psychology, 82*, 33-40. doi:10.1037/0022-0663.82.1.33

Pintrich, P. R., Wolters, C., & Baxter, G. (2000). Assessing metacognition and self-regulated learning. In G. Schraw, & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43-97). Lincoln, NE: Buros Institute of Mental Measurements.

Pressley, M., & Gaskins. I. W. (2006). Metacognitively competent reading comprehension is constructively responsive reading: how can such reading be developed in students? *Metacognition and Learning, 1*, 99-113. doi:10.1007/s11409-006-7263-7

Pressley, M., & Ghatala, E. S. (1990). Self-regulated learning: monitoring learning from text. *Educational Psychologist, 25*, 19-33. doi:10.1207/s15326985ep2501_3

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: how much is enough? *Journal of Experimental Psychology: General, 140*, 283-302. doi:10.1037/a0023956

Redish, E. F. (2003). There's more than content to a physics course: the hidden curriculum. *Teaching Physics with the Physics Suite* (pp. 51-68). Hoboken, NJ: Wiley. Retrieved from http://www2.physics.umd.edu/~redish/Book/

Rohwer, W.D. Jr., & Thomas, J.W. (1989). Domain specific knowledge, metacognition, and the promise of instructional reform. In C. McCormick (Ed.), *Cognitive strategy research: from basic research to educational applications* (pp. 104-132). New York: Springer-Verlag.

Rosenshine B., & Meister, C. (1994). Reciprocal teaching: a review of the research. *Review of Educational Research, 64*, 479-530. doi:10.1016/0959-4752(91)90012-W

Sarkisian, E. (1997). *Working in groups: a note to faculty and a quick guide for students.* Retrieved from Harvard University, Derek Bok Center for Teaching and Learning: http://bokcenter.harvard.edu/fs/html/icb.topic58474/wigintro.html

Schoenfeld, A. H. (1985). *Mathematical problem solving.* Orlando, FL: Academic Press.

Schraw, G., & Dennison, R. S. (1994) Assessing metacognitive awareness. *Contemporary Educational Psychology* 19*,* 460-475.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351-371. doi:10.1007/BF02212307

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111-139. doi:10.1007/s11165-005-3917-8

Slavin, R. E., Hurley, E. A. and Chamberlain, A. (2003). Cooperative learning and achievement: theory and research. In I. B. Weiner (Ed.), *Handbook of psychology* (Three, pp. 177-198). Wiley Online. 10.1002/0471264385.wei0709

Strain, A. C., Azevedob, R., & D'Mello, S. K. (2013). Using a false biofeedback methodology to explore relationships between learners' affect, metacognition, and performance. *Contemporary Educational Psychology, 38*, 22-39. doi:10.1016/j.cedpsych.2012.08.001

Tanner, K. D. (2012). Promoting student metacognition. *CBE Life Sciences Education, 11*, 113-120. doi:10.1187/cbe.12-03-0033 Retrieved from http://www.lifescied.org/content/11/2/113.full

Tanner, K., Chatman, L. S., & Allen, D. (2003). Approaches to cell biology teaching: cooperative learning in the science classroom — beyond students working in groups. *Cell Biology Education, 2*, 1-5.

doi:10.1187/cbe.03-03-0010 Retrieved from https://ncbi.nlm.nih.gov/pmc/articles/PMC152788/

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28*, 129-160. doi:10.1016/S0361-476X(02)00011-5

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning, 1*, 3-14. doi:10.1007/s11409-006-6893-0

Volet, S. E. (1991). Modelling and coaching of relevant metacognitive strategies for enhancing university students' learning. *Learning and Instruction, 1*, 319-336. doi:10.1016/0959-4752(91)90012-W

Vrugt, A., & Oort, F. J. (2008). Metacognition, achievement goals, study strategies and academic achievement: pathways to achievement. *Metacognition and Learning, 3*(2), 123-146. doi:10.1007/s11409-008-9022-4

Wellman, H.M. (1985). The origins of metacognition. In D. L. Forrest-Pressley, G. E. McKinnon, & T. G. Waller (Eds.), *Metacognition, cognition and human performance* (pp. 1-31). Orlando: Academic Press.

Willingham, D. T. (2007). Critical thinking: Why is it so hard to teach? *American Educator, 31*(2), 8-19. Retrieved from http://aft.org/pdfs/americaneducator/summer2007/Crit_Thinking.pdf

Wilson, N. S., & Bai, H. (2010). The relationships and impact of teachers' metacognitive knowledge and pedagogical understandings of metacognition. *Metacognition and Learning, 5*, 269-288. doi:10.1007/s11409-010-9062-4

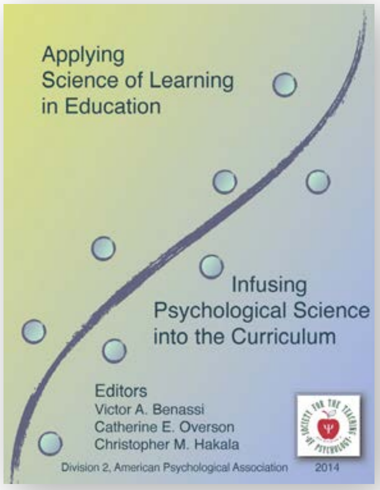
Yarrow, F., & Topping, K. J. (2001). Collaborative writing: The effects of metacognitive prompting and structured peer interaction. *British Journal of Educational Psychology, 71*, 261-282.

Zimmerman, B. J., & Martinez-Pons, M. (1992). Perceptions of efficacy and strategy use in the self-regulation of learning. In D. H. Schunk & J. L. Meece (Eds.), *Student perceptions in the classroom* (pp. 185-207). Hillsdale, NJ: Erlbaum.

Zohar, A. (1999). Teachers' metacognitive knowledge and the instruction of higher order thinking. *Teaching and Teacher Education 15*, 413-429. doi:10.1016/S0742-051X(98)00063-8

Zohar, A., & Ben-David, A. (2008). Explicit teaching of meta-strategic knowledge in authentic classroom situations. *Metacognition and Learning, 3*, 59-82. doi:10.1007/s11409-007-9019-4

Zull, J. E. (2011). The connecting thread: metacognition and the integrated mind. *From brain to mind: Using neuroscience to guide change in education* (pp. 257-284). Sterling, VA: Stylus.



Operation ARA: A Serious Game that Combines Intelligent Tutoring and Learning Principles to Teach Science

Keith Millis
Northern Illinois University

Arthur C. Graesser
University of Memphis

Diane F. Halpern
Claremont McKenna College

Consider the following conversation between two students and a teacher. The topic was a multiple-choice question on a test in a psychology research methods course. The question asked about the definition of instrumentation, which occurs when measurement devices (machines, people) become less accurate over time. Both students got the answer wrong, which led the teacher to hold a brief discussion. Let's listen in on their conversation:

> **Teacher**: What is meant by instrumentation?
>
> **Student 1**: You use instruments to measure what you want.
>
> **Teacher**: No. That's not really what is meant by instrumentation. Can you try again?
>
> **Student 1:** It means there is a problem.
>
> **Student 2:** I do not understand.
>
> **Teacher**: This might help. As you think about the meaning of the term instrumentation, consider how research equipment is sometimes used for long periods of time.
>
> **Student 1:** Right, there is a problem with the measurement.
>
> **Student 2**: I do not understand.
>
> **Teacher:** All right. Try this. The idea with instrumentation is that over time a piece of research equipment becomes less what?
>
> **Student 1**: Accurate
>
> **Teacher**: Good job! It becomes less accurate.

Student 2: I see now. Instrumentation refers to errors that can occur over time as an instrument becomes less accurate. For example, a blood pressure cuff could become less accurate if it is not re-calibrated after being used so many times.

The conversation has a number of principles of learning. Some of these are obvious, others more subtle. For example, 1) the teacher corrects a misconception regarding instrumentation, an important principle of feedback; 2) the teacher tries to elicit the correct response from the students by providing hints, a principle of promoting student generation of information; and 3) a once-bewildered student gains understanding after having received more specific and discriminating content.

What is particularly noteworthy is that the dialog above did not occur in a classroom among three living human beings. Instead, it occurred between one live human (student 1) and two computerized entities called animated pedagogical agents, sometimes referred to as "talking heads". Pedagogical agents are animated virtual entities (e.g., people, insects, inanimate objects) that instruct in digital learning environments by talking, gesturing, showing emotions and achieving different levels of understanding (Johnson, Rickel, & Lester, 2000; Moreno, 2005). Animated agents may inhabit different roles, such as a teacher and student (Baylor & Kim, 2005).

This "trialog" (what we refer to as a 3-way tutorial dialog; Cai, et al., 2011) occurred when a college student was playing the web-based game "Operation ARA." Operation ARA is a serious game developed by the authors, and is currently owned by Pearson Education (note that Operation ARA was previously named Operation ARIES; Millis, Forsyth, Butler, Wallace, Graesser, & Halpern, 2011). The game teaches aspects of the scientific method to high school and college students. Serious games are games that help students learn, contrasted with games that primarily focus on providing entertainment (Ritterfeld, Cody, & Vorderer, 2009). Educational games have increased in popularity throughout the last decade, being adopted by private enterprise, the military, and educators in a number of disciplines (Annetta, 2008). The "ARA" in the game's title is an acronym for **A**cquiring **R**esearch **A**cumen. Operation ARA is unique in that it is a serious game that employs an intelligent tutoring system (ITS) called AutoTutor in its architecture (Graesser, D'Mello, Hu, Cai, Olney, & Morgan, 2012; Graesser, Lu, Jackson, Mitchell, Ventura, Olney, & Louwerse, 2004; Graesser, Person, & Harter, 2001). ITSs are computerized learning environments that teach content and skills, often by mimicking the processes exhibited by human tutors (Woolf, 2009).

In this chapter, we describe the game, AutoTutor, tactics in how the game implements some learning principles and major research findings that support those principles. We close by offering advice to teachers who might consider including serious games into their pedagogy.

Overview of Operation ARA

The goal of Operation ARA is to teach scientific enquiry skills, and, in particular, the ability to evaluate research summaries. Understanding proper research methodology is not only important to science students learning their craft, but also to the general public who are exposed to scientific summaries in magazines, newspapers, TV and the Internet. For example, a new study may report the health benefits of a certain food. However, sometimes studies may be done poorly (e.g., poor experimental design), or they may be well-executed but misinterpreted by the author (journalist, blogger) of the write-up. We believe that a person who can properly critique science found in the popular press is both inoculated against believing unfounded interpretations, and accepting of valid conclusion(s) when warranted.

Embedded in Operation ARA is a storyline that is carried throughout the game. At the beginning of the game, the player is told via a video that the “Federal Bureau of Science” (FBS) has alien creatures called Fuaths in captivity. The Fuaths, it seems, are disguised as Human Beings and are covertly publishing bad science in hopes of “dumbing down” Earth’s occupants and turning us into mindless consumers. In addition, there are mysterious events happening across the world, such as the disappearance of oil tankers and large bodies of water, which escalate over the course of the game. The FBS has launched “Operation ARA” in an effort to find and stop the alien intrusion. The goal of the secret operation is for citizens to become FBS agents by teaching them the scientific method. By knowing the scientific method, FBS agents will be able to identify and arrest Fuaths who are authoring the bad research.

There are three main modules (or levels) in the game. In the first, called Basic Training, the aspiring FBS agent (player) reads an ebook about the scientific method. It is essentially a brief research methods book. To fit into the storyline, the player is told that the ebook was written by Fuaths for their Fuath spies. The FBS had obtained and translated a copy of the book. The book describes aspects of good science so that their operatives can violate them and write poor science for human consumption (to “dumb them down”). The ebook is segmented into brief lessons, each of which describes a main topic. Main topics include the need for control groups, random assignment, accuracy and sensitivity of the dependent variable, objective scoring, experimenter bias, sample size, sample selection, participant bias, mortality/attrition, replication, and confusing causality and correlation. Each lesson includes multiple choice questions for assessment of learning, and the questions are accompanied by brief tutorial trialogs (Figure 1 presents a screen shot of Basic Training). The trialogs include the player, Dr. Quinn (the teacher) and Glass Tealman (fellow student). The trialog that began this chapter was an example tutorial trialog taken from the game. We will discuss the tutorial dialogs and trialogs more deeply below.

Figure 1: Basic Training

Whereas the purpose of Basic Training is to provide the player with the basic concepts of scientific thinking, the goal of the second module, called Proving Ground, is to give the player practice in identifying flaws in research descriptions. Figure 2 presents a screen shot of Proving Ground. The research descriptions are written like newspaper and magazine articles, and cover topics in psychology, biology, and chemistry. For example, one article described a new procedure to teach dancing, and in another, an experiment testing whether a new plastic is biodegradable in landfills. The former study suffered from no control group, a small sample size, experimenter bias and poor sample selection; whereas the latter study used a dependent variable that lacked sensitivity. The player competes against a snooty and irritating pedagogical agent named Tracy, who is a fellow student. (Glass Tealman is absent from this module because he is looking for his lost brother, which is part of the story line). Tracy and the human player take turns identifying flaws, and they earn and lose points as they produce correct and incorrect answers. Besides Tracy, there are two other pedagogical agents in Proving Ground: Dr. Quinn and Broth, who is a Fuath defector. Both of these agents serve as teaching agents, helping the player to correctly identify flaws. Similar to Basic Training, the players receive brief tutorial dialogs to help their performance.

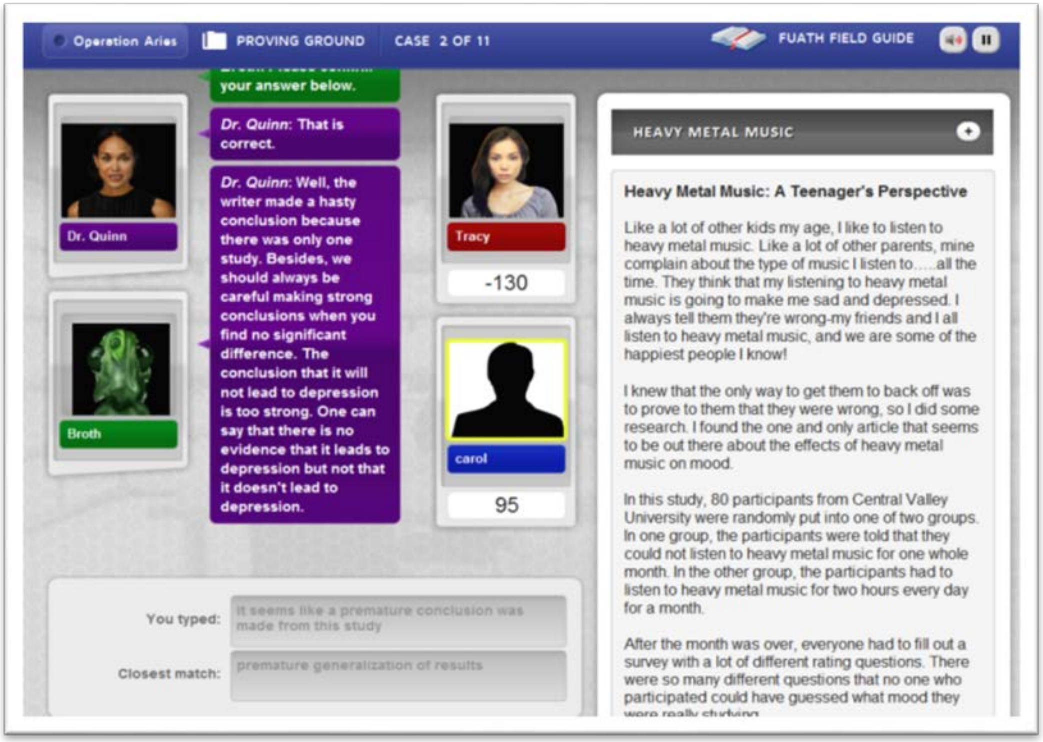


Figure 2: Proving Ground

The third and last module is called Going Active (see Figure 3 for a screen shot). By the time players reach this module, they have graduated and are now “in the field” as special agents of the FBS. The story line grows more intense. The world authorities have conducted a raid in which a number of suspected Fuaths have been captured. Unfortunately, it is thought that human scientists have also been arrested, a consequence of the Fuaths’ ability to look and act human. It is the player’s job in this module to interview each suspect regarding their research to determine if it is flawed and whether the scientist follows and understands the scientific method. If there is a flaw or a lack of understanding of the scientific method, then the suspect should be deemed an alien by the player, and if not, the suspect should be deemed to be human by the player and set free. In this module, the player types in questions to be asked of the suspect to Scott, who is an animated agent playing the role of interrogator. Scott interprets the question passed to him by the player and asks the suspect the question. For example, if the player types “Was there a comparison group?” Scott may ask the suspect “I have been meaning to ask you. Was there a control group?” After the suspect answers, Dr. Quinn who is also overseeing the interrogation may pipe in with a clarifying comment. For example, if the suspect answered that there was no control group, Dr. Quinn may say “This is a correlational study, and these usually do not contain control groups. So, I would say “no flaw.” One-half of the suspects reveal flaws if asked the “right” questions during interrogation (i.e., they are aliens) whereas one-half do not (i.e., they are humans).

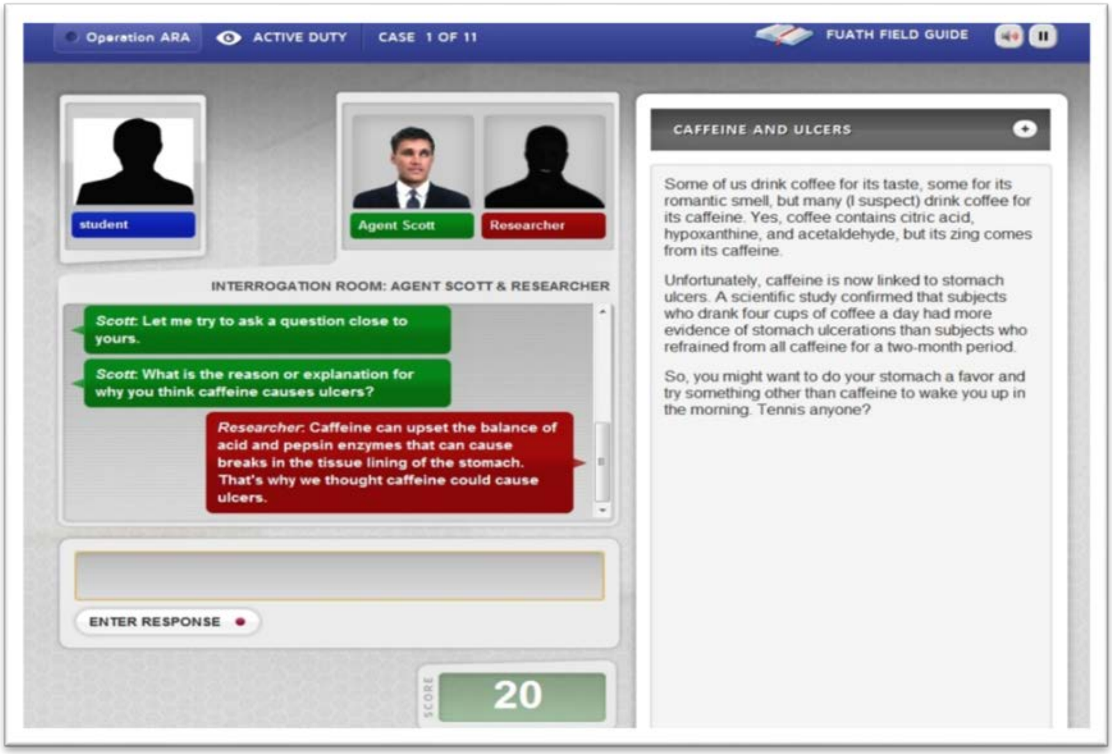


Figure 3: Active Duty

In sum, Operation ARA begins with teaching the declarative knowledge (encyclopedic knowledge) of main concepts in scientific enquiry in Basic Training. The player then receives guided practice in identifying flawed instances of the concepts in Proving Ground. Lastly, the player learns to distinguish flawed from unflawed research by actively asking questions. These activities occur across the backdrop of a story line in which the player helps to save the world from disaster.

AutoTutor: The Intelligent Tutor Behind Operation ARA

The goal of AutoTutor, as with other ITSs, is to teach the content and skills of a targeted domain to the students (Sleeman & Brown, 1982; Woolf, 2009). It presents problems and questions to the user, and evaluates the quality of the answers generated by the user. When appropriate, the program will give corrective and elaborative feedback. One feature of AutoTutor that is present in many but not all ITSs is that the majority of the pedagogical activities occur through tutorial dialogs using natural language. AutoTutor uses animated pedagogical agents to give content aurally (along with other digital media, such as text, animations and movies) via dialogs and monologues. In this way, it mimics real tutors who pose problems and questions to their clients, giving corrective feedback (whether the answer is correct or not), elaborative feedback (hints and brief explanations to help the tutee form a more complete or clearer answer). People are accustomed to having conversations, so, an ITS, such as AutoTutor, that capitalizes on conversations may be appealing due to its natural conversational flow. Another advantage, more relevant to this chapter, is that people learn from tutorial conversations (Graesser, Person, & Magliano, 1995; Graesser, D'Mello, & Cade, 2011). Central to these tutorial conversations is learner interactivity. Interactivity in this regard refers to the extent that both tutor and tutee contribute to the answer in an interdependent fashion. Interactivity is low if the tutor merely gives a lecture while the tutee listens. Interactivity is higher if the tutee answers and asks questions and makes contributions that fit in the dialog. Of course, people may learn from the content of the conversation as they do from reading a text or listening to a lecture, but the interactivity in a dialog appears to have its own effect (VanLehn, Graesser, Jackson, Jordan, Olney, & Rose, 2007). Studies have shown positive correlations between the proportion of words produced by the tutee and learning gains (Katz, Allbritton, & Connelly, 2003). In a series of seven experiments that compared tutorial dialogs with reading text which were largely equated on content, VanLehn et al., (2007) report advantages for the tutorial dialogs which encouraged interactivity. Interestingly, the effect occurred when the material was slightly more difficult than what was known by the tutee. For example, novices learned more when the material was geared toward an intermediate level of knowledge than when it was geared at a novice level. The authors pointed out that this finding was consistent with the zone of proximal development which we discuss later in this chapter.

Why do people learn from engaging in tutorial conversations? Engaging in conversations requires several processes that are associated with learning. These include keeping track of content, answering and asking questions, monitoring understanding, reasoning, explaining, making inferences, retrieving information from long-term memory. All of these processes require the person to be actively engaged, co-constructing a coherent mental representation from the contributions made in the conversation. Because ideas are actively constructed, linked and repeated in working memory, they are learned and better remembered than if they were passively received.

The structure of the dialogs in AutoTutor is very similar to dialogs in real human tutors. The structures are based upon the work of Graesser and Person (1994; Graesser, Person, & Magliano, 1995) who systematically analyzed dialogue patterns of typical human tutors in middle school and in college. What they and others have found is a list of what tutors have in mind as they begin a tutoring session as well as a series of steps that frame the session. First, tutors or their clients typically have questions or problems that they will have the tutee solve. Second, the tutor will have a set of anticipated good answers for each question. These are called *expectations*. Tutors will also have incorrect assertions in mind called *misconceptions*. Depending on the problem, there may be a dozen or more expectations and a few misconceptions. Larger and complicated problems will have more expectations than smaller and simpler problems. The dialogs in Operation ARA fall more into the latter camp. The problem about instrumentation that produced the trialog in the beginning of the chapter (“What is meant by instrumentation?”) only had one expectation: “instrumentation is when a measurement device becomes less accurate with time” and one misconception (“instrumentation is when instruments and measures are used”). In contrast, a physics problem may have a dozen expectations and several misconceptions.

Based on the work on real tutors (Graesser, Person, & Magliano, 1995), AutoTutor employs a 5-step tutoring frame: 1) Present a question to the student. In Basic Training, the questions are based on the multiple choice questions which appear in each lesson. These involve the definition of the concept (“What is attrition?”), the importance of the concept (“Why is attrition important to science?”), and why a particular answer was either incorrect or correct (“Why was the answer about the video tape the correct answer?”). In the Proving Ground module, 1) the question for each summary is: “What are the flaws in this study?”; 2) The student gives an initial answer, such as the definition or flaws that the student sees; 3) The tutor gives brief feedback on the initial answer. In the example presented at the start of this paper, the tutor responds with “no” because the initial answer was incorrect and also conveyed a misconception; 4) The tutor attempts to get the student to express the expectation across a number of turns. In our example, this began with the tutor giving a “pump” (“Can you try again”, “What else?”) to get the student to provide information. AutoTutor is constantly assessing whether the content typed in matches the current expectation using algorithms from computational linguistics and artificial intelligence; 5) If it does not meet the threshold of an acceptable answer, then the tutor will use a "hint -> prompt -> assertion" cycle. A hint is a general assertion that tutors use to nudge the student to express the expectation. In the trialog that began this chapter, the hint was "As you think about the meaning of the term instrumentation, consider how research equipment is sometimes used for long periods of time." If the hint is unsuccessful in eliciting the expectation, then a prompt is given. A prompt is an open-ended question which attempts to get the student to utter a single word or a short phrase that belongs to the expectation. In our example, the prompt tried to get the student to express the word effective or accurate (e.g., "The idea with instrumentation is that over time a piece of research equipment becomes less what?" answer: accurate. If the prompt is successful, then an assertion is provided, which is often a paraphrase of the expectation (e.g., Student 2: "I see now. Instrumentation refers…"). If it is not successful, then AutoTutor may provide another "hint -> prompt" cycle, depending on the authors of the tutorial program. AutoTutor has an authoring tool which allows the designers to provide content and to modify tutorial structures. The fifth step is to have the student assert whether they understood or to have them do another task which illustrates understanding.

Operation ARA employs the 5-step procedure described above in Basic Training when players answer the multiple choice questions in a lesson and in Proving Ground when the player has not yet identified remaining flaws in a case. However, what is unique about the implementation of AutoTutor in Operation ARA is that instead of only using a computer tutor + human tutee configuration, it adds another pedagogical agent to the mix. Hence, we have *trialogs*: tutorial conversations in the game between a teacher agent (Dr. Quinn), a student agent (Glass Tealman) and the human player. We modified the 5-step procedure (particularly step 4) to fit different conversational structures based on the prior knowledge of the player regarding the current topic. Prior knowledge for a topic is assessed in Basic Training by the performance on the multiple choice questions. When prior knowledge is scored as low, AutoTutor chooses a structure in which Dr. Quinn tutors Glass and the human is asked to verify Glass's understanding in Step 5. We call this the *vicarious* trialog. When prior knowledge is scored as higher, then the player is tutored by Dr. Quinn (*standard* trialog) as Glass chimes in periodically, or by Glass asking the player for help (a *teaching* trialog). In teaching trialogs, the player tutors Glass while Dr.Quinn chimes in when a misconception arises. The same expectations, hints and prompts are used in each trialog form (for a particular multiple choice question), although they differ in how they are expressed. For example, in the teaching trialog, Glass introduces the question, "[Player's name], I had trouble with this question. Can you tell me what instrumentation means?" and hints ("I remember reading something about how instrumentation is related to how research equipment is sometimes used for long periods of time.") and prompts ("I think I remember that over time a piece of research equipment becomes "less something." What was that word?").

Learning Principles and Operation ARA

AutoTutor and its game-like off-spring Operation ARA implement a number of learning principles. The reason that we focus on tutorial dialogs is that dialog interactivity is a key principle whereby people learn from holding conversations. Below, we highlight four additional learning principles that function within the AutoTutor environs: the zone of proximal development, self-explanation, spacing effect, and encoding variability.

The zone of proximal development

This concept is borrowed from developmental psychology and refers to the conceptual distance between the current knowledge and skill level of an individual when working on a problem and what could be learned from more capable peers (Vygotsky, 1978). When individuals are placed outside of this zone, they may experience boredom if the problem is too easy, or frustration or disengagement if the problem is too difficult.

Operation ARA relies on the zone of proximal development in global and local ways. Globally, the ordering of the modules helps to maintain individuals in the zone by going from easier to harder tasks. That is, instruction begins with the concepts (Basic Training) before the individual applies the concepts to examples (Proving Ground) which occurs before the player learns to make fine-tuned discrimination (flawed versus not flawed) via questioning scientists. Another global process is the selection of the next problem or main question designed to extend the learner's knowledge by selecting problems and questions that fill gaps and correct difficulties. Difficulty increases with each module because each requires knowledge and skills taught and assessed in the preceding module. The principle works locally within the trialogs that occur during Basic Training. Recall that when players have low knowledge regarding a concept, they will receive a vicarious trialog rather than the more demanding standard or teaching trialog. If low-knowledge students had received one of the latter trialogs, it is likely that they would be frustrated. Similarly, if high-knowledge students had received a vicarious trialog, it is likely that they would be bored. Empirically, this strategy is based on the work of VanLehn et al. (2007) who found the most learning from dialogs occurred when the material was a bit more advanced than the current level of the student. Therefore, the game presents the vicarious trialogs only when there is good reason to believe that the player has low-level knowledge. When a player has at least a moderate level of knowledge, the trialogs will be at or above that level, ensuring that the zone of proximal development applies.

Self-explanation

Self-explanation refers to the act of a person (reader, player) explaining the materials to one's self (See the chapter by Chiu & Chi, in this volume). Self-explanation includes cognitive activities such as reasoning, and identifying causes, antecedents, and boundary conditions. Self-explanation increases comprehension (Palinscar & Brown, 1984) and metacognition (Griffin, Wiley, & Theide, 2008), which is addressed in the chapter by Girash (this volume). Students engage in self-explanation in Operation ARA during trialogs, and periodically in Proving Ground when they document identified research flaws.

Spacing effect

Memory increases when the material is encoded (Bahrick & Hall, 2005) and tested (Roediger & Karpicke, 2006) over time (see the chapter by Carpenter, this volume). For example, in Operation ARA, players are exposed to instances of the same concept (e.g., attrition) in the ebook (Basic Training), and during research summaries in both Proving Ground and Going Active modules. Consequently, players will see and be tested on the concepts across the game and in different contexts.

Encoding variability

An ultimate learning goal is that students are able to transfer learned information to new and novel situations. Encoding variability refers to the finding that transfer to novel situations increases when there is variability in the materials during learning (Bransford, Sherwood, Hasselbring, Kinzer, &

Williams, 1990; Halpern, 2002). Built into Operation ARA are three different phases, providing, perhaps the most dramatic implementation of presentation variability. Another source of variability in Operation ARA occurs when players read and practice identifying the target concepts in research summaries related to a variety of topics, including psychology, biology, sociology, and chemistry.

Learning from AutoTutor and Operation ARA

Graesser and colleagues have reported that various instantiations of AutoTutor have lead to significant learning gains (Graesser et al., 2012; VanLehn et al., 2007). Based on more than 20 experiments, they report an average effect size of .80 (Graesser et al., 2012). This represents a "large" effect size (Cohen, 1988) and roughly corresponds to a letter grade in the United States. The experiments that fed into this average comprised different topics (physics, computer literacy), and different control groups (do nothing, read equivalent material). Therefore, there is ample evidence that content and skills when delivered by AutoTutor leads to significant learning.

The AutoTutor assessments only partially apply to Operation ARA because the game employs AutoTutor in only two of its modules (Basic Training and Proving Ground). In addition, there are many other features of the game (e.g., points, storyline) which may theoretically distract from learning. For example, the storyline itself may disrupt learning because following it may take cognitive resources away from learning target concepts (Adams, Mayer, MacNamara, Koenig & Wainess, 2012) and function as a "seductive detail." The seductive details effect occurs when interesting but irrelevant adjuncts (pictures, inserts) in text lead to decreased learning (Garner, Gillingham, & White, 1989). One reason for the effect is that learners organize the material around the adjuncts rather than the important causal information in the body of the text (Harp & Mayer, 1998). Unfortunately, the research on seductive details and split attention effects only apply to short interventions of 10 minutes to 90 minutes. The conclusions may not scale up to interventions of 10 to 40 hours, when motivation plays a central role in learning. We believe that the storyline keeps student learners motivated because they will want to find out what happens next to the characters in the story.

To examine if the storyline and other game features in Operation ARA increase, decrease or have no impact on learning, we manipulated the presence of game features in the Proving Ground module (Millis, Timmins, Wallace, & Graesser, 2011). Students in an undergraduate research methods course were in one of three learning groups: (1) *Game,* in which they played the Proving Ground module; (2) *No Game Control,* in which they completed the same activities (with the same materials) except without game play (i.e., no graphics, storyline, and competition); (3) *Do Nothing Control,* in which they did neither. All students took a pretest and posttest (separated by 2 weeks, counter-balanced) in which they were required to critique research. Students learned in both Game and No Game Control groups, however, student performance in the Game group was greater than in the No Game Control ($p < .05$, $\eta 2=.10$). As expected, students in the Do Nothing group showed no learning. These results demonstrated that although the materials and feedback contributed to learning, the gaming activity produced an added value.

We have also shown that students learn from playing the Basic Training module (Halpern, Millis, Graesser, Butler, Forsyth, & Cai, 2012). Students enrolled in a community college, a state university, and in a private liberal arts college either played or did not play the Basic Training module. Students took a pretest and a posttest comprising multiple-choice questions and constructed responses (short answer questions) that were counter-balanced. We computed proportional learning gains (posttest – pretest/1 – pretest) for each participant. Overall, students who played the game had significantly higher proportional learning gains (*M* = .19) than students who did not (*M* = -.10), *d* = 1.40. Statistically, there was no difference in learning gains among students in the three types of college, suggesting that students in different types of colleges benefit from playing the game.

More recently, we have found that students at a state university learn from playing the entire game. In this experiment, Introductory Psychology students at Northern Illinois University played Operation ARA as part of their course. All students took a pretest and a posttest, labeled in the game as “application test” and “exit test,” respectively. Students in a control condition answered the same questions on a web-based questionnaire. The tests comprised 24 multiple-choice questions, two for each concept. Students in “experimental” classrooms (*N* = 196) played Operation ARA in-between the pretest and posttest, whereas students in “control” classrooms (*N* = 236) played Operation ARA after the posttest. Students in the experimental and control conditions did not differ on their pretest scores (60% vs 56% correct), but students in the experimental condition had greater posttest scores than the control condition (71% vs 60% correct), leading to a significant interaction ($p < .01$, *d* = .47) (greater learning gains on the posttest for students in the experimental condition). Mixed models analyses confirmed that the interaction remained significant when the effect of the different classrooms and instructors were taken into consideration. Although the effect size is smaller in this experiment compared to others (*d* = .47), it is still considered a “medium” effect (Cohen, 1988).

The above findings are consistent with a meta-analysis on serious games recently conducted by Wouters, van Nimwegen, van Oostendorp, and van der Spek (2013; also see Sitzmann, 2011, and Vogel, Vogel, Cannon-Bowers, Bowers, Muse & Wright, 2006). The Wouters et al. (2013) analysis included 38 studies in which students learned content with a serious game alone or in combination with other instructional methods, and in which the study included a control group that was equated on content. They found that teaching with serious games led to significantly greater learning than conventional methods (d = 29). This advantage occurred when the game was presented alone (*d* = .20) or when it was supplemented with instructional methods (d = .41). The authors speculated that combining serious games with other instructional methods increased the effect because the supplemental materials might have offered opportunities for reflection and the articulation of knowledge, thereby facilitating learning (Mayer, 2004). In addition, the advantage for serious games over conventional methods was maintained when delayed tests were used in the study (*d* = .36). The authors also reported significantly more learning with serious games when students were engaged in multiple sessions than when they were engaged in a single session ($p < .01$).

Interestingly, Wouters et al. (2013) found that serious games did not significantly increase motivation, as compared to conventional teaching methods. They report an effect size of *d* = .29, but the effect did not reach statistical significance. This is an important finding because serious games are thought to increase motivation, thus being attractive to educators (Garris, Ahlers, & Driskell, 2002). Wouters et al. speculated that the null effect might have occurred because (1) serious games are typically chosen by teachers (or supervisors) rather than by students, and that this might have undermined intrinsic motivation (Deci, Koestner, & Ryan, 1999), and (2) game design with its emphasis on entertainment is not well integrated with instructional design. Indeed, it is difficult to develop a game that integrates learning and entertainment, and one that does each well. More research is needed to examine under what conditions serious games affect motivation.

Advice to Teachers

There are a myriad of issues that teachers must consider when thinking about adopting new technology as part of the curriculum. Choosing to have students interact with an ITS or a serious game is no exception. Certainly one of the most pressing concerns is whether the technology increases students' learning and (hopefully) motivation. Obviously, there is no "magic bullet" if one adopts a serious game for their class. Everyone will not enjoy and learn from the game. In fact, research with AuoTutor has shown a significantly negative relation between versions that promote deep learning and versions undergraduate students rate as enjoyable (Jackson & Graesser, 2007). Deep learning by students it would seem is not an enjoyable struggle. This may be true of serious games as well, but the alternative is equally plausible if games can turn work into play. We recommend that teachers attempt to gauge the extent that students' learn from the game, and the extent that they find the game motivational, engaging, usable (usability issues), and relevant to their learning goals. We do not suggest that teachers base adoption only on affective dimensions (e.g., liking, motivation). If learning is the primary objective, then we think it is important to (1) examine the empirical evidence for learning from a given technology before adoption, (2) attempt to assess learning from the game as it is implemented in the course, and based on Wouters et al. (2013), and (3) augment the game with other materials in the classroom.

In deciding whether to adopt and use new technology, like an ITS or serious game, we recommend that the teacher address the following questions (de Freitas & Oliver, 2006). First, does the technology teach the content and skills of interest? The technology is irrelevant if it does not. Second, is the technology age-appropriate? The content and/or the interface might be too difficult for younger students, or perhaps too easy for older students. Ideally, we would want the students to be in the zone of proximal development when interacting with the technology. Third, does the technology take into consideration the different skill and knowledge levels of the student? Maintaining a student in the "zone" is difficult to achieve so it is important that the technology contains some mechanism to do so. Fourth, what resources are available for the students and teachers regarding the technology? Teachers and students do not want to be involved in a flurry of reports or emails that the technology is broken.

References

Adams, D.M., Mayer, R.E., MacNamara, A., Koenig, A., & Wainess, R. (2012). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology, 104*, 235-249. doi:10.1037/a0025595

Annetta, L.A. (2008). Video games in education: Why they should be used and how they are being used. *Theory Into Practice, 47*, 229-239. doi:10.1080/00405840802153940

Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education, 15*, 95–115.

Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language, 52,* 566-577. doi:10.1016/j.jml.2005.01.012

Bransford, J.D., Sherwood, R.S., Hasselbring, T.S., Kinzer, C.K., & Williams, S.M. (1990). Anchored Instruction: Why we need it and how technology can help. In D. Nix & R. Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 115-141). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cai, Z., Graesser, A.C., Forsyth, C., Burkett, C., Millis, K., Wallace, P., Halpern, D.H., & Butler, H. (2011). Trialog in ARIES: User Input Assessment in an Intelligent Tutoring System. In W. Chen & S. Li (Eds.), *Proceedings of the 3rd IEEE Intern. Conf. on Intelligent Computing and Intelligent Systems* (pp.429-433). Guangzhou: IEEE Press.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Lawrence Erlbaum Associates.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation*. Psychological Bulletin, 125*, 627-668. doi:10.1037/0033-2909.125.6.627

de Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education, 46*, 249-264. doi:10.1016/j.compedu.2005.11.007

Garner, R., Gillingham, M., & White, C. (1989). Effects of "seductive details" on macroprocessing and microprocessing in adults and children. *Cognition and Instruction, 6*, 41-57. doi:org/10.1207/s1532690xci0601_2

Garris, R., Ahlers, R., & Driskell, J.E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*, 441-467. doi:10.1177/1046878102238607

Graesser, A. C., D'Mello, S.K., & Cade, W. (2011). Instruction based on tutoring. In R.E. Mayer and P.A. Alexander (Eds.), *Handbook of Research on Learning and Instruction* (pp. 408-426)*.* New York: Routledge Press.

Graesser, A. C., D'Mello, S. K., Hu. X., Cai, Z., Olney, A., & Morgan, B. (2012). AutoTutor. In P. McCarthy and C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation, and resolution* (pp. 169-187). Hershey, PA: IGI Global.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods*, *Instruments*, *and Computers*, *36*, 180–193. doi:10.3758/BF03195563

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, *31*, 104–137. doi:10.3102/00028312031001104

Graesser, A. C., Person, N., & Harter, D. (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*, *12*, 257–279.

Graesser, A.C., Person, N.K. & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9*(6), 495-522.

Griffin, T. D., Wiley, J. & Thiede, K. W. (2008) Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition, 36*, 93-103.

Halpern, D. F. (2002). Teaching for critical thinking: A four-part model to enhance thinking skills. In S. Davis & W. Buskist (Eds.), The teaching of psychology: Essays in honor of Wilbert J. McKeachie and Charles L. Brewer (pp. 91-105). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Halpern, D.F., Millis, K., Graesser, A.C., Butler, H., Forsyth, C. & Cai, Z (2012). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity, **7***, 93-100.

Harp, S. F. & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 3,* 414-434. doi:10.1037/0022-0663.90.3.414

Jackson, G. T., & Graesser, A. C. (2007). Content matters: An investigation of feedback categories within an ITS. In R. Luckin, K. Koedinger, & J. Greer (Eds.), *Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work* (pp. 127–134). Amsterdam: IOS Press.

Johnson, W. L., Rickel, J. F., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education, 11*, 47-78.

Katz, S., Allbritton, D., & Connelly, J. (2003). Going beyond the problem given: How human tutors use post-solution discussions to support transfer. International Journal of Artificial Intelligence in Education, 13, 79-116.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59*, 14-19. doi:10.1037/0003-066X.59.1.14

Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A., Halpern, D. (2011). Operation ARIES!: A Serious Game for Teaching Scientific Inquiry. In M. Ma, A. Oikonomou, and L. Jain (Editors), *Serious Games and Edutainment Applications*. (pp. 167-194). Springer-Verlag, UK.

Millis, K., Timmins, T., Wallace, P., & Graesser, A. (2011). Serious games for serious learning. Presented at *21th Annual Meeting of the Society for Text and Discourse, Poitiers, France*.

Moreno, R. (2005), Multimedia Learning with Animated Pedagogical Agents. In R. Mayer (Ed.) *The Cambridge Handbook of Multimedia Learning*, Cambridge University Press, pp. 507-524.

Palincsar, A.S. & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction,* **1***,* 117-175. doi:10.1207/s1532690xci0102_1

Ritterfeld, U., Cody, M. & Vorderer P. (2009) *Serious games: Mechanisms and effects* (pp. 103-116).

Roediger, H. L. III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210. doi:10.1111/j.1745-6916.2006.00012.x

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*, 489-528. doi:10.1111/j.1744-6570.2011.01190.x

Sleeman, D., & Brown, J. S. (1982). Intelligent tutoring systems. New York, NY: Academic Press.

VanLehn, K., Graesser, A.C., Jackson G.T., Jordan, P. Olney, A., & Rose, P.A. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science,* 2007, **30**, 1-60.

Vogel, J.J., Vogel, D.S., Cannon-Bowers, J., Bowers, C.A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-Analysis. *Journal of Educational Computing Research, 34*, 229-243. doi:10.2190/FLHV-K4WA-WPVQ-H0YM

Vygotsky, L. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: President and Fellows of Harvard College.

Woolf, B. (2009). Building Intelligent Interactive Tutors: Student-centered strategies for revolutionizing e-learning. Burlington, MA: Morgan Kaufmann Publishers.

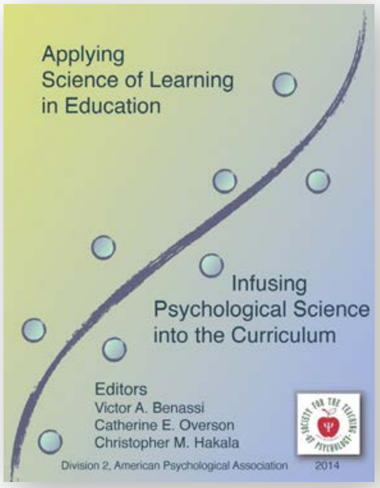
Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology, 105, 249-265. doi:10.1037/a0031311

Part 2

Preparing Faculty to Apply the Science of Learning

Division 2,
American Psychological Association

2014



Assessing the Impact of Instructional Methods

Regan A. R. Gurung
University of Wisconsin-Green Bay

Once you discover the rich findings from the science of learning (ScoL) and adapt your instructional methods accordingly, it is time to assess the educational impact of the changes you have made. Practicing clinicians need to be able to evaluate evidence-based treatments and new developments in the field and incorporate them into their own practices. Similarly, good teachers need to be able to incorporate the latest findings into their classes. What can teachers do to assess the kind of science of learning based instruction covered in this book? How can you create and use educationally relevant outcome measures? What are some simple methods of research design that you can develop and apply to be able to attribute learning outcomes to the educational interventions you use? Not all instructors consider such questions. Part of the reason is that for some disciplines these questions come naturally. Another part of the reason is that many of us do not always take a research or scholarly approach to our teaching although we may be otherwise well published scholars. Regardless of why, it is important to be able to take the fruits of the science of learning's labors and use them to make systematic, intentional changes to our pedagogy.

Conducting Scholarship of Teaching and Learning (SoTL)

The good news is that there is an active community of instructors who provide models for how to modify instructional methods and assess their effectiveness. When a teacher critically assesses her own teaching by focusing on student learning, using robust research methods and then sharing the results in peer-reviewed formats, she is said to be doing the *scholarship of teaching and learning* (SoTL). SoTL entails intentional, systematic reflections on teaching and learning resulting in peer-reviewed products made public (Gurung & Schwartz, 2012; Potter & Kustra, 2011). By taking ScoL results, applying them to your classroom, assessing the effectiveness of the results and disseminating the results, you will be conducting SoTL. Boyer (1990) popularized the term "scholarship of teaching" although caring teachers have practiced the kind of work to which it refers for many years. SoTL can be an integral part of every academic's life. SoTL "brings powerful new principles and practices" to position decisions about key academic questions such as what students need to know and should be able to do (Hutchings, Huber, & Ciccone, 2011, p. 3). Some faculty have argued that pedagogical research of this sort should be seen as part as one's professional responsibility as a teacher (Bernstein & Bass, 2005). But if you do not assess the effect of your innovation or pay close attention to the different ways the change can make a difference, you are not being an ethical teacher or using SoTL ethically (Gurung, 2012).

Criteria for SoTL

Glassick, Huber, and Maeroff (1997) identified criteria for evaluation that apply equally to traditional scholarship and to SoTL. The results include the following six standards to be applied to assess the quality of research from any type of scholarship: Clear goals, adequate preparation, appropriate methods, significant results, effective communication, and reflective critique. Wilson-Doenges and Gurung (2013) identified a continuum of SoTL and demarcated aspirational benchmarks that also serve as guidelines for research design. Whereas qualitative and quantitative data and methods all have a place in SoTL, the benchmarks provide clear-cut standards of design and analysis. Similar to social science's methodology for research in general, SoTL should also aim for similar standards that are theory-based and intentionally designed using the best models for methodological and statistical rigor. The benchmarks are divided into three main levels with the final level comprising the gold standard for research. This increased focus on strengthening the scientific rigor of SoTL should benefit how this form of research is recognized across disciplines in general. Similarly, Felten (2013) provides a universal set of guidelines to conduct SoTL.

Getting the Most Out of Published Science of Learning Research

Although one can look to blogs, twitter, or general media sources to learn about the most recent findings from the science of learning, the best way to get the latest pedagogical research findings is to read the relevant journal articles. You should be reading for ideas for research and examples of research designs, and also for different statistical analyses. The first key step in using ScoL results is to know where to locate them. Although there is no single outlet for this work, you can find useful articles in journals such as *Psychological Science* (e.g., Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013), *Journal of Experimental Psychology: Learning, Memory, and Cognition* (e.g., Karpicke & Roediger, 2007), *Memory* (e.g., Karpicke, Butler, & Roediger, 2009), *Journal of Educational Psychology* (e.g., McDaniel & Donnelly, 1996), and *Applied Cognitive Psychology (*e.g., Kang & Pashler, 2012), among many others.

An article published on a topic you are interested in can be a gold mine of information. The introduction should provide you with a wealth of relevant citations together with an immediate sense of how up to date you are on your topic of interest. If you are keeping up with the literature, you should recognize many of the citations in the article. With consistent reading, you should run into the usual suspects in terms of citations. It is a great feeling to read an article hot off the press and recognize most of the research discussed in the introduction section. Of course, this also gives you a good "to read" list.

The Method section of an article provides a nice opportunity to test your own research skills and your knowledge of the subject area. Once you are cognizant of the hypotheses being tested from the Introduction section, you can now see one way (the authors') of testing the hypotheses. Reading and understanding a variety of designs gives you the basic building blocks for when you want to design your own study. If you can understand *what* exactly researchers did (i.e., their research design), you will be in a better position to evaluate the results and critique the methods. It is always a fun exercise to see if you can come up with better ways to test the hypotheses. It is a lot easier to critique someone else's work than your own.

The Discussion section provides you with great fodder for future research. Most good articles have a strong Limitations section as well as a Future Directions section. Not only will you be able to generate a number of your own future directions, but it will also be useful to ponder those outlined by the authors. Ask yourself if you agree with what the authors propose as next steps. Do the suggestions follow directly from the findings discussed in the article? Such mental exercises will make you a more robust researcher.

Together with strong research skills, it is important to have good references to guide your statistics. One of the most readable books on statistics is by Field (2013), who not only gives you the basics behind the major statistical methods you will use but also provides the commands and results tables from the most commonly used statistical package (SPSS, the Statistical Package for the Social Sciences). Research and statistical skills are intertwined.

Being comfortable with statistics goes hand in hand with having strong research skills. Many schools yoke research methods classes with statistics classes and also require students to take statistics before methods. The reason for this pairing is simple. Different types of research designs need different statistical methods. Sometimes poor research design requires complex statistical methods to clean things up. For example, you may have grossly unequal number of students in the two groups you tested. This and many more issues can be somewhat rectified by appropriate statistical tests. Conversely, the cleaner the research design, the cleaner the statistics needed, whether quantitative or qualitative.

Quantitative analyses form the backbone of research in the natural and social sciences, although both these areas also incorporate qualitative methodologies, which are often critical to theory building and preliminary explorations. You should be well informed about both to prepare you for what your research will need. One of the main drawbacks of quantitative analyses is that they sometimes make it easy to forget about theory. Teachers are sometimes so excited to have numbers to play with that the theory that guided the research in the first place takes second place to the emerging patterns in the numbers. (There is a time and a method for letting the data shape your theory.) Indeed, quantitative researchers do collect numerical information and use statistical analyses to determine whether there are relationships between the numbers, but this does not have to be as cold or difficult a process as it sounds. Furthermore, qualitative researchers do “process data” as well. You may want to introduce a classroom innovation based on the ScoL and ask your students if their learning improved or if they found it easier to understand the concept. This focus group or interview method generates qualitative evidence similar to essay answers or the use of open ended questions, and can also lead to insights. Qualitative data processing brings order to observations and often involves coding, classifying, or categorizing the observations. In contrast to quantitative analyses, in which a standardized unit (e.g., a score on a questionnaire) is used, coding in qualitative analyses focuses on the concept as the organizing principle. There is a nice middle ground where one can do quantitative analyses on what started out as a qualitative investigation.

Qualitative analyses often involve closer connections among theory, data collection, and analyses and are most distinguishable from quantitative research by not always being driven by a formal hypothesis about how variables are related. For example, you do not have to assume that a given method of instruction is better than another or that certain classroom factors will facilitate learning more than others. Instead, the main goal may be to describe the situation or process in as rich detail as possible and to explore possible reasons why certain outcomes may occur. Like an undercover cop, the qualitative researcher often immerses herself in the setting. Hypotheses and theory then emerge from these observations. An example is the Grounded Theory Method (e.g., Urquhart, Lehmann, & Myers,

2010) of qualitative analysis. This stresses the importance of building an inductive-based theory using observations to reveal patterns in behavior and then building the theory from the ground up without preconceptions.

Main Steps in Assessing Instructional Methods

The steps for conducting research on teaching and learning mirror most of the steps used to conduct research on any topic (Sansone, Morf, & Panter, 2003). First, the teacher identifies a question of interest and then reviews what has been published on the topic. Next, the teacher ascertains how to answer the research question. One common approach to this sort of research is to measure relevant aspects of what the students are learning, make a change or introduce a new method or assignment, and then measure students' learning again to determine the extent to which the manipulation affected it (see Bishop-Clark & Dietz-Uhler, 2012; Gurung & Schwartz, 2012, for exemplars of conducting SoTL research). This process becomes even easier if you identify a ScoL finding or prescription (e.g., repeated testing aids learning) and want to change your class design accordingly.

Similar to conducting psychological science research in general, you can start with descriptive studies (e.g., how are my students learning?), move on to correlational designs (e.g., what is student learning associated with?), and then design experiments (e.g., if I change this assignment, or use a ScoL finding, or introduce a new way of talking about this concept, will learning change?). In designing classroom experiments, teachers can choose from a wide variety of models. Bartsch (in press) presents a fuller description of different research designs and particular adaptations for classroom research.

At its core, research consists of testing hypotheses. Researchers may choose from three distinct approaches to hypothesis testing: validation, falsification, and qualification. The development and nature of the experimental paradigm began a little over 80 years ago with the work of Fisher, who first formalized essential elements of research, including the manipulation of independent and dependent variables, and randomization (Pelham & Harton, 2006**).** Other essential characteristics of research you need to be familiar with are validity, reliability, and measurement scales (Morling, 2012). Validity and reliability are crucial to both experimental and passive observational research (naturalistic research that does not involve the manipulation of variables).

The best way to assess the use of a ScoL finding is to conduct an experiment. A form of research that affords the most control is the true experiment, in which the researcher has control over all of the independent variables of interest and uses random assignment. The independent variable is what is manipulated. For example, if you were to have one section of your class take multiple quizzes based on your reading of the ScoL literature on repeated testing, and compare their exam scores with another section that did not take multiple quizzes, the independent variable would be 'use of multiple quizzes'. True experiments are further broken down into one-way designs and factorial designs. In some studies, such as repeated measures designs, researchers hold some control over individual differences by exposing participants to more than one level of an independent variable. For example, you may use a pre and post test of learning (i.e., the repeated measure) to compare a section of a course that takes three quizzes as a study aid, with a section of a class that takes five practice quizzes. You cannot always manipulate factors such that one course section or class gets nothing as you may be keeping something valuable from that section. There are research designs to help control as many factors as possible and do research in the real world setting of the classroom (Morling, 2012). This is the tip of the iceberg in terms of research methods terminology, representing some of the most common elements of research design. You can familiarize yourself with many of these terms, issues, and research skills by reading peer-reviewed scholarly articles published in journals in your field. Reading a research methods book provides an essential foundation, but reading journal articles is what keeps your research mind sharp.

Basic Ways to Assess Pedagogical Innovations

One easy way to start is to pick a course in which you test students many times or courses that are offered consistently every semester and year. This type of Repeated Measures Design (RMD) works well in course designs that have several similar exams or assignments. The term “RMD” is used in research when the assessment is identical such as when department learning outcomes or course objectives are measured year after year in the same college student sample. Unlike the example in the previous section, where the pre and post test repeated measure is used with the same group of students, this form of design tests different cohorts. In such designs, the key is to identify changes in responses to similar question(s) over time. The measure used repeatedly consists of the same number of questions asked in the same order, and differences in the responses will be taken to indicate changes in knowledge.

However, using identical questions is not always practical or possible in most courses. To avoid this problem, many teachers modify the RMD to include a pretest and a posttest. For example, many large general education courses give the same test at the beginning and end of the semester (e.g., a test of knowledge of governmental policy in a political science course). To test whether learning is changing over the course of the semester, the teacher can test if the class average is changing over time, test if learning has changed from the beginning of the semester (using a class average) or even compare a single student’s score to his or her previous score to determine if a student is improving over the course of the semester. If there is a significant difference in student learning between the two assessments, it could be due to the instruction the teacher provided in between the pretest and the posttest. Of course, a teacher cannot be sure that the change was due only to instruction unless she has measured and controlled for many other possible factors such as how much and how the student studied. If, while holding other variables constant, a teacher finds a significant difference between the pretest and posttest measures, then she may be confident that her finding is a good indicator that instructional changes produced increases in learning.

This basic idea – measure learning (pre-test), introduce a change (e.g., repeat quizzing), measure learning again (post-test) – allows you to test the utility of any pedagogical innovation. The ScoL provides a number of good practices (Dunlowsky, Rawson, Marsh, Nathan, & Willingham, 2013); you can take any one of them, introduce it into your class and see how learning changes. Although many such investigations can be conducted on a single class in a single semester, a pretest/posttest approach introduces the problem of testing effects when the original test affects later measurements. For example, within a social psychology class, you may first give students a 15-item measure assessing understanding of the cognitive dissonance or the bystander effect. You then discuss the concept in class using a new approach you read about in the ScoL literature and give them the same test to see if their knowledge increases. The problem is that taking the first learning measure, students might perform better on the second measure simply because they had taken the first measure and not due to your instruction. To avoid students recognizing a past measure, a good solution is to have two measures for the same construct (e.g., a form A and form B of a measure with different questions on each, Bartsch, Engelhardt Bittner, & Moreno, 2008).

Sometimes a teacher may not even use a pre measure of learning. The simplest type of design for comparing two-groups is called a two-group post-only design. In the two-group post-test only design, one group receives the ScoL innovation and another does not, and then both are measured on the same assessment. The two groups could be two sections of a class or one class divided into two or the same class over two consecutive semesters. For example, in a test of whether giving PowerPoint lecture notes to students before the lecture was given in class, Noppe gave one section of her class notes in advance and did not provide notes to the other section (Noppe, 2007). She used two separate sections of the same course but used the same exam. The class that did not receive the notes scored lower on the exam. There was no pretest of learning used in this case but Noppe checked to see if the grade point averages of students in both sections were similar at the start of the class to eliminate the possibility that students in the section that received the notes were not as good students, which could have accounted for the same result.

Sometimes you can make sure all students get the innovation, or different levels or types of an intervention. This type of design is called a within-participants design. This type of design is useful when you want to test different variations on a theme. For example, one study examined the effectiveness of PowerPoint presentations (Bartsch & Cobern, 2003). In this study researchers compared the effectiveness of different forms of presentation format and content, a popular object of study in the ScoL (e.g., Mayer, 2011). Bartsch and Cobern used three forms of delivery: overhead transparencies, plain PowerPoint slides, and PowerPoint slides with pictures, graphs, transitions, and sound effects. Conditions rotated each week, and each week researchers quizzed the students.

Designing a sound way to test if your innovation worked is important, but do not forget to take the time to also design sound learning assessments. There are also good practices for writing test questions, essay prompts, and other forms of tests. Taking the time to fine tune one's rubrics for grading and aligning assessments with one's learning outcomes is time well spent. Stevens and Levi (2011) and Suskie (2009) provide indispensable guides for rubric development and assessment.

Did Your Innovation Work?

Once you incorporate a ScoL finding into class and design a study to test if it works, you then need to pay close attention to factors that can influence whether the innovation 'worked.' Social science research nicely alerts us to the fact that the changes we observe in our students' learning may be due to a whole host of factors. One of these factors indeed may be what we do as instructors or what the students have done (perhaps because of our instruction), but observable changes may also be due to other naturally occurring factors. People change over time (i.e., maturation). Factors outside our awareness influence results (i.e., history). It is possible that changes that we see in our student learning are due to these natural changes or factors external to our instructional interventions. Social science methodology alerts us to these two confounds, to many others, and most importantly, to ways to avoid them. Watching out for these confounds is critical when assessing the effectiveness of changes to instructional methods. Detailed descriptions of these confounds can be found in research methods books and are worth the perusal (Creswell & Piano Clark, 2012; Morling, 2012).

It is also important to ask if the changes are statistically significant. Social science methodology involves significance testing. Assessing change is one type of pedagogical endeavor that necessitates the quantitative method, regardless of discipline. If you want to know whether students' performance improved after a change you made (e.g., a new assignment, an innovative presentation, group work, flipping your class), you need to know if that change would have happened by chance or if any other factors could account for it. This requires quantification of the evidence (e.g., themes brought up in a close reading, concepts used in essays, levels of meaning). When social scientists ask if the change is statistically significant, they want to ensure that the change is due to what was done, and not just due to chance. Stated in this way, it is seems hard to not care about statistical significance. If you have worked hard to change your instruction and improve student learning, it is important to know whether that change would have happened by chance and without your intervention in the first place. Before one spends more time and energy on changing instruction or even trying to get others to also change instruction based on the changes you have seen, you should be sure your changes are not random. Statistical testing does this for you.

The most common statistical tests used to assess the effectiveness of instructional changes are *t*-tests and analyses of variance (ANOVA). These statistical analyses test for differences between groups – the class or section that got an innovation and the class or section that did not. There are no gray areas when it comes to testing. The statistical program (e.g., SPSS) provides a test statistic (e.g., an *F* or *t* test) and a probability value (i.e., *p* value). If the probability value is less than .05, the difference between groups is considered statistically significant and you can assume your innovation is important. Note that a *p* value is influenced by the size of your sample (the number of students in your groups) so some tests fail to be significant because they lack sufficient statistical power due to a small sample size. That said, for starters you want to see a *p* value less that .05. For step by step directions on conducting statistical analyses for SoTL or an easy to read exposition of statistical testing, see Gurung and Schwartz (2012) and Field (2013).

Note that statistical significance need not be the ultimate and only criterion for SoTL but it is certainly something to be considered for appropriate and relevant research designs and questions. Furthermore, statistical significance should not be confused with or taken to be synonymous with ‘significant’ as used in everyday life (i.e., to mean important).

There are a number of additional research designs (see Morling, 2012). Each has pros and cons. Although there may be better and worse ways to design and conduct a study, there will never be *one* clear right way (Bartsch, 2013). The best we can do is to control for as many factors as possible in a systematic and intentional way. Starting with disciplinary methodologies and approaches you are comfortable with (and know how to use) makes great sense, but why stop there? If your question is learning in general, it makes better sense to develop your question and then pick the best methodology for the question regardless of the discipline. This may involve collaboration and reviewing literature on how other disciplines conduct questions surrounding teaching and learning (Gurung, Chick, & Haynie, 2009).

The science of learning provides the teacher with a wide variety of findings ripe for modifying and using in the classroom. It is imperative to assess the effectiveness of any change in a robust way and be prepared for the possibility that what may work in a cognitive psychology lab may not necessarily work in a classroom. Even some science of learning results from other classroom studies may not necessarily translate to your classroom. Using basic research design methods well will at least ensure that you get a sense of what does work.

References

Bartsch, R. A. (2013). Designing studies to measure student success. In R.A.R. Gurung & J. H. Wilson (Eds.). *Doing the Scholarship of Teaching and Learning: Measuring systematic changes to teaching and improvements in learning (pp. 00-00)*. San Francisco, CA: Jossey-Bass.

Bartsch, R. A., & Cobern, K. (2003). Effectiveness of PowerPoint presentations in lectures. Computers & Education, 41, 77-86.

Bartsch, R. A., Engelhardt Bittner, W. M., & Moreno, J. E. (2008). A design to improve internal validity of assessments of teaching demonstrations. *Teaching of Psychology*, *35*(4), 357–359. doi: 10.1080/00986280802373809

Bernstein, D. & Bass, R. (2005, July-August). The scholarship of teaching and learning. Academe, 91, 37-43.

Bishop-Clark, C., & Dietz-Uhler, B. (2012). Engaging in the scholarship of teaching and learning: A guide to the process and how to develop a project from start to finish. Sterling, VA: Stylus.

Boyer, E. L. (1990). Scholarship reconsidered: Priorities of the professoriate. San Francisco, CA: Jossey-Bass.

Creswell, J. W., & Piano Clark, V. L. (2012). Designing and conducting mixed methods research. Thousand Oaks, CA: Sage.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. Psychological Science in the Public Interest, 14(1), 4-58. doi:10.1177/1529100612453266

Felten, P. (2013). Principles of good practice in SoTL. Teaching and Learning Inquiry, 1, 121-126.

Field, A. (2013). Discovering statistics using SPSS. Thousand Oaks, CA: Sage.

Glassick, C. E., Huber, M. T., & Maeroff, G. I. (1997). Scholarship assessed: Evaluation of the professoriate. San Francisco: Jossey-Bass.

Gurung, R. A. R. (2012). Consuming the Scholarship of Teaching and Learning: Using evidence-based pedagogy ethically. In R. E. Landrum & M. A. McCarthy (Eds.), Teaching ethically: Challenges and opportunities (pp. 67-76). Washington, DC: APA.

Gurung, R. A. R., & Schwartz, B. M. (2012). Optimizing teaching and learning: Practicing pedagogical research. Malden, MA: Wiley-Blackwell.

Gurung, R. A. R., Chick, N., & Haynie, A. (Eds.) (2009). Exploring signature pedagogies: Approaches to teaching disciplinary habits of mind. Sterling, VA: Stylus.

Hutchings, P., Huber, M. T., & Ciccone, A. (2011). The scholarship of teaching and learning reconsidered: Institutional impact. San Francisco, CA: Jossey-Bass.

Kang, S. K., & Pashler, H. (2012). Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast. Applied Cognitive Psychology, 26(1), 97-103. doi:10.1002/acp.1801

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. Journal of Memory & Language, 57(2), 151-162. doi:10.1016/j.jml.2006.09.004

Karpicke, J. D., Butler, A. C., & Roediger III, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? Memory, 17(4), 471-479. doi:10.1080/09658210802647009

Mayer, R. E. (2011). Applying the science of learning. Boston, MA: Pearson.

McDaniel, M. A., & Donnelly, C. M. (1996). Learning with analogy and elaborative interrogation. Journal of Educational Psychology, 88(3), 508.

Morling, B. (2012). Research methods in psychology: Evaluating a world of information. New York: Norton.

Noppe, I. (2007). Powerpoint presentation handouts and college student learning outcomes. International Journal for the Scholarship of Teaching and Learning, 1. Retrieved from www.georgiasouthern.edu/ijsotl

Pelham, B. W., & Harton, B. (2006). Conducting research in psychology: measuring the weight of smoke. (3 ed.). Belmont, CA: Wadsworth.

Potter, M. K., & Kustra, E. (2011). The relationship between scholarly teaching and SoTL: Models, distinctions, and clarifications. International Journal for the Scholarship of Teaching and Learning, 5. Retrieved from www.georgiasouthern.edu/ijsotl

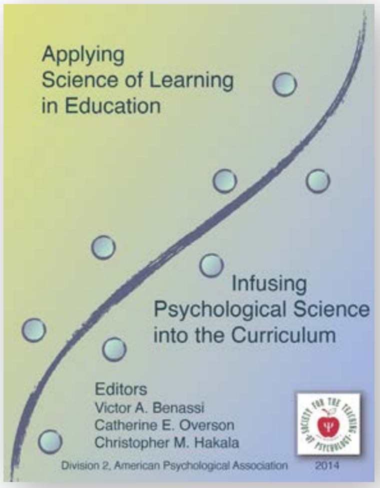
Sansone, C., Morf, C., & Panter, A. (2003). The Sage handbook of methods in social psychology. Thousand Oaks, CA: Sage.

Schwartz, E., & Gurung, R.A. R., (2012). *Evidence-based teaching in higher education*. Washington, DC: American Psychological Association.

Stevens, D. D., & Levi, A. (2011). Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning. Sterling, VA: Stylus.

Suskie, L. (2009). Assessing student learning: A common sense guide (2nd ed.). San Francisco, CA: Jossey-Bass/Wiley.

Wilson-Doenges, G., & Gurung, R. A. R. (2013). Benchmarks for scholarly investigations of teaching and learning. Australian Journal of Psychology, 65(1), 63-70. doi:10.1111/ajpy.12011



Applying the Science of Learning: The Cognition Toolbox

Victor A. Benassi, Elizabeth M. Tappin, Catherine E. Overson, Michael J. Lee, Edward J. O'Brien, Barbara Prudhomme White
University of New Hampshire

Jennifer J. Stiegler-Balfour
University of New England

Christopher M. Hakala
Western New England University

Basic and applied research, starting in the early 20th century and continuing to the present, has documented that learning (ranging from acquisition of rote to higher order cognitive skills) can be positively and sometimes dramatically influenced by certain conditions in which the learning environments are structured. Until relatively recently, however, there has been little systematic examination of the principles that come from this research in the context of postsecondary institution courses and curricula. A lot is known about how people learn and about how instructional environments can be modified to maximize their learning, retention, and transfer of knowledge. However, based on our experiences with faculty and academic administrators over the many years, we see little reason to modify the view expressed by Diane Halpern and Milton Hakel in 2002: “It is sadly true that most of the way we teach and learn is uninformed by laboratory findings in human cognition” (Halpern & Hakel, 2002, p. 1). Moreover, there is scant evidence of anything approaching the systematic application of science of learning principles in higher education (or at any academic level for that matter). The Cognition Toolbox program described in this chapter represents one effort to address this issue (see also Swaboda, this volume).

Overview of Cognition Toolbox Program

The University of New Hampshire (UNH) Center for Excellence in Teaching and Learning (CETL) launched an initiative in 2007 to undertake a new approach in assisting faculty who want to improve students’ learning, retention, and transfer of course material. Since then, we have worked with faculty to apply in their courses one or more tools, based on the science of learning framework (Mayer, 2010), from our *Cognition Toolbox.* We then assess the impact of these interventions using direct measures of learning, retention, and transfer of knowledge. The overarching goal of our program has been to design, implement, evaluate, and disseminate our Cognition Toolbox in a broad range of academic courses.

Approach

We have followed the same basic approach for all our projects, as described below.

- We begin by advertising projects to all UNH teaching faculty, and identifying certain types of courses for inclusion in the various projects (e.g., general education courses, linked courses in majors).
- Once we have identified a particular course, we meet with the instructor and determine what learning issue(s) on which the teacher wishes to focus. The project team then reviews all course-related materials (syllabus, assignments, learning assessments, etc.) and proposes an intervention based on an appropriate science of learning principle. Finally, we design a strong learning outcomes assessment protocol in order to collect evidence about the impact of the intervention.
- Prior to or at the start of a course, a project staff person works with the instructor to ensure the proper implementation of the cognitive principle. Contact with the instructor during the semester ranges from minimal to moderate, depending on particular circumstances. One of our overriding objectives is to minimize additional work on the part of the teachers who participate in the project. If there are to be clear benefits to student learning, we need to develop interventions that teachers will reliably use in their courses.
- About half way through a course offering, a project staff person often, but not always, conducts a midcourse assessment to determine the extent to which students perceive the benefits from the cognitive tool(s) used in their course. For information about our midcourse assessment process, go to:

  http://www.unh.edu/teaching-excellence/MidAssessmentProcess/index.html
- Project staff members are responsible for designing and implementing the assessment of the impact of the Cognition Toolbox intervention in each course. We have developed over the course of the last six years a strong set of evaluation designs to assess learning outcomes that our interventions are designed to affect.
- When a course is finished, we provide a report to the instructor. Often we continue to work with teachers in following semesters, making changes in our interventions and assessments based on what we learned during the semester.

Additional details on the Cognition Toolbox may be found at:

http://www.unh.edu/teaching-excellence/Cognitiontoolkit/index.html

Who Has Participated?

We open our call to teachers from across the university, from all fields and disciplines. If teachers have learning issues they want to address with assistance from CETL, we do our best to accommodate them. As a result, we have worked with teachers in a wide variety of courses on a wide range of learning issues, representative examples of which are shown below:

English Composition	Social Psychology	Soils Science	Community Psychology

General Chemistry	Chemistry for Engineers	American Government	British Literature
Human Reproduction	Macro Economics	Epidemiology	Consciousness
Introduction to Justice Studies	Science and Nature of Human Stress	Principles of Animal Physiology	Introduction to Geographic Information Systems
Molecular and Cellular Biology	Health Care Financial Management	Science and Practice of Strength Training	Evolution, Biodiversity and Ecology
Statistics	Human Occupation	Introductory Psychology	Introduction to Physics

Rationale for Our Approach

Our approach to working with faculty differs from an approach often used by staff of teaching and learning centers—one-shot workshops. A workshop is designed and presented to a group of teachers on a particular approach and/or technique—for example, on promoting deep learning, the use of worked examples, the testing effect. Participants learn about the approach/technique, perhaps do some work at the workshop on designing some instruction that will use the approach/technique, and then go back to their daily routines. Our experience is that such workshops, if well designed and executed, can be highly motivating. Teachers may leave such workshops with the best of intentions. However, for a variety of reasons (e.g., insufficient time to develop the new instruction, problems that come up when applying the new approach), systematic follow through does not often occur.

In our approach, we work with teachers from the beginning to the end of their project. We help them to operationalize the learning issue they want to address; we identify the principle(s) that will be used to address their issue; we design the intervention that will be used (including the development of a strong experimental design to evaluate the contribution of the intervention to any changes in learning that are observed); we work with teachers to implement the intervention (e.g., we redesign multimedia slide presentations to adhere to Mayer's [this volume] multimedia learning principles); we administer and score online quizzes intended to produce a testing effect; we collect, analyze, and report data; after a course is over, we consider developing a follow-up project and/or giving teachers suggestions for working on their own.

We do not expect that our work with faculty will make them experts in science of learning; however, some faculty with whom we have worked have taken a keen interest in science of learning and, over the course of our work with them, have become colleagues in our research efforts to develop, implement, and assess the impact of science of learning principles in academic courses.

We do expect that, as a result of their work with project staff, teachers will be able to use on their own the approach/technique that was the topic of our project with them. We have seen this to be the case with most teachers with whom we have worked.

Principles and Interventions

Our projects with teachers have involved the application of one or more principles. Below are some examples.

- Desirable Difficulties (Clark & Bjork, this volume)
- Test-enhanced Learning (Pyc, Agarwal, & Roediger, this volume)
- Generation Effect (Bertsch & Pesta, this volume)
- Retrieval Practice (Nguyen & McDaniel, this volume)
- Self-Explanation (Chiu & Chi, this volume)
- Distributed Practice (Carpenter, this volume)
- Interleaving (Carpenter, this volume)
- Expertise Reversal Effect (Lee & Kalyuga, this volume)
- Worked Examples (Renkl, this volume)
- Coherence Principle (Mayer, this volume)
- Elaborative Interrogation (Pazicni & Pyburn, this volume)
- Transfer Appropriate Instruction

A selection of interventions based on some of the above principles and others include: A teacher of a course in exercise science wanted her students to do better at learning and remembering anatomical structures so they could develop better exercise programs for athletes. We thus developed an intervention based on the testing effect. A teacher of an advanced psychology course wanted her students to perform better on exam essay questions that required conceptual understanding of material. We thus developed an intervention that required they write answers to *guiding questions* while working on assigned readings. A teacher of a civil engineering course wanted his students to perform better on exams that required them to solve structural engineering problems. We thus developed an intervention in which the teacher developed sets of homework problems informed by research on worked examples. Several teachers wanted to improve the *PowerPoint*© slides they used in their courses. We thus worked to redesign their existing slides based on "research-based principles for designing multimedia instruction" (Mayer, this volume). Several colleagues who teach biological science courses to first-year students were concerned about the poor study skills of many of their students. We thus developed and have administered a study skills training module that includes instruction on many of the principles listed above. And so on.

To learn about the above and additional principles, visit the University of Memphis Department of Psychology website and select the title "25 Learning Principles to Guide Pedagogy and the Design of Learning Environments" (http://www.memphis.edu/psychology/teaching/content/misc.php).

We incorporate a research design into most of our projects. As noted in the introductory chapter to this volume, although principles related to learning, retention, and transfer of information and knowledge have been well-established in laboratory studies, with regard to many of those principles, little has been done in real academic settings. For example, although we know a lot from lab-based experiments about effects of interleaving of practice and of testing we have been interested in how these techniques operate in authentic learning environments (college and university academic courses). Doing research in academic courses, however, is "noisy," meaning we lack the kind of tight experimental control that is the hallmark of lab-based research (cf., Daniel, 2012). In addition, simply using a technique (e.g., quizzing) and measuring students' performance on a later high stakes exam does not allow us to isolate the contribution of quizzing to any exam performances we observe. As illustrated in the three studies described later in this chapter, we have incorporated one or more "control" conditions to assess possible effects of our interventions. Once we firmly establish that our interventions are having a desired effect on students' academic performance, we then remove the control conditions and urge teachers to use the principles that have been shown to work. Our empirically- and experimentally-grounded approach stands in contrast to the approach that often guides teachers' selection and use of teaching methods—namely, "this method seems to work well." When we hear teachers make this type of statement, we are always wondering, "compared to what?"

Example Projects

Three chapters in the final section of this volume present descriptions of studies that were part of the Cognition Toolbox project: The testing effect: an intervention on behalf of low-skilled comprehenders in general chemistry; Using guiding questions to advance reading comprehension and student learning; Applying multimedia principles to slide show presentations. These studies examined important learning outcomes: problem solving; problem-solving and conceptual essay questions; conceptual essay questions; and factual and conceptual questions, respectively. In the project described below, the learning outcome measure was exam performance based on multiple-choice questions (factual, conceptual, application questions). Because of the large enrollment in this course, grading short-answer or essays questions was not feasible.

Effects of Online Quizzing in a Large Enrollment Course on the Science and Nature of Stress

The instructor (Dr. Barbara White) of a course on *Science and Nature of Human Stress* contacted project staff and told us she had been giving weekly online quizzes for several semesters. She asked us to work with her to assess whether these quizzes were actually helping students to learn course material and whether taking these quizzes had a measurable impact on midterm and final exam performance.

After meeting with the instructor, we determined that our first effort would focus on assessing whether students performed better on the midterm exam on topics for which they answered quiz questions compared with exam performance on topics for which they were not quizzed. If there was no educationally important difference, we might want to reconsider whether quizzing was worth the time and effort. If there was a difference, we could then go on to explore what it was about the quizzing that made the difference.

The studies reported below represent an exploration of the well-documented "testing effect." Taking a test after being exposed to material has been shown to have a positive effect on the retention of that material at a later date (compared to the retention of material that is not tested or material that is restudied) (Pyc, et al., this volume).

Study 1

We randomly assigned half of the 132 students enrolled in the course to either Group A or Group B in the course *Blackboard©* site. During each of the five weeks leading up to the midterm exam, students completed a multiple-choice quiz online via the *Blackboard©* site. Each week, students in Group A and Group B completed a quiz consisting of 12 questions. For both groups, six of the questions were unique to their group's quiz, and the remaining six questions appeared on the quiz for both groups. The quizzes consisted of material covered in assigned reading for the week. After a weekly quiz became unavailable, students were permitted to review the correct answers to each quiz item in *Blackboard©*. Quiz scores contributed to approximately 20% of students' final grade in the course.

During week six of the semester, students completed a midterm exam consisting of 70 questions. There were 9 questions randomly selected from Group A's quizzes and 9 questions that were randomly selected from Group B's quizzes. The exam questions for these items were rewritten in the form shown below.

Question as it appeared on the quiz (correct answer in blue)	Question as it appeared on the midterm exam (correct answer in blue)
The *Yerkes-Dodson principle* suggests that: A. not enough stressors can put someone at risk for disease. B. **a certain amount of stress is linked with adaptive, competent performance** C. after a certain point, stress can improve one's health D. all stress is bad for your overall health.	What principle suggests that a certain amount of stress is linked to adaptive, competent performance? A. the self-regulation principle B. the ACTH principle C. **Yerkes-Dodson** D. none of the above are correct
If your mom did not get enough to eat during her pregnancy, your metabolism would likely adjust in what way to deal with the external environment? A. Your metabolism does not prepare you well to deal with life outside of the womb.....you might tend to be overweight or skinny dependent ONLY on how much you eat. B. It probably doesn't have much of an effect on you later. C. It might become "thrifty" meaning that you are not very efficient in calorie use and are likely to be skinny. D. **It might become "thrifty," meaning that you use every nutrient efficiently. This could lead to obesity later in life if you are around plenty of food.**	Your metabolism might become "thrifty," meaning that you use every nutrient efficiently, if during her pregnancy your mother: A. was highly stressed B. was depressed C. **had inadequate food intake** D. gained a lot of weight.

Thus, there were 18 items on the midterm exam that had been quizzed for either Group A or Group B. We also calculated the mean midterm exam score for questions that did not appear on either of the groups' quizzes and for midterm exam items that appeared on both of the groups' quizzes.

We were interested in determining whether students performed better on midterm exam items for which they received quiz items on the same content. For Group A, students performed better on exam questions for which they had received quiz items (M = .89, SD = .12) than on exam questions for which they did not receive quiz items (M = .72, SD = .17), F (1, 65) = 72.90, p < .001, η_p^2 = .53. For Group B, students performed better on exam questions for which they had received quiz items (*M* = .85, *SD* = .16) than on exam items for which they did not receive quiz items (*M* = .77, *SD* = .19), F (1, 65) = 11.17, $p <$ .001, η_p^2 = .15.

Students in Group A and Group B performed equally well on exam questions that were not quizzed for either group, on exam questions that were quizzed for both groups, and on their overall exam scores (*p*s ranged from .15 to .83). Thus, quizzing worked: Group A performed better on exam questions that were covered on quizzes, as did Group B. These results were consistent with lab-based findings on the testing effect (see Pyc, et al., this volume, for a review of this literature).

One obvious possible explanation of our results, unrelated to the testing-effect, is signaling. Group A and Group B received different quiz items. Perhaps the items they received served as a signal to them to especially focus on the material covered on their quiz. ("The teacher quizzed us on this material, so she must think it is important and will cover it on the exam.") This is not an unreasonable speculation. However, to the extent that students focused on some material (quizzed material) and not on other material (not quizzed material), their exam performance *should* be lower on the latter material. After consultation with the instructor, we explored this possibility in Study 2.

Study 2

In our second project, we switched from a between-students design to a within-students design. That is, each student was exposed to each experimental condition of the study (see Gurung, this volume, for more details on this and other experimental designs).

In this study, we wanted to ensure that students were exposed during quizzing to *all* of the topics that might be covered on the midterm exam (to avoid cueing as a possible explanation for the results we found in Study 1).

As in Study 1, there were weekly quizzes leading up to the midterm exam. There were 167 students who completed all the weekly quizzes and took the scheduled midterm exam. The quizzes were administered via the course *Blackboard©* site, as in Study 1. For each quiz, students answered two types of questions. For one type of question (recall quiz condition), students received a question that required them to type an answer. After they provided an answer, or indicated that they did not know the answer, they clicked on the Submit button in the *Blackboard©* testing mechanism. A multiple-choice question on the topic covered in the recall question immediately appeared. For example, students received the following recall item:

Cortisol is regulated in large part by what kind of mechanism?

After students typed a response in the textbox, or typed 'don't know,' they clicked the Submit button. As soon as they did, the following multiple-choice question appeared:

Cortisol is regulated in large part by what kind of mechanism? (Correct answer in blue)

A. A positive feedback loop meaning that the system shuts down if cortisol levels become too high.

B. **A negative feedback loop, meaning that when cortisol levels get too high, the system shuts down production.**

C. A neutral feedback loop, meaning when levels get too low, the system increases cortisol.

D. A positive feedback loop meaning when levels get too high, the system produces more cortisol.

For the second type of question (study condition), students received a statement. They were instructed to read the statement and then click the Submit button. As soon as they clicked on Submit, a multiple-choice question appeared. Students were instructed to select the answer that appeared in the statement they just read. Using the item above as an example, students were asked to read the following statement:

Cortisol is regulated in large part by a negative feedback loop, meaning that when levels get too high, the system shuts down production.

After they read the statement, they clicked the Submit button, and the following multiple-choice question appeared:

Cortisol is regulated in large part by what kind of mechanism? (Correct answer in blue)

A. A positive feedback loop meaning that the system shuts down if cortisol levels become too high.

B. **A negative feedback loop, meaning that when cortisol levels get too high, the system shuts down production.**

C. A neutral feedback loop, meaning when levels get too low, the system increases cortisol.

D. A positive feedback loop meaning when levels get too high, the system produces more cortisol.

On the midterm exam, 24 questions appeared that had been presented as recall/multiple-choice items on the quizzes, and an additional 24 questions appeared that had been presented as study/multiple-choice items on the quizzes (using the rewritten items format used in Study 1). In the "study" condition, students were exposed to content and they were then tested on that content using a multiple-choice question. In the "recall" quiz condition, students were required to try to recall a correct answer and then were tested with a multiple-choice question. After students finished and submitted a quiz, they were given access in *Blackboard©* to the entire quiz (recall and study question conditions), their answers and correct answers to multiple-choice questions, and their total score on the quiz. [Note: Because the study condition items provided students the correct answer to the related multiple-choice questions, we expected that students would be virtually perfect in correctly answering these questions and, in fact, they were. Our primary reason for including multiple-choice questions in the study condition was to check on whether students were attending to the material presented in that condition.]

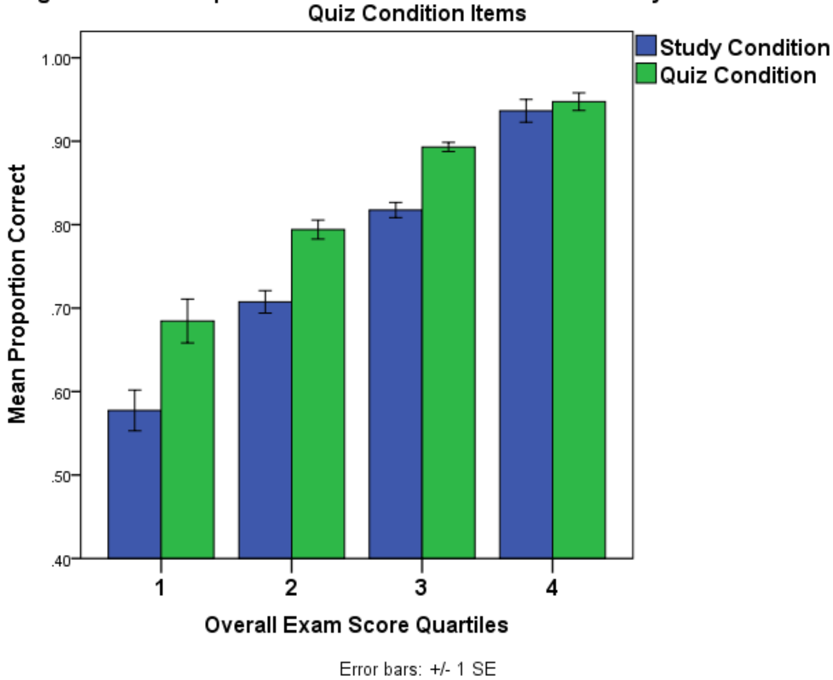
We were interested in students' performance on midterm exam questions that had been either presented as study trials or recall quiz trials on the weekly online quizzes. Students performed, on average, seven percentage points better on exam questions for which they had received recall quiz items (M = .84, SD = .11) than they did on material in the study condition (M = .77, SD = .13), F (1, 166) = 76.48, $p < .001$, $\eta_p^2 = .32$.

This finding is remarkable. For items in the study condition, students received information, were immediately asked a multiple-choice question about that information, and then received correct answer feedback. For items in the recall quiz condition, students attempted to recall the answer to a question, were then asked a multiple-choice question about that information, and then received correct answer feedback. In other words, for study condition items, students were given complete and accurate information. Yet, on the exam, students performed better on items that were in the recall quiz condition. Thus, our result can rule out cueing as a possible explanation for the improved scores in the quizzed “recall” condition.

Because the instructor’s course had a large enrollment, we were able to ask an additional question: was the quizzing condition equally effective for students at all levels of exam performance? As discussed in Lee and Kalyuga (this volume), instructional interventions may not be equally effective for low-skill and high-skill students. Perhaps quizzing has more positive impact for students with poorer background knowledge or skill. If that turned out to be the case, the course instructor’s practice of administering online quizzes may have the added benefit of helping students who need it the most. We address this possibility below.

We assigned students to one of four groups (quartiles) based on their overall exam score. As shown in Figure 1, the lower the overall exam performance, the larger the benefit for exam items in the recall quiz condition relative to items in the study condition. From bottom to top quartiles, students performed 11, 8, 7, and 1 percentage point(s) better on exam items that were in the quiz condition.

Figure 1. Mean Proportion Correct on Midterm Exam for Study Condition and Quiz Condition Items



Studies 1 and 2, established that quizzing had an educationally positive impact on students’ exam performance in professor White’s *Science and Nature of Human Stress* course. This finding supported a large body of prior research, most of which was conducted in tightly controlled laboratory experiments (Pyc, et al., this volume). We were pleased to have extended this research by showing a robust testing effect in a real academic course. An important novel contribution of Study 2 was that we found that the testing effect was larger for students whose overall performance on the midterm exam was lower. In other words, quizzing helped students who needed the most help.

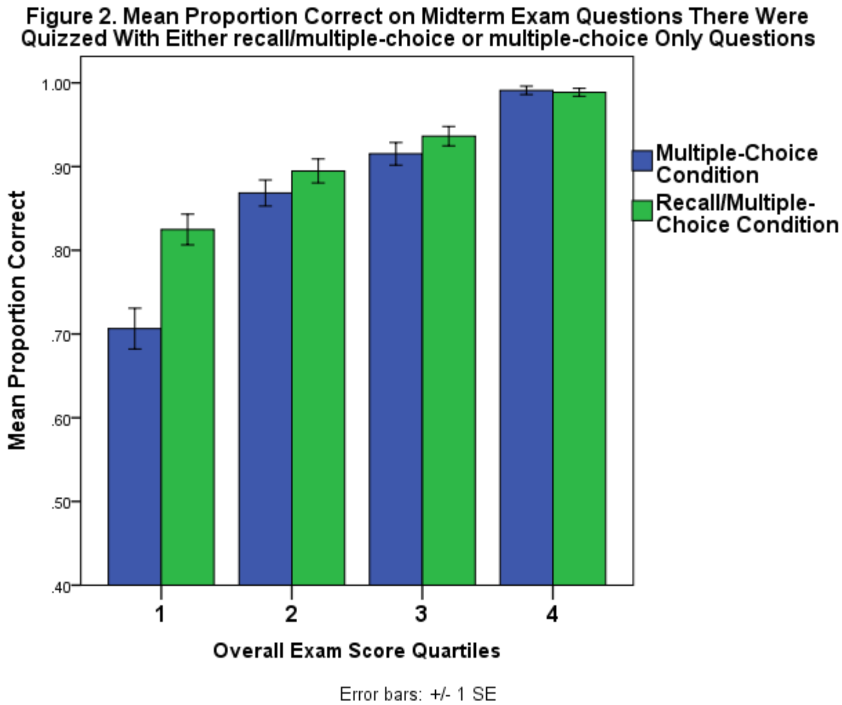
One finding that is often obtained in laboratory studies is that people perform better on a final test when they are quizzed with retrieval quizzes than when they are quizzed with recognition quizzes. That is, recalling the answer to quiz questions will often produce a larger testing effect than answering multiple-choice quiz questions. In our original discussions with professor White, CETL staff suggested that she administer recall-type quiz questions. Although she appreciated the suggestions, she decided against this approach because of the time it would take to score student responses to this type of quiz (which were graded). Her class had more than 150 students most semesters. We addressed this issue in Study 2 by asking students to first attempt to recall the answer to a quiz question and then to answer a multiple-choice question on the item. Only answers to the multiple-choice items contributed to the quiz grade; not the recall answers. As reported above, we found in Study 2 that students performed better on midterm exam questions for items quizzed in this manner (84%) than for items in the study condition (77%).

Study 3

In this Study, as in Study 1 and Study 2, students completed weekly 12-item quizzes, administered via the course *Blackboard©* site, leading up to the midterm exam. There were 156 students who completed all the weekly quizzes and took the scheduled midterm exam. As in Study 2, we again used the recall/multiple-choice condition. For a comparison condition, we included quiz items that were administered in a multiple-choice only format. We wanted to assess whether we could boost midterm exam performance for items that were quizzed in a recall/multiple-choice format relative to items that were quizzed with only a multiple-choice question. We also included on the midterm questions that were not quizzed in any format. The weekly quizzing and midterm exam protocols were the same as in Study 1 and Study 2.

Students performed better on midterm exam questions for which they had received recall/multiple-choice quiz items (M = .91, SD = .10) than they did on questions for which they had received multiple-choice only quiz items (M = .87, SD = .15) and on questions for which they were not quizzed (M = .77, SD = .08), F (2, 310) = 137, p < .001, η_p^2 = .47. Each mean differed significantly from each other mean, all *p*s < .001. Most important, students performed four percentage points better on exam items in the recall/multiple-choice condition than on items in the in the multiple-choice only condition, F (1, 155) = 15.48, p < .001, η_p^2 = .09.

As in Study 2, we asked whether between-condition effects were larger for the students whose overall performance in the midterm exam was relatively poor. We assigned students to one of four groups (quartiles) based on their overall score on the exam. As shown in Figure 2, the lower the overall exam performance, the larger the benefit for exam items in the recall/multiple-choice condition relative to items in the multiple-choice only condition. From bottom to top quartiles students performed 12, 3, 2, and 0 percentage point(s) better on exam items that were in the recall/multiple-choice condition.



This project represents an excellent example of how we address a learning issue posed by a teacher. Over three offerings of the course, we learned that online quizzing in a large enrollment general education course had systematic positive effects on midterm exam performance. An important new finding was that the effects of quizzing were larger for students who performed less well overall on the exams. In a different project (Pazicni & Pyburn, this volume), we found that quizzing had a larger positive effect on exam performance in a general chemistry course for students with relatively poor reading comprehension skills.

Dissemination

We have worked closely with colleagues at two postsecondary institutions in the New England region of the USA—University of New England (Biddeford, Maine) and Western New England University (Springfield, Massachusetts). Professor Jennifer Stiegler-Balfour and Professor Christopher Hakala have invited project staff to their institutions (UNE and WNEU, respectively) to develop projects at their universities.

In an effort to develop interest in our approach at other institutions, we have met with faculty and administrators at colleges/universities in New England (Lesley University, Merrimack College, Massachusetts College of Pharmacy and Health Sciences University, Wheaton College, Clark University) and beyond (Westminster College of Salt Lake City; Utah Valley University). We have also given presentations to college and university academic leaders at New England Association of Schools and Colleges annual meetings, at American Psychological Association conventions, and at the annual National Institute for the Teaching of Psychology.

Conclusion

We have completed dozens of course-level studies over the past six years, as illustrated by the example provided above and the chapters in the last section of this volume by Pazicni and Pyburn, Stiegler-Balfour and colleagues, and Overson. Many of our projects are conceptual replications of other projects. For example, we have investigated the testing effect in courses in introductory psychology, genetics, science and practice of strength training, health care financial management, and statistics in psychology. One of the questions that we had when we began our project was whether the science of learning principles that we were going to apply would "work" the same way as they do in laboratory studies. As noted in our introductory chapter, most of these principles have a strong base of support in tightly controlled laboratory experiments. Few of them have been systematically examined in actual academic courses (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Daniel (2012) recently argued that much additional course-level translational research is needed before "promising principles" should be promoted to teachers as "best practices" (p. 251). Our work at UNH over the past six years has focused on evaluating a variety of science of learning principles in a wide range of academic disciplines and course formats.

Acknowledgments

This work is supported by grants from the Davis Educational Foundation (The Cognition Toolbox, 2009-2012; Teaching and Learning with Multimedia, 2012-2015). The Foundation was established by Stanton and Elisabeth Davis after Mr. Davis's retirement as chairman of Shaw's Supermarkets, Inc. With the support of the Davis Educational Foundation, we have had the resources required to work with faculty before, during, and after they taught a course included in the project.

Support has also been provided by the Office of the Provost and Vice President for Academic Affairs, University of New Hampshire.

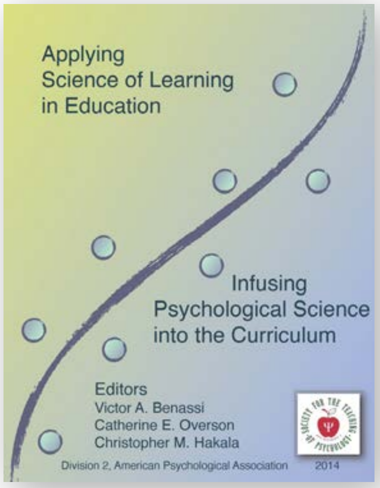
References

Daniel, D. B. (2012). Promising principles: Translating the science of learning to educational practice. *Journal of Applied Research in Memory and Cognition, 1*, 251–253.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. Psychological Science in the Public Interest*, 14*, 4-58.

Halpern, D. F., & Hakel, M. D. (Eds.), (2002). *Applying the science of learning to the university and beyond.* New Directions for Teaching and Learning. San Francisco: Jossey-Bass.

Mayer, R. E. (2010). *Applying the science of learning*. Upper Saddle River, NJ: Pearson.



Applying Evidence-Based Principles of Learning to Teaching Practice: The Bridging the Gap Seminar

Debra Swoboda
York College of The City University of New York

In the past decade, scholars have engaged in interdisciplinary research that relates the fields of cognitive psychology and neuroscience to education, although getting teachers to appreciate and apply the results of this research is an evolving experiment at many institutions. At The City University of New York (CUNY), discussions among Center for Teaching and Learning (CTL) directors about how to equip teachers to utilize this knowledge were propelled by the recognition that “it can be tough to find a balance between concepts drawn directly from ‘hard research’ vs. ideas that are so digested for practitioners’ use that the underlying research is completely obscured” or lead to “just another set of random-seeming ‘teaching tips’” (Girash, 2011, p.1). This chapter introduces the faculty development program to forge this divide − the Bridging the Gap Seminar – a faculty learning community that was implemented at six CUNY campuses to bridge the gap between evidence-based principles of learning and instructional practice. This chapter describes the faculty members who participated in the Bridging the Gap (BtG) Seminar and the format and curriculum of BtG sessions, which focused on common readings from *How Learning Works: Seven Research-Based Principles for Smart Teaching* (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010). In addition, this chapter presents assessment results collected from a pre-post survey and participant implementation plans to unpack the challenges and outcomes associated with participating in the BtG Seminar. These findings are intended for educators interested in applying research findings on learning in their teaching and for faculty development coordinators wishing to develop a similar program or learning community.

Literature Review

Educational neuroscience brings together cognitive science, biology, and linguistics to inform teaching and learning. Although the neurosciences have been raising questions for decades about how people learn (Byrnes & Fox, 1998; Gazzaniga, 2004; Petittot & Dunbar, 2004; Willingham & Lloyd, 2007), there is a growing movement to elucidate the learning process that includes academic societies and journals, as well as a popular industry in brain-related educational workshops marketed to teachers and institutions (Coch, Michlovitz, Ansari, & Baird, 2009). Because this body of work presents a broad array of topical, methodological, and philosophical perspectives with divergent lines of research, distilling its implications for instructional practice is challenging. Nonetheless, this research is consistent in what it suggests for teachers:

> It calls for, among other things, instructional practices that move the instructor away from being the center of the activity and instead placing learners at the center. It focuses on learning outcomes rather than on teaching inputs. It calls for collaboration rather than competition among learners. It indicates a need for teachers to change from their overdependence on traditional lecture to more interactive lecture and other instructional strategies that engage students in their learning. And it encourages teachers to make use of the most current research on human learning (Feden, 2012, p.6)

Educational neuroscience, therefore, provides a scientific foundation for the design of effective instruction.

Although surveys of teachers regarding their opinions of this research indicate that they are enthusiastic about understanding the role of cognition in educational activities (Hook & Farah, 2012; Pickering & Howard-Jones, 2007), teachers do not simply want to be told what works (Goswami, 2006), but instead want to refine their understanding using this research to decide what works in their own classrooms. Thus, the goals of any faculty development program on this research are to help teachers make sense of this body of work and to recognize the instructional implications of how people learn.

Bridging the Gap Seminar Design

Participants

During the 2012-13 academic year, six CUNY CTLs – including York College, the campus where the author served as CTL director – offered BtG programs. A total of 46 faculty members participated in the CUNY BtG Seminar, 6-12 participants in each cohort. Participants in the cohorts reflected a diversity of ranks and department affiliations and included faculty members from across the disciplines as well as from the professional programs (e.g., Business, Health Professions). On some campuses, an open invitation to participate was extended to all faculty members. In other cases, faculty members competitively applied to participate and in some cohorts, participants received a small stipend ($500) for their BtG participation. At York College, for example, nine teachers were selected from two dozen applicants to participate in the BtG Seminar, including two veteran faculty members, a number of professors without tenure, a librarian with classroom teaching responsibilities, and a lecturer. The courses chosen by participants across the cohorts for Seminar discussion and strategy implementation (participants were asked to focus on teaching and learning in a single course they were teaching) also were diverse in nature, ranging from large introductory courses to capstone courses to online/hybrid courses.

The BtG Curriculum

All CUNY BtG Seminars followed a similar syllabus and assessed participants' experiences using similar instruments and procedures. The BtG Seminar typically consisted of five 2-hour sessions (a small set of sessions was deemed the best format to obtain faculty participation), facilitated by the campus CTL director. Common readings included chapters from *How Learning Works: Seven Research-Based Principles for Smart Teaching* (Ambrose et al., 2010), although additional scholarship describing educational theory and research (e.g., scholarly articles, website links) was used to supplement ideas discussed in the text when appropriate. Although the sum of chapters selected from the text for discussion varied across cohorts, all cohorts at minimum discussed three topics: 1) understanding the role of prior knowledge in learning; 2) facilitating deep vs. surface learning and the transition from novice to expert learner; and 3) enhancing learning through practice and feedback. Each CUNY campus BtG Seminar added to this base of topics as desired. For example, while the York College cohort discussed additional chapters on the topics of motivation and metacognition, other CUNY cohorts discussed chapters on the topics of course climate or self-directed learning.

A set of achievable and assessable outcomes was established for the BtG program. In each of the cohorts, participants were asked to:

1. Explore and discuss empirical evidence and research-based theories about prior knowledge, deep vs. surface learning, and practice and feedback, drawing on the multiple disciplines of cognitive science— including psychology, neuroscience, linguistics and education;
2. Reflect on their own teaching practices in the context of research on how people learn;
3. Design a new pedagogical approach or strategy for implementation in an upcoming course; and
4. Implement a new pedagogical approach or strategy the semester after the seminar and assess its impact.

These learning outcomes in turn provided a blueprint for assessment of participants' experiences in the BtG faculty development program.

The BtG Seminar also operated using a similar format for delivery across all campuses. Each BtG session began with discussion of situations presenting student learning challenges familiar to all teachers (e.g., utilizing prerequisite knowledge, demonstrating higher-order understanding, and expressing adequate written communication skills). Working from a particular chapter in the Ambrose et al. text, participants then explored a principle of learning and the supporting research informing how learning works to explain these challenges. In the course of discussion, the facilitator (the CTL director) acted to ensure that the participants understood the specific concepts and ideas informing the learning principle being examined. This meant that in practice, discussion of the concept of prior knowledge, for example, began by exploring vignettes that illustrated the idea that learning requires activation of accurate, sufficient, and appropriate prior knowledge. Aided by the facilitator, participants then discussed the principle that "students' prior knowledge can help or hinder learning" (Ambrose et al., 2010, p.13) and the implications of this principle for identifying strategies that ameliorate inadequacies in students' prior knowledge. Similarly, discussion of deep vs. surface learning began by examining classroom scenarios illustrating the challenges students face in organizing and connecting elements of information in meaningful ways, and how experts organize knowledge differently than novices. Participants then deliberated the learning principle that "how students organize knowledge influences how they learn and apply what they know" (p.44) and identified strategies to be used to help students organize knowledge in deeper and more meaningful ways. Discussion of the concepts of practice and feedback likewise began by having participants consider challenging teaching situations best explained by the principle that "goal-directed practice coupled with targeted feedback enhances students' learning" (p.125). Through facilitated discussion, participants then examined the research implications of this principle and identified strategies to enhance learning through practice and feedback outlined in the text.

To help BtG participants view what they learned as more than just another set of teaching tips, participants completed assignments for most sessions. One assignment, for example, asked participants to evaluate the sources of students' prior knowledge by having them make a list of the information and skills they expected students to have upon entering one of their courses and where students should have acquired this prior knowledge. Another assignment involved having participants create a cognitive map of the central concepts and ideas they believed were important to learn in one of their courses and asking students in the course to do the same for the purpose of comparison. Yet another assignment asked participants to reflect on the ‘expert processes’ they routinely used to analyze a primary source in their field (e.g., vocabulary use, concept understanding, critical thinking, and synthesis) and how they had acquired these abilities in order to explicitly describe these processes to students.

In addition to these assignments, participants were asked to develop an implementation plan describing a teaching approach or strategy to improve student learning to be employed in one of their classes the following semester. For their implementation plans, participants were asked to outline:

1. The specific approach or strategy to be used (it could be more than one and could include several iterations);
2. The learning principle(s) underpinning the approach/strategy (e.g., prior knowledge, deep vs. surface learning, practice and feedback);
3. he expected learning outcome; and
4. An assessment plan (evaluation of how the approach/strategy improved student learning).

Asking participants to identify the approach/strategy they would use to improve student learning and the principle(s) informing why this approach/strategy was appropriate was intended to help them digest the principles and underlying research discussed. In addition, expecting participants to evaluate if their use of a new or different approach/strategy improved student learning was expected to help them realize that teaching and learning improvement is a process that requires assessment.

Outcomes Assessment

In order to evaluate the impact of the BtG Seminar on participants, assessment results were obtained through a pre-post survey as well as from participant implementation plans. Surveys assessed participants’ knowledge and practice of learning principles and strategies, significant learning experiences, and opinions concerning BtG program design. Forty-six BtG participants across the CUNY cohorts completed both the pre- and post-survey. The survey administered at the outset of the BtG Seminar consisted of 23 questions using a 5-point Likert scale (i.e., ‘strongly disagree’ to ‘strongly agree’) or a 3-point scale (i.e., ‘a great deal’ vs. ‘some’ vs. ‘not at all’) asking participants to rate their understanding and classroom application of the concepts of prior knowledge, knowledge organization, and practice/feedback. The pre-survey also contained three open-ended questions asking participants to report specific strategies they used to address each concept. The survey completed at the end of the Seminar included 33 questions asking participants to rate the knowledge they gained about these three concepts and related strategies on a 3-point Likert scale (i.e., ‘a great deal’ vs. ‘some’ vs. ‘not at all’). For the purpose of pre-post comparison, many of these questions were similar to those asked in the initial survey. The survey also contained five questions using a 5-point Likert scale (i.e., ‘very high’ to ‘very low’) and five open-ended questions assessing participants’ significant learning experiences and their attitudes about BtG program design. In addition to analyzing the pre-post survey data, implementation plans were reviewed to evaluate how participants applied BtG knowledge and practices. Statements made in response to open-ended questions were triangulated with the above data.

Knowledge and Intentions

Pre-post survey results indicate that participants developed knowledge about learning principles and planned to implement strategies to improve student learning as a result of their participation. Regarding the notion of prior knowledge, at the outset of the BtG Seminar a majority of the respondents (61%, n=28) classified over a third students in their classes as unprepared due to insufficient, inappropriate, or inaccurate prior knowledge. Even so, less than half of the respondents (41%, *n*=19) reported that they employed activities at least some of the time to attend to these prior knowledge deficits. In contrast to these findings, post-survey respondents reported that they learned a great deal about the impact of prior knowledge on learning and 83% (*n*=38) of them reported they intended to utilize various strategies (e.g., diagnostic tests, brainstorming sessions, linking new and previous material) to address prior knowledge deficiencies at least some of the time. Responses to open-ended questions support these findings. Respondents stated that discussion of prior knowledge taught them to "ask students to tell what they already know before getting into the material" and to "not make assumptions about students' background knowledge or skills." One respondent indicated this shift in understanding by stating, "Reflecting on what I assumed students knew versus what they actually know was eye-opening. I think reading about blind spots really helped me understand why some things were happening in my classes."

Pre-post survey results regarding the concept of knowledge organization also reflect a change in attitudes and intentions regarding teaching practices over the course of the Seminar. In the initial survey, many respondents (67%, *n*=31) reported that they took adequate steps to support students' organization of their knowledge by providing detailed syllabi, distributing handouts, affording opportunities for revision, or by teaching study tips. At the same time, however, a significant portion of pre-survey respondents (20%, *n*=9) stated that they did not understand what 'knowledge organization' meant. In addition, a few respondents (13%, *n*=6) acknowledged their deficiencies in addressing students' knowledge organization. Statements such as "I do let students know how the course will be organized, but beyond that, I'm not doing much," and "I focus a lot on constructing activities that help students learn content but I spend little or no time on how students organize information…or model what expert knowledge is and how to develop it" reflect these deficiencies. Post-survey results, however, show respondents' newfound appreciation of the role of knowledge organization in learning. As one respondent noted, "Before the BtG program, I had little or no knowledge of strategies that enable students to see the 'big picture' and link theories, concepts and related information." Results also reflect respondents' recognition that they need to augment their repertoire of teaching strategies to help students structure information. In fact, 100% of respondents (*n*=46) indicated that they would use various strategies to improve students' knowledge organization (e.g., concept maps, examples of shallow vs. deep knowledge) at least some of the time. Several respondents also acknowledged that they would implement approaches to support student learning of 'expert processes,' stating that they were more attuned to "the differences between novice and expert thinking patterns" and "the importance of viewing assignments and tasks from students' perspectives."

Pre-post survey results regarding the concepts of practice and feedback indicate that at the outset of the Seminar respondents believed they were using sufficient best practices, but at the close of the Seminar believed they should employ additional pedagogies more often. In the pre-survey, most respondents (85%, *n*=39) reported that they provided adequate tools at least some of the time for students to practice what they learned, such as examination, discussion, and group work, and a majority of respondents (63%, *n*=29) reported that they provided frequent, specific feedback at both the individual and group level at least some of the time. In comparison, however, 89% (*n*=41) of post-survey respondents reported that they learned a great deal about the role of feedback in learning and 91% (n=42) of them indicated they intended to provide more opportunities for students to practice what they learned at least some of the time. "I learned that establishing goals and giving exams was not enough because students need more time and exercises to really learn what they should," concluded one respondent. These findings suggest that the BtG Seminar advanced participants' understanding that deep learning requires sufficient and targeted practice and feedback. As one respondent stated, "Understanding the implications of the research on learning helped me be more explicit about performance criteria and the strategies I can use to help students transfer and apply their learning."

Teaching Approaches and Strategies

Implementation plans submitted by Seminar participants (*n*=32) indicate that they intended to use the knowledge they gained in the BtG Seminar to implement approaches and strategies linked to evidence-based learning principles in an intentional and explicit fashion. Many of the plans outlined the implementation of more than one approach or strategy, and several plans discussed implementation in terms of more than one learning principle. The majority of these plans (53%, *n*=17), however, primarily outlined participants' intentions to apply approaches increasing practice and feedback, bolstered by statements that they understood the need to provide students with ample practice of tasks, with feedback at each step. In this set of implementation plans, participants proposed use of approaches that included instituting exam wrappers, scaffolding assignments, employing peer grading, and having students keep journals, among others. A smaller portion of the submitted plans (28%, *n*=9) primarily outlined participants' intentions to focus on students' knowledge organization, most often by providing more facilitation than lecture. Approaches identified included "being more explicit about what they did and did not want students to do for assignments," "modeling successful examples of coursework," "discussing a task in the framework of becoming an expert," or "spending more class time discussing the process of and rationale for assignments." The rest of the submitted implementation plans (19%, *n*=6) primarily discussed participants' recognition that students' prior knowledge affects their comprehension of course material, expectations for success, and performance in examinations. These plans typically outlined participants' intentions to use diagnostic assessments or concept maps to assess students' prior knowledge.

The implementation plans also described a range of assessment instruments, both quantitative and qualitative, that participants intended to use to measure changes in student learning in response to their implementation efforts. Some of these assessment methods included: 1) reviewing the content of student self-report questionnaires administered at the beginning, middle, and end of classes; 2) evaluating students' reflective essays on learning strategies; 3) analyzing students' annotated portfolios describing their understanding of concepts; and 4) examining trends in the amount of time students report they spent preparing for exams and comparing these preparation tallies to exam scores.

In addition to above description of the various approaches and strategies to be implemented, a significant portion of the plans (34%, *n*=11) also discussed participants' efforts to "rethink the syllabus" and revise course learning goals and link assignments to goals more explicitly. In describing their implementation plans, participants cited their awareness of the importance of creating clear, appropriate learning outcomes in course design, and the challenges of mapping learning outcomes to assignments and exams in ways that reflected what they wanted students to learn. In the words of one participant, "I had not thought deeply about how to translate learning objectives into activities and evaluations but my new course explicitly ties these together and informs students how each piece of the class relates to one or more of the objectives."

Scholarly Inquiry

Responses to open-ended survey questions regarding significant learning experiences, as well as statements made during Seminar sessions and in implementation plans, indicate that the BtG Seminar had a profound impact on participants' ideas about teaching and learning. Comments reflect that participants were enthusiastic about understanding the role of cognition in instruction, and several participants stated that they enjoyed the scholarly process of examining the links between the learning principles and associated research. Statements indicate that participants also appreciated being "somewhat guided through the research." In this respect, they found the Ambrose et al. text to be both "scholarly" and "practical," a comprehensive but accessible discussion of the neuroscience of learning and its implications. At the same time, participants did not simply accept what was presented in the text, but rather sought to refine their understanding. Their responses suggest that they evaluated how well the learning research applied to their students and determined for themselves which strategies and approaches were appropriate to address the challenges they faced in their courses.

Responses also indicate that participants valued the way the Seminar operated as an inquiry process − a process of evaluating the connections between research outcomes and instructional practice in order to select strategies that support student learning. For most participants, the text became a baseline of information they could use to evaluate student learning in their classrooms and identify what they needed to do to help students engage in deep learning. Statements reflect, for example, that participants believed that studying the research on learning provided a "framework for critically evaluate teaching practices in a much more constructive way" or "the means to stop and think about pedagogy and learning in a concerted manner." Practicing this process of inquiry led one participant, for example, to "want to continue to understand more about what I'm doing, why I'm doing things, and find that there's been an improvement.'' Participants' comments also suggest that they believed that an essential component of the inquiry process was its collaborative nature. Several participants, for example, stated that it was helpful to discuss learning in a collaborative environment because they previously were not aware of "how similar their experiences with student learning were to those of colleagues in completely different areas" or because they "could see that there were teaching strategies that needed to be practiced in all the disciplines." "A significant experience for me," stated one participant, "was the opportunity to learn that my colleagues across the disciplines grapple with some of the very same issues that I grapple with in the classroom." In addition, several participants acknowledged the importance of routinely assessing student learning as an essential aspect of inquiry into teaching and learning, and their intention to continue to "ask questions, make observations, and find solutions." These responses indicate that the process of inquiry employed in the BtG Seminar became an organizing framework participants planned to continue to practice.

Discussion

Research on cognition and learning in educational psychology has been productive for advancing scientific understanding and for informing educational practice. One of the promising ideas in this research is that principles of learning can serve as the foundation of effective instruction. Nonetheless, faculty development efforts to have teachers apply these findings have only just begun to occur. This chapter describes such an effort – the BtG Seminar – designed to help teachers develop understanding of the cognitive and neuroscience research on learning and use this knowledge to implement approaches and strategies that promote deep learning.

Outcomes regarding the knowledge and practices developed by participants in the BtG Seminar indicate that the Seminar can serve as an instructional model of how to help teachers apply evidence-based principles of learning to instruction in an intentional and explicit fashion. Findings suggest that teachers hoping to appropriate practical information from the dense body of neuroscience research on learning are well-served by the use of a text – in this case, *How Learning Works: Seven Research-Based Principles for Smart Teaching* (Ambrose et al., 2010) – that distills and organizes this information for utilization.

Participant evaluations of the BtG Seminar also indicate that unpacking learning principles and their implications for effective instruction is best done through an inquiry process. Results show that teachers enjoy the process of scholarly inquiry involved in appraising the practical implications of educational research, but they also want to be able to articulate and refine the strategies and approaches that work best for their students and courses. In addition, findings suggest that deepening teachers' comprehension of the research on learning is best facilitated through collaborative discussion of the challenges faced by all instructors, as well as through completion of assignments and implementation plans that require teachers to digest, apply, and assess what they learn.

Encouraging teachers to apply evidence-based principles of learning to teaching practice, as the BtG curriculum was designed to do, also appears to transform teachers' views of effective pedagogy and their role in the teaching process. As results from BtG pre-post surveys and implementation plans attest, strategy implementation that increases practice and feedback, improves knowledge organization, or assesses prior knowledge leads teachers to employ active learning pedagogies. Statements from several BtG participants also show that development of a working knowledge of the research on learning leads teachers to examine their role in the classroom and move from seeing themselves as a lecturer or dispenser of knowledge to seeing themselves as a facilitator of student learning.

The outcomes of the BtG Seminar presented in this chapter suggest that improvement of teaching practices requires transformation of teachers' understanding of the principles of learning, and that cognitive and neuroscience research can assist in identifying these principles. These findings are intended for educators interested in applying research findings on learning in their teaching and for faculty development coordinators wishing to develop a similar program or learning community.

Acknowledgements

The author would like to thank the other CUNY CTL directors who facilitated BtG Seminars on their campuses and shared their findings (Janine Graziano-King, Kingsborough Community College, Julia Jordan, City College of Technology, Nelson Nunez-Rodriguez, Hostos Community College, Michelle Piso, LaGuardia Community College, Gina Rae-Foster, Lehman College) and the CUNY administrators and others who provided support for this faculty development program (Mosen Auryan, Hunter College, Harriet Shenkman, Bronx Community College, Karrin Wilks, CUNY Dean for Undergraduate Studies).

References

Ambrose, S., Bridges, M., DiPietro, M., Lovett, M., & Norman, M. (2010). *How learning works: Seven research-based principles for smart teaching.* San Francisco: Jossey-Bass.

Byrnes, J., & Fox, N. (1998). The educational relevance of research in cognitive neuroscience. *Educational Psychology Review, 10*(3), 297–342. doi:10.1023/A:1022145812276

Coch, D., Michlovitz, S., Ansari, D., & Baird, A. (2009). Building mind, brain, and education connections: The view from the upper valley. *Mind, Brain, and Education, 3*(1), 27-33. doi:10.1111/j.1751-228X.2008.01050.x

Feden P. (2012). Teaching without telling: Contemporary pedagogical theory put into practice. *Journal on Excellence in College Teaching, 23*(2), 5-23.

Gazzaniga, M. (Ed.). (2004). *The cognitive neurosciences III* (3rd ed.). Cambridge, MA: MIT Press.

Girash, J. (2011). Talking with faculty about cognitive science and learning. *Essays on Teaching Excellence, 22*(3), 1-5. Retrieved from http://www.podnetwork.org/publications/teachingexcellence/10-11/V22_N3_Girash.pdf

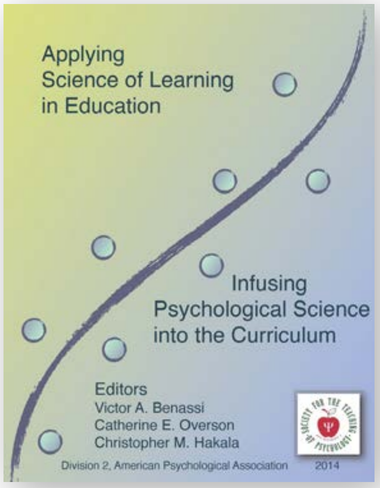
Goswami, U. (2006). Neuroscience and education: From research to practice? *Nature Reviews Neuroscience, 7*(5), 406-413. doi:10.1038/nrn1907

Hook, C., & Farah, M. (2012). Neuroscience for educators: What are they seeking, and what are they finding? *Neuroethics, 6*(2), 331-341. doi:10.1007/s12152-012-9159-3

Petitto, L., & Dunbar, K. (in press). New findings from educational neuroscience on bilingual brains, scientific brains, and the educated mind. In K. Fischer & T. Katz (Eds.), *Building usable knowledge in mind, brain, and education*. Cambridge: Cambridge University Press.

Pickering, S., & Howard-Jones, P. (2007). Educators' views on the role of neuroscience in education: Findings from a study of UK and international perspectives. *Mind, Brain, and Education, 1*(3), 109–113. doi:10.1111/j.1751-228X.2007.00011.x

Willingham, D., & Lloyd, J. (2007). How educational theories can use neuroscientific data. *Mind, Brain, and Education, 1*(3), 140–149. doi:10.1111/j.1751-228X.2007.00014.x



Helping Students to Get the Most Out of Studying

Stephen L. Chew
Samford University

This chapter explores how teachers can help students become more effective learners. I first outline the obstacles and challenges that students encounter to developing a deep understanding of concepts, such as misconceptions, multi-tasking, and ineffective study strategies. Then I discuss the research on how to help students overcome those obstacles, such as deep processing and formative assessments. Teachers can have a major impact on how well their students learn both in terms of course content and in terms of study skills. I argue that effective teachers address both issues.

For students to be critical thinkers, they must first be critical learners. They must have the study skills to learn information efficiently and effectively. Learning efficiently means students must be able to get the optimal amount of learning from their time studying. Learning effectively means several things. First, they must be able to discern the critical concepts from the less important and tangential information. Second, they must develop a schematic understanding of the subject. That means students must not only learn the concepts, but, in addition, they need to develop a framework of connections among the concepts; they need to know how ideas are linked together and what makes them similar or distinct. Third, they must be able to retrieve and apply the information appropriately. They must recognize when the information is relevant to a novel situation, and be able to use the information to reason, make decisions, or solve problems in that situation. Finally, their schematic understanding must be generative; it must lay the foundation for further, more sophisticated learning.

Should teachers be concerned with their students' ability to learn? The answer to that question depends on a teacher's belief about the primary goal of teaching. If one believes that the primary goal of teaching is to present information and that it is the sole responsibility of the students to learn the information, then whether or not students learn is not relevant to the teacher. For these teachers, their primary responsibility is to present accurate information in a clear, interesting, and well-organized manner. If however, a teacher believes that the goal of teaching is to develop student understanding, then whether and how students learn is a major concern. Teaching success is measured by student learning, not by the content of the teaching presentation. For these teachers, learning is a shared responsibility between teacher and student. No matter how brilliant the presentation, if students fail to learn, then the teaching is not successful.

Educational research clearly shows that teachers can have a huge impact on the efficiency and effectiveness of student learning through the design of the pedagogy, learning activities, and assessment methods (Hattie, 2008; Ambrose, Bridges, DiPietro, Lovett, Norman, 2010). The research also indicates that learning is a complex interaction of many factors (Chew, et al., 2009). A problem with any one of the factors can create potential obstacles to student learning and understanding. Thus, the teacher has a role in helping students to diagnose and correct student learning difficulties. Below are some the potential problem areas that can undermine learning.

Student Mental Mindset

Efficient and effective learning starts with a proper student mindset. Several lines of research indicate that student beliefs about learning can have positive or negative consequences on learning (e.g. Dweck, 2002). Teachers must be aware of student misconceptions that undermine learning and try to correct them. They include:

- Learning is fast
- Knowledge is composed of isolated facts
- Being good at a subject is a matter of inborn talent rather than hard work
- I'm really good at multi-tasking, especially during class or studying

Schommer-Aikins (e.g. Schommer-Aikins & Easter, 2006) has identified several beliefs related to poor student performance. First, students believe that learning occurs much more quickly than it really does. They may not understand the difference between skimming a chapter and closely reading a chapter for comprehension. They may believe that a single reading is sufficient for comprehension. Thus, they may plan their study schedule to finish reading all the material for the first time right before the exam, leaving no time for review. Because of their flawed judgment, students start work on assignments or preparing for exams too late to do an adequate job. They rush to finish papers at the last minute and "cram" before exams. Students must learn that there are no shortcuts to reading for comprehension. They must realize that they learn more as they review material that they have already read than they do the first time they read it. They must learn to set study goals with realistic deadlines that include time to spare in case the work takes longer than they estimate. The teacher can help by providing a desired timeline for certain steps of an assignment to be completed. For example, students should complete a draft of a paper three days before the final draft is due, or they should finish reading all the chapters on an exam a week before they take the exam so they can spend the last week in review. Ideally, providing the structure will be sufficient for students to benefit, and the teacher need not monitor or grade students for achieving these intermediate steps.

A second student belief that undermines learning is that knowledge is composed of isolated facts (Schommer-Aikins & Easter, 2006). Students with this belief memorize key definitions in isolation of other concepts. The use of note cards lends itself to this generally poor study strategy because the easiest way to use note cards is to write down and then memorize definitions. Students never develop a connected understanding or how to reason with and apply concepts. Students can use note cards to accomplish deep understanding if they try to connect information on one card to other concepts, but there are other methods of study that reinforce the connectedness of knowledge, such as concept maps. (Berry & Chew, 2008).

Dweck and colleagues (e.g. Dweck, 2002; Blackwell, Trzesniewski, & Dweck, 2007) have demonstrated how student views on the nature of intelligence can have a strong influence on how they study and how they deal with setbacks. Dweck distinguishes between a belief that intelligence is fixed and inborn versus the belief that it is malleable and grows with practice and study. Those who believe intelligence is inborn believe that people are naturally good or bad at a subject. Statements such as "I'm bad at science," or "I'm good at math" illustrate this mindset. A subject either comes easily or with difficulty to a person and no amount of practice can change that. A growth mindset is reflected in a statement like, "I have to work hard at science," and "I've spent a lot of time on math so it comes easily to me." Students with a fixed mindset tend to avoid challenges because failure would be threatening to their self-concept. They put forth less effort in subjects that they feel should be easy for them and they give up easily if a task becomes too difficult. They cope poorly with failure. Students with a growth mindset work towards mastery, are more open to criticism, and are more likely to persevere through challenges and setbacks. We promote a growth or fixed mindset by the feedback we give. Teachers must be aware of these beliefs and encourage a growth mindset through their instruction and feedback.

Multitasking is the curse of the connected, digital world. We are surrounded by the siren song of technology ready to distract us. Students may not understand how harmful multitasking is to the efficiency and effectiveness of their study. Attention, the ability to focus and concentrate, is a major constraint on the cognitive system. The evidence is clear: trying to perform multiple tasks at once is virtually never as effective as performing the tasks one at a time focusing completely on each one (Sanbonmatsu, Strayer, & Medeiros-Ward, Watson, 2013). There are several reasons for this. Divided attention occurs when attention is split between tasks. Divided attention diminishes performance even when people try to ignore the distraction. When students who are supposed to be taking notes on their laptops or tablets choose to distract themselves by shopping or checking social media, they distract not only themselves, but also those around them, diminishing learning for everyone (Sana, Weston, & Cepeda, 2013).

The second problem with multi-tasking is inattentional blindness (Beanland, & Pammer, 2012; Bredemeier & Simons, 2012). Attending to one object in a scene diminishes the ability to perceive other objects in the scene. In other words, we are incapable of detecting all objects in our environment. Attention allows us to select what we will perceive, but it also often prevents us from perceiving much else. As Simmons and his colleagues have shown, observers miss large, salient objects in a scene when attention is directed elsewhere, but they have the sense that they have seen the entire scene (e.g. Bredemeier & Simons, 2012). The problem of not knowing what we missed is that we believe we haven't missed anything. Inattentional blindness is relevant to students when they allow themselves to be distracted. They necessarily miss objects they are not attending to and are unaware that they missed them.

The third problem with multitasking is the attentional blink (Beanland, & Pammer, 2012). Attentional blink refers to the fact that switching attention is time consuming and effortful. When we allow ourselves to be distracted from what we are supposed to be studying, it takes several minutes to refocus on the material and fully concentrate on it. There is no such thing as a momentary distraction.

Cognitive Load and Mental Effort

Attention allows us to select relevant information from irrelevant information, but it also has a second, equally important function. It allows us to concentrate on the relevant information. Concentration, or mental effort, can be another obstacle to student learning, according to Cognitive Load Theory (e.g. van

Gog, Paas, & Sweller, 2010). Cognitive load refers to the amount of concentration required by a task. Some tasks, like doodling, have low cognitive load; other tasks, like taking a final exam, require a great deal of mental effort and have a large cognitive load. Mental effort is the amount of concentration that people have available to them. Mental effort is always a limited resource; a person only has a finite amount of concentration that he or she can use to perform tasks. The key is that the total cognitive load of all the tasks that a person is trying to do cannot exceed the available mental effort. If it does, performance on all tasks will suffer. Tasks with a high cognitive load, like studying conceptually difficult material, require all available mental effort. Any distraction will carry some amount of cognitive load, will take away from the mental effort dedicated to studying, and cause a decrement in learning. Furthermore, if students are presented with difficult concepts or required to complete complex activities that exceed their available mental effort, then they will be overwhelmed and they will not be able to learn. Teachers must be aware of the cognitive load of different concepts and different learning activities. Learning is hard work, but not all hard work leads to learning.

Metacognition and Self-Regulation

Two related concepts that are relevant to student learning are metacognition and self-regulation. Metacognition, also referred to as self-awareness, refers to a person's awareness of his or her own thought processes. In the case of learning, it refers to a student's awareness of his or her own level of understanding of a concept. Students with good metacognition know when they have mastered material sufficiently to perform well on an exam. Students with poor metacognition tend to be overconfident (Dunning, Heath, & Suls, 2004; Askell-Williams, Lawson, & Skrzypiec, 2012, see also Ehrlinger & Shain, this volume). They believe they have a good understanding when their knowledge is superficial and full of gaps. They stop studying before they have truly mastered material. On exams, they do not realize they are missing items because their knowledge is too superficial to realize they are missing nuances and fine distinctions. Thus they often leave exams confident they have done well but are then stunned when they do poorly.

Self-regulated learning is a model developed by Pintrich (e.g. Pintrich, 2004) in which students plan for learning, implement study strategies and monitor their progress, and finally evaluate what they have learned in relation to their goals. Effective learning requires regulation of learning strategy, motivation, and affect. Good self-regulation means that students have multiple strategies that they can bring to bear for learning in a particular context. Self-regulated learning is strongly associated with student learning (Clark, 2012).

Ineffective Study Strategies

Students are often most enamored with study strategies that are the least effective (Rohrer & Pashler, 2010; Arnott & Dust, 2012). For example, Arnott and Dust (2012) found that a massed review activity just before an exam was significantly less effective for student exam performance than spaced review activities after each chapter, but the massed review was overwhelmingly preferred by students. Students often prefer study strategies that are intuitive or easy to do, regardless of whether or not they are effective. For example, the notion of "learning styles," which states that each person learns optimally in one modality, such as visual or acoustic, remain popular despite the lack of any evidence for the validity of such extreme positions (Pashler, McDaniel, Rohrer, & Bjork, 2008).

Cognitive and educational psychologists have established the effectiveness of many study strategies (e.g. Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013), at least in short-term, controlled studies;

but these strategies are often difficult and counterintuitive. Hyde and Jenkins (1969) established that deep, meaningful rehearsal leads to better recall of word lists compared to shallow, meaningless rehearsal, regardless of whether students intend to learn or not. They also established the idea of an orienting task, which is a task that induces participants to rehearse information at a shallow (e.g. checking for a certain letter in the spelling) or deep (e.g. rating the pleasantness of a word) level. Deep level orienting tasks lead to better recall than shallow tasks. This work paved the way for the levels of processing framework (Craik, 2002), which posited that the method of encoding or rehearsal was the critical element for later recall. The lesson from levels research is clear: bad study strategies trump good intentions.

Misconceptions

Students are not blank slates, but possess intuitive or popular beliefs about psychology, many if not most of which are simplistic or incomplete, or completely wrong (e.g. Lilienfeld, Lynn, Ruscio, & Beyerstein, 2009). These beliefs turn out to be highly resistant to change through pedagogy (e.g. Chew, 2005). Some are popular misconceptions promulgated by the media, such as developing total retrograde amnesia through a knock on the head and the existence of paranormal powers such as ESP. Many are intuitive, such as the idea that blind people develop extraordinary hearing to compensate for the loss of vision. Others are just highly confusable or difficult to understand, such as the difference between negative reinforcement and punishment. Such misconceptions are widespread and difficult to correct.

Student Fear and Mistrust

Cox (2011) documents how student fear and misunderstanding can undermine learning. The goals of students often differ from the goals of teachers, and conflict and miscommunication can result, even when both the student and teacher are well intentioned. Students may fail to perceive the goals of an assignment and see it as “busy work” or a hoop to jump through in order to get a passing grade. They may see the teacher as an obstacle to their goal of obtaining a degree rather than a facilitator of learning. They may see critical feedback not as an opportunity to reflect and improve, but a personal attack on their competence. Students shift easily from a learning mindset to one in which they simply want to do whatever is necessary to pass the course, to give the teacher what he or she wants. Learning becomes irrelevant. For optimal learning, the students must have trust in the teacher; they must believe that the teacher wants them to learn and all the assignments and activities are designed to help them learn. Students will work harder and persevere longer for teachers they trust.

Transfer of Learning

The goal of pedagogy is to have students transfer their learning appropriately. Traditionally, this means that students will appropriately generalize their learning from courses to relevant situations and apply the information correctly to help them understand or reason through the situation. While the goal is virtually universal in education, the evidence suggests that it is seldom achieved (e.g. Bransford & Schwartz, 1999; Bransford, Brown, & Cocking, 2000). New learning is highly context dependent. Students learn concepts within a specific context, both in terms of subject matter and in a class, and they think of it within that context (e.g. Thomas & McDaniel, 2007). They do not automatically generalize information to other contexts. Thus, learning may easily become inert, in that students fail to generalize it to applicable situations (Bransford, et al., 2000). Information must be taught and learned in a way that promotes appropriate transfer to relevant situations.

How to Help Students Become Better Learners

Effective teaching that supports student learning is a complex process (Chew, et al. 2009), yet there is much that teachers can do to help students learn more effectively. I've already discussed how to try to correct some of the misconceptions in student mindset. The importance of undivided attention during class and study cannot be overemphasized. Faculty can set policies about inappropriate use of electronic devices during class. Students can be encouraged to turn devices off or put them out of reach during study.

Teachers can develop activities and assignments that promote deep processing and thus enhance learning. Assignments and activities are basically orienting tasks that make students think about the material in certain ways. If the assignments require reflection and meaningful analysis, then the assignment will facilitate learning. If the assignment allows students to complete it with only shallow processing, then it may actually undermine learning and understanding. Although deep processing is desirable for learning, it is also highly effortful. Teachers must be aware of cognitive load in designing pedagogy, activities, and assignments to prevent overwhelming students (Clark, Nguyen, & Sweller, 2006).

Teachers who have developed expertise over many years can easily underestimate the cognitive load required for understanding certain concepts or completing certain activities. It is easy to forget how challenging learning a concept for the first time can be. Teachers must take care to not overwhelm students; they can adjust the amount of information presented, the time allowed for certain activities, or provide scaffolding for certain assignments. Furthermore, teachers must be careful to design their presentations, especially multimedia presentations, to minimize distraction and reduce irrelevant cognitive load. Mayer (2009) has articulated a theory of effective multimedia learning (see also Mayer, this volume). He has demonstrated how the design of multimedia presentations can help or hinder student learning.

Researchers have approached the problem of developing activities that lead to deep processing in two ways. Learning scientists have documented cognitive principles that can guide development of effective pedagogy (e.g. Ambrose, et al. 2011; Dunlosky, et al., 2013; McDaniel & Wooldridge, 2012). These include spacing learning out versus massed learning, interleaving topics, and testing effects (see also Carpenter, this volume; Pyc, Agarwal, & Roediger, III, this volume). The second approach is the scholarship of teaching and learning, which documents and assesses effective pedagogical practices. This can include effective study strategies (e.g. Berry & Chew, 2008) or effective teacher practices (e.g. Bain, 2004; Cox, 2011). Both approaches are useful to teachers interested in improving their effectiveness.

Another method to help students is to instruct them explicitly in effective study strategies (Chew, 2010). When students are explicitly taught a conceptual framework of the cognitive basis of effective learning that includes concepts such as metacognition, self-regulation, and deep study strategies, their learning improves. (Arnott & Dust, 2012; Askell-Williams, Lawson, & Skrzypiec, 2012). I have created a set five brief videos that help teach students the cognitive basis of effective study. They are available at http://www.samford.edu/how-to-study/.

Formative assessments are a powerful means of promoting desirable study habits and learning. A formative assessment can be any low stakes assessment activity that gives both the teacher and the student feedback about the level of student understanding (Clark, 2012). They come in many forms, such as think-pair-share items, conceptests (e.g. Chew, 2005), or classroom assessment activities (e.g. Angelo & Cross, 1993), but their goal is the same, to provide feedback that both teacher and student can use to improve student learning and understanding. When implemented properly, they are a highly effective means of improving student learning (Clark, 2012). Formative assessments can be used to promote improved metacognition, deep learning, and appropriate retrieval and application. They can be used to detect and correct student misconceptions (Chew, 2005). They force students to practice recall, which promotes learning. They can demonstrate appropriate transfer of concepts to novel domains. Furthermore, they can promote student trust and rapport with the teacher. They show that the teacher wants to help students learn. They model the kind of thinking and understanding that the teacher expects before high stakes exams occur. There are a wide variety of formative assessments that can be adapted to virtually any situation (e.g. Hammer & Giordano, 2012).

Teaching effectiveness should be measured in terms of student learning because teachers have a huge impact on it, for better or worse. Poor teaching can actually hinder learning. Truly effective teaching makes student learning almost unavoidable. There is an important distinction between teaching that makes it easy for students to learn and teaching that makes it easy for students to get good grades. Good teachers focus on the former. They do so both by designing effective pedagogy and teaching students to be effective learners.

References

Ambrose, S. A, Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching.* San Francisco, CA: Jossey-Bass.

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco, CA: Jossey-Bass.

Arnott, E., & Dust, M. (2012). Combating unintended consequences of in-class revision using study skills training. *Psychology Learning and Teaching, 11*, 99-104. Retrieved from http://dx.doi.org/10.2304/plat.2012.11.1.99

Askell-Williams, H., Lawson, M. J., & Skrzypiec G. (2012). Scaffolding cognitive and metacognitive strategy instruction in regular class lessons. *Instructional Science, 40*, 413–443. doi 10.1007/s11251-011-9182-5

Bain, K. (2004). *What the best college teachers do.* Cambridge, MA: Harvard University Press.

Beanland, V., Pammer, K. (2012). Minds on the blink: The relationship between inattentional blindness and attentional blink. *Attention, Perception, & Psychophysics, 74*, 322-330. doi 10.3758/s13414-011-0241-4

Berry, J. W., & Chew, S. L., (2008). Improving learning through interventions of student-generated questions and concept maps. *Teaching of Psychology, 35*, 305-312. doi: 10.1080/00986280802373841

Blackwell, L., Trzesniewski, K., & Dweck, C.S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*, 246-263. doi: 10.1111/j.1467-8624.2007.00995.x

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*, 61-100. doi: 10.2307/1167267

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school* (Expanded edition). Washington, DC: National Academy Press.

Bredemeier, K., & Simons, D. J. (2012). Working memory and *inattentional blindness*. *Psychonomic Bulletin & Review, 19*, 239-244. doi: 10.3758/s13423-011-0204-8

Chew S. L., Bartlett R., Dobbins, J., Hammer E. Y., Kite, M., Loop, T., McIntyre, J., Rose, K. A. (2009). Contextual approach to teaching: Bridging goals and outcomes. In D. Halpern (Ed.), *Undergraduate education in psychology: A blueprint for the future of the discipline* (pp. 95-112), Washington D.C.: APA.

Chew, S. L. (2005). Seldom in Doubt but Often Wrong: Addressing Tenacious Student Misconceptions. In D. S. Dunn & S. L. Chew (Eds.), *Best practices in teaching general psychology* (pp. 211-223). Mahwah, NJ: Erlbaum.

Chew, S. L. (2010). Improving classroom performance by challenging student misconceptions about learning, *APS Observer, 23,* 51-54. Retrieved from http://goo.gl/zEc2Pf

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24,* 205–249. doi 10.1007/s10648-011-9191-6

Clark, R., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: Evidence-based guidelines to manage cognitive load.* San Francisco, CA: Pfeiffer.

Cox R. D. (2011). *The college fear factor: How students and professors misunderstand one another*. Cambridge, MA: Harvard University Press.

Craik, F. I. M. (2002). Levels of processing: Past, present . . . and future? *Memory, 10*, 305–318. doi: 10.1080/09658210244000135

Dunlosky J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology, *Educational Researcher, 39*, 406–412. doi: 10.1177/1529100612453266

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5,* 69-106. doi: 10.1111/j.1529-1006.2004.00018.x

Dweck*,* Carol S. (2002). Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways). In J. Aronson (Ed.), *Improving academic achievement*: *Impact of psychological factors on education* (pp. 37-60). San Diego, CA, US: Academic Press.

Hammer E. Y., & Giordano, P. J. (2012). Active learning. In V. A. Benassi & W. Buskist (Eds.), *Effective college and university teaching: Strategies and tactics for the new professoriate* (pp. 99-106). Los Angeles, CA: Sage.

Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology, 82,* 472-481. doi: 10.1037/h0028372

Lilienfeld, S. O., Lynn, S. J., Ruscio, J., & Beyerstein, B. L. (2009). *50 Great Myths of Popular Psychology: Shattering Widespread Misconceptions about Human Behavior*. Hoboken, NJ: Wiley-Blackwell.

Mayer, R. E. (2009). *Multimedia learning (2nd ed.)*. New York: Cambridge University Press.

McDaniel M. A., & Wooldridge, C. (2012). The science of learning and its applications. In V. A. Benassi & W. Buskist (Eds.), Effec*tive College and University Teaching: Strategies and Tactics for the New Professoriate* (pp. 49-60). Los Angeles, CA: Sage.

McDaniel, M. A., Roediger, H. L., & McDermott, K. B., (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14*, 200-206. doi: 10.3758/BF03194052

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence, *Psychological Science in the Public Interest, 9*, 105-119.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16,385*-407.

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher, 39*, 406-412. doi: 10.3102/0013189X10374770

Sana, F., Weston, T., Cepeda, N. J. (2013). Laptop multitasking hinders classroom learning for both users and nearby peers. *Computers & Education 62,* 24–31. doi: 10.1016/j.compedu.2012.10.003

Sanbonmatsu D. M., Strayer D. L., Medeiros-Ward, N., & Watson, J. M. (2013). Who multi-tasks and why? Multi-tasking ability, perceived multi-tasking ability, impulsivity, and sensation seeking. *PLoS ONE, 8*, 1-8. doi:10.1371/journal.pone.0054402.

Schommer-Aikins, M., & Easter M. (2006). Ways of knowing and epistemological beliefs: Combined effect on academic performance, *Educational Psychology*, *26*, 411-423. doi: 10.1080/01443410500341304

Thomas, A. K., McDaniel, M. A. (2007). Metacomprehension for educationally relevant materials: Dramatic effects of encoding--retrieval interactions. *Psychonomic Bulletin & Review, 14*, 212-218. doi: 10.3758/BF03194054

van Gog, T., Paas, F., & Sweller J. (2010). Cognitive load theory: Advances in research on worked examples. *Educational Psychology Review, 22*, 375–378. doi 10.1007/s10648-010-9145-4

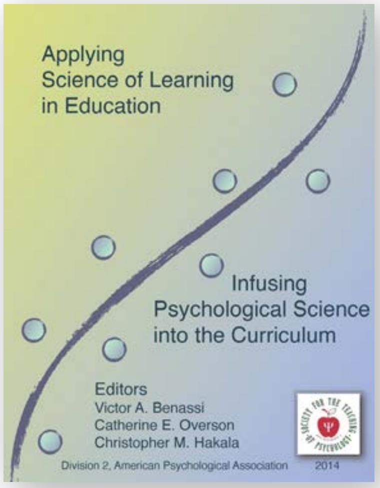
Part 3

Putting the Science of Learning Into Practice

Division 2,
American Psychological Association

Society for the Teaching of Psychology
APA Division 2

2014



The Effects of Memory Retrieval, Errors and Feedback on Learning

Nate Kornell
Williams College

Janet Metcalfe
Columbia University

Abstract

In recent years, a great deal of cognitive science research has focused on principles of learning that can be used to enhance education. And yet these principles have rarely been tested in real-world educational settings. For example, asking students to retrieve answers from memory enhances learning in laboratories—but what happens when a student fails to answer, or makes an error, or gets distracted? Can feedback compensate for such errors? We implemented a computer-based study program to help children learn, and in doing so assessed three cognitive principles: Does retrieving answers help? Do errors hurt, if they are corrected? And what is the effect of feedback? We found that retrieval helped, though less robustly than we expected; making errors, if they were corrected, caused no measurable harm; and feedback was unconditionally beneficial.

Introduction

As cognitive science has matured over the past 50 years, its potential contribution to educational practice has matured with it. But promising laboratory results do not always translate into successful interventions in real-world settings. We set out to examine three ways of enhancing learning that seemed, based on principles of cognitive science, potentially very powerful: Retrieving answers from memory (and actively participating in learning), avoiding errors, and giving feedback to correct errors when they were made.5

From a cognitive science perspective, the average student's self-regulated study habits are often flawed. When people study, they are subject to a number of biases and illusions that prevent them from studying optimally (Koriat, 1997). For example, students often believe they have learned something when in fact they have not (Bjork, 1999; Metcalfe 1998). They have faulty understandings of how memory works, and how to optimize it (Koriat, Bjork, Sheffer, & Bar, 2004; Kornell & Bjork, 2009). And study decisions made based on faulty memory monitoring are likely to be faulty themselves (Benjamin,

---

[5] Metcalfe and Kornell (2007) presented brief summaries of some of the experiments reported here.

Bjork, & Schwartz, 1998; Bjork, Dunlosky, & Kornell, 2013; Kornell & Bjork, 2007). Supervised learning has weaknesses as well. For example, in a foreign language textbook, presenting a word and its translation side by side makes self-testing difficult, and produces the sort of conditions that result in inaccurate metacognitive judgments (Nelson & Dunlosky, 1991) and overconfidence (Kelley & Jacoby, 1996). In the current experiments we investigated cognitive principles that we expected to improve learning.

The Retrieval Effect

One way to improve self-regulated study is to allow people to try to retrieve answers from memory. Retrieval increases the accuracy of metacognitive monitoring (Dunlosky, Hertzog, Kennedy, & Thiede, 2005; Dunlosky & Nelson, 1992), and students, at least at the college level, seem to want to test themselves when they study (Kornell & Bjork, 2007). More importantly for the current research, trying to retrieve information from memory directly enhances learning (Glover, 1989; Roediger & Karpicke, 2006b).

The benefits of attempting to retrieve information from memory are variously labeled as testing effects (Hogan & Kintsch, 1971; Roediger & Karpicke, 2006a), generation effects (Hirshman & Bjork, 1988; Jacoby 1978; Slamecka & Graf, 1978), and retrieval effects (Landauer & Bjork, 1978; Karpicke & Roediger, 2008). There are methodological and definitional differences between the three effects, but all three effects share an essential underlying characteristic: active, self-generated attempts to retrieve information from memory enhance memory (especially in the long term), as compared to passivity. We use the term retrieval (as opposed to generation or testing) to refer to the underlying processes that appear to underlie all three effects. The benefits of retrieval, of course, depend on other variables that we also examined in the current experiments, namely the effect of making errors, and feedback to correct those errors.

Errors and Feedback

The effects of retrieval are closely intertwined with the success of the retrieval attempt. What happens when one retrieves an incorrect answer, or no answer at all? Psychologists have long believed in the promise of errorless learning, which was championed by Skinner, among others (e.g., Guthrie, 1952; Skinner, 1958; Terrace, 1963). If one recalls an incorrect response, most memory models would predict that the incorrect response will be strengthened in memory, and as a result, it should be more likely to be recalled again on a later test (but see Mozer, Howe, & Pashler, 2004). Indeed, incorrect answers selected on multiple-choice tests do tend to persist (Roediger & Marsh, 2005; Marsh, Roediger, Bjork, & Bjork, 2007)—although the overall effect of testing, even in those situations, is positive. Recent research has shown that unsuccessful retrieval attempts, if they are followed by feedback, can be more effective than simply studying (Arnold & McDermott, 2012; Grimaldi & Karpicke, 2012; Hays, Kornell, & Bjork, 2013; Huelser & Metcalfe, 2012; Izawa, 1970; Knight, Ball, Brewer, DeWitt, & Marsh, 2012; Kornell, Hays, & Bjork, 2009; Richland, Kornell, & Kao, 2009; Vaughn & Rawson, 2012).

The effects of retrieval attempts and errors depend crucially on feedback. Most teachers would agree that if, during a study session, a student cannot think of the answer to a question, it would be wrong to simply leave the student wondering and move on. Instead, feedback should be given, in the form of more time to figure out the answer, a hint, the correct answer, or at the very least, telling the student whether they were right or wrong. (Different kinds of feedback undoubtedly have different outcomes; in the current experiments we defined feedback as a presentation of the correct answer.) The effects of feedback have been addressed in educational settings (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991;

Butler & Winne, 1995), but except for a few recent exceptions (see Metcalfe, Kornell, & Finn, 2009; Pashler, Cepeda, Wixted, & Rohrer, 2005; Pashler, Zarow, & Triplett, 2003), cognitive psychologists have paid little attention to the topic. Moreover, there is research in motor-skill learning that suggests that too much feedback can impair learning (Bjork, 1999; Schmidt & Bjork, 1992). Thus we investigated the effects of feedback.

The Previous Bronx Study

Our own previous research has shown that implementing cognitive principles in educational settings can dramatically improve student learning (Metcalfe, Kornell, & Son, 2007). We designed a computer program to help students study in an at-risk school in New York City's South Bronx. The program was successful, in large part, because it took advantage of a number of cognitive principles: retrieval, feedback, hints that allowed students to answer every question correctly, multi-modal processing, focused time allocation guided by ongoing evaluations of a student's progress, spacing, and rewards for effort. To maximize the benefits of the program, all of these principles were intentionally confounded, and implemented together. The program was compared to a control condition in which the students were given study tools such as flashcards, markers, paper, etc., and allowed to regulate their own study. The benefits of the program were large (as compared to self-guided study, computer-guided study resulted in an improvement on the order of 400-600% in scores on a delayed post-test), but a necessary drawback was that we could not isolate and measure the effect of any one principle by itself.

In the current experiments, we sought to systematically investigate three of the principles we had previously relied on: memory retrieval, feedback, and avoidance of errors. We conducted several experiments in an after-school program at the same poorly performing middle school in the South Bronx, in tandem with the sister experiments conducted under more controlled conditions using college undergraduates. It was our hope that by answering these questions, we would be able to revise the original study program to do an even better job of enhancing children's learning.

We conducted 7 experiments to address these issues. The first two experiments focus on the effects of retrieval, errors, and feedback. To foreshadow, we found a large benefit of feedback but no effect of making errors (which was replicated in Experiment 4) and an unexpectedly small benefit of retrieval. We investigated the retrieval effect further in Experiment 3, 5, 6, and 7. We explored four hypotheses concerning why the effect of retrieval was elusive: The first, the mixed list hypothesis, was that mixing items that had to be retrieved with items that did not have to be retrieved within the same list eliminated the benefits of retrieval. The second, the definition hypothesis, was that retrieval is not beneficial when definitions are used as stimuli. The third, the activity hypothesis, was that requiring participants to make decisions, in a non-retrieval condition, about their own self-paced study time, made them engage in the kind of active processing that mimics the benefits of an active retrieval condition. Finally the fourth, the retrieve-before-reading hypothesis, was that our procedure in the first three experiments—which included a brief pause, in the non-retrieval conditions, after the question was presented but before the answer was shown— turned the non-retrieval conditions into retrieval conditions and thus minimized the retrieval effect.

Experiment 1

In Experiment 1 we investigated the effects of memory retrieval, errors, and feedback using Grade 6 children from a poorly performing school in New York City's South Bronx as participants. The experiment involved a 3 x 2 factorial design. The first factor was retrieval condition: Free retrieval (i.e., answer only if confident), Forced retrieval (i.e., answer whether certain or not), and Read-only. This manipulation allowed us to address the effect of committing errors, by comparing the free retrieval conditions, which we expected to produce few errors, with the forced retrieval conditions, which we expected to produce many errors (see also Pashler, Rohrer, Cepeda, & Carpenter, 2007). Furthermore, comparing the retrieval conditions to the Read-only condition allowed us to measure the effect of retrieval. The second factor was feedback. On half of the trials, feedback was given after each response (i.e., the correct answer was presented for 2.5 s). On the other half of the trials, no feedback was given. (No feedback was given in the Read-only condition.) We predicted that retrieving information would benefit memory, that errors would have a detrimental effect on learning, and that feedback would have particularly important effects when the children were committing many errors.

Method

Participants. The participants were 16 children in 6th or 7th grade at a poorly performing middle school in New York City's South Bronx. They participated as part of an after-school program.

Setting. Each child participated one day per week (although they were sometimes absent and had to make up sessions). The children were volunteers who chose to participate after school. On average there were two experimenters and 4-6 students per session. The experiment took place in an otherwise unused classroom in the basement of the children's school. The experimenters, whom the children referred to as teachers, collected the children from their classrooms and brought them to the experimental room each day. The children were provided with a healthy snack and socialized for about 10 minutes before beginning the experiment. During the experiment, the children were seated at individual computers. The use of computers paralleled the increasingly common practice of using computers to aid student learning in schools. The students wore headphones and were discouraged from interacting with each other until the experiment was over, although inevitably there were times when they could not resist distracting each other. Thus the experiment was conducted in a classroom setting, but one in which the experimenters were able to maintain more control than would be possible in a normal classroom.

Design. The experiment was a 3 (Free-retrieval, Forced-retrieval, Read-only) x 2 (Feedback, No feedback) within-participants design. Feedback in the Read-only condition was effectively meaningless, because giving feedback would have meant presenting the answer immediately after presenting the answer. Therefore, we did not manipulate feedback in the Read-only condition, and there were five conditions: Read-only (presentation with no feedback), Free-retrieval/Feedback, Free-retrieval/No feedback, Forced-retrieval/Feedback, Forced-retrieval No feedback. Equal numbers of items were assigned to each condition.

Materials The materials were 120 definitions chosen to be consistent with the students' grade level, for example, "Not letting light through – Opaque." Participants studied, and were tested on, 30 definitions during each of the four sessions. The words were randomly assigned to session and condition separately for each participant.

Procedure. The experiment took place over the course of 4 sessions. The sessions occurred approximately once per week. They were essentially the same, except that different definitions were used in each session. Each session was split into two blocks of 15 definitions each. One block was presented under Free-retrieval conditions, and the other was presented under Forced-retrieval conditions. The blocks were used to avoid confusion about the free/forced instructions. After completing the first block, the participants were allowed to play the computer game Tetris for two minutes, as a break, and then the second block began. Thus each child participated in 8 blocks of study/test trials across four days, four blocks under Free-retrieval instructions, and four under Forced-retrieval instructions. Feedback, No feedback, and Read-only trials were mixed within each block.

The procedure during a given block consisted of three phases: (1) initial study, (2) manipulation, and (3) cued-recall test. During initial study all 15 definitions were presented one at a time. First the definition was presented alone for 1 second, after which the word appeared and a recording was played of the word being spoken aloud. The definition and word remained visible for 4 seconds, after which they disappeared and the next definition was presented.

During the manipulation phase, each item was tested and/or presented once. In the Read-only condition, the definition was presented alone for one second, and then the target word was presented until the participant pressed return, after which the pair remained visible for an additional 2.5 s and then the next trial began. In all other conditions, the definition was also presented alone for one second, but then an answer box was shown, and the participant was asked to type in the target word. In the Free-retrieval condition, participants were instructed to answer only if they were sure. In the Forced-retrieval condition, they were instructed to answer whether they knew the answer or not. After pressing return, in the Read-only and Feedback conditions the correct answer was presented for 2.5 seconds, while in the No feedback condition, the 2.5 seconds elapsed before the next item was presented, but the correct answer was not shown.

The third phase was the test phase (thus study and test took place during the same session). Each definition was presented, one at a time, and participants were asked to type in the target word and press return.

A fourth and fifth phase followed, during which the items were presented again (like in phase 2) and then tested again (like in phase 3). Phases four and five were included in the experiment so that participants would experience success and attain a sense of accomplishment. Moreover, they increased the participants' chances of remembering and using the vocabulary outside of the after-school program. We did not include the fourth and fifth phases in the data analysis, however, because the fourth phase served as a sort of delayed feedback for the third phase, even in the No feedback condition, and the test in the third phase served as a test even for the Read-only items.

Throughout the experiment, whenever a participant was prompted to enter a response, if they had not responded after 12 seconds a recorded voice said, "Hurry!" After another eight seconds, the voice said, "Please respond now!" Four seconds after that, the voice said, "Next question," and the computer automatically went on to the next question. The participants were told in advance to expect these warnings. As soon as the participant entered the first letter of their answer, the timing and voices stopped.

In the Forced-retrieval condition, participants had to give a response on every trial. If they left the answer blank, the computer prompted them to enter a response. To discourage nonsense answers, participants were asked to revise their answer if it did not contain a vowel, was shorter than three letters, or included a word from the definition. If any of these conditions was violated, participants were prompted to enter an English word that was not contained in the definition.

Results

To examine the effects of errors and feedback, we conducted a 2x2 ANOVA, excluding the Read-only condition. Test performance was scored (in all experiments) using a lenient scoring algorithm (created by Brady Butterfield) that counted correct but misspelled answers as correct.

Retrieval. To analyze the effect of retrieval, we compared the Read-only condition to the Free-retrieval with feedback condition. We chose the Free-retrieval condition as a comparison because it is most similar to previous experiments on retrieval effects (and because there was little difference in performance between the Free- and Forced-retrieval conditions). Although the Free-retrieval condition resulted in numerically better performance than the Read-only condition (Figure 1), as predicted by the retrieval effect, we were surprised to find that the retrieval effect was not significant, $t(15) = 1.59$, $p > .10$.

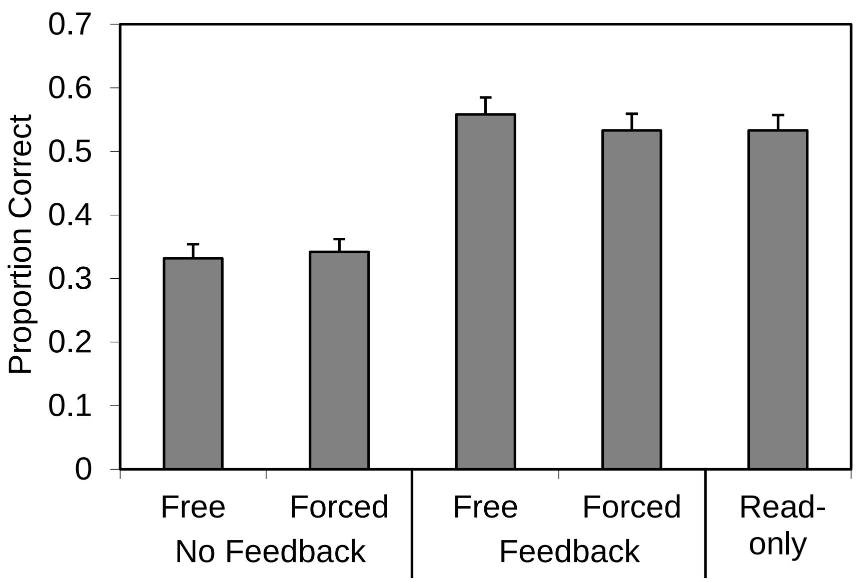


Figure 1. Proportion correct on the final test, as a function of study condition, in Experiment 1. The participants were middle school children. Error bars, in this and all other figures included herein, represent standard errors.

Errors. There was no significant effect of making errors. As Figure 1 shows, the Free- and Forced-retrieval conditions were not significantly different, $F(1, 15) = .07$. $p > .10$. The manipulation was effective, however, in making participants produce errors during the manipulation phase; commission errors accounted for significantly more trials in the Forced-retrieval condition (70% and 66% in the Feedback and No feedback conditions, respectively) than in the Free-retrieval condition (13% and 14%, respectively), $F(1, 15) = 172.01$, $p < .0001$, $\eta_p^2 = .92$.

Feedback. As Figure 1 shows, feedback had a large and significant effect on test performance, $F(1, 15) = 29.27$, $p < .0001$, $\eta_p^2 = .66$. Feedback did not interact with instruction condition (Forced-retrieval vs. Free-retrieval), $F(1, 15) = 1.05$, $p > 1.0$.

Discussion

The children learned more when they were given feedback than when they were not, as predicted, but only feedback had a significant effect. Making errors did not seem to affect performance at all.

Retrieving answers, although beneficial, did not have a significant effect. This finding was surprising, and (as described in the introduction) we tested a number of hypotheses about why it occurred in the experiments that follow.

Experiment 2

Experiment 2 was a replication of Experiment 1, using Columbia University undergraduates instead of middle school students as participants. The purpose of Experiment 2 was to verify the surprising findings from Experiment 1, and to determine whether the effects of retrieval, errors, and feedback remain constant or differ across age groups. In Experiment 2a, the procedure was essentially the same as in Experiment 1, except that the experiment took place in a single session; in Experiment 2b, the procedure was modified slightly to control for study time.

Method

The participants in Experiment 2a were 52 Columbia University undergraduates. The methodology of Experiment 2a was the same as the methodology of Experiment 1, with the following exceptions. Instead of taking place in four separate sessions, the experiment took place in a single session, which consisted of two blocks of 60 items each. The stimuli were age-appropriate definitions taken from GRE study guides (e.g., "Warm or glowing praise; eulogy; panegyric—Encomium").

The participants in Experiment 2b were 26 Columbia University undergraduates. The procedure differed from the procedure in Experiment 2a in one respect. In Experiment 1 and 2a, participants were shown feedback (in the feedback conditions) for a fixed 2.5 seconds, whereas in the Read-only condition, they were allowed to look at the answer for as long as they wanted. In Experiment 2b, participants were allowed to look at the answer for as long as they wanted in the Read-only condition and the feedback conditions. This was done to address a potential concern with Experiments 1 and 2a, namely that performance in the Read-only conditions might have been boosted artificially by the provision of participant controlled—and thus potentially unlimited—study time.

Results

The pattern of findings with college students was very similar to the findings with children from Experiment 1. There were no significant differences in test performance between Experiments 2a and 2b, so the results have been combined.

There was not a significant retrieval effect: Test performance in the Free-retrieval and Read-only conditions did not differ, $t(77) = 1.43$, $p > .10$ (Figure 2). Again, numerically, performance was better in the Free-retrieval condition than the Read-only condition.

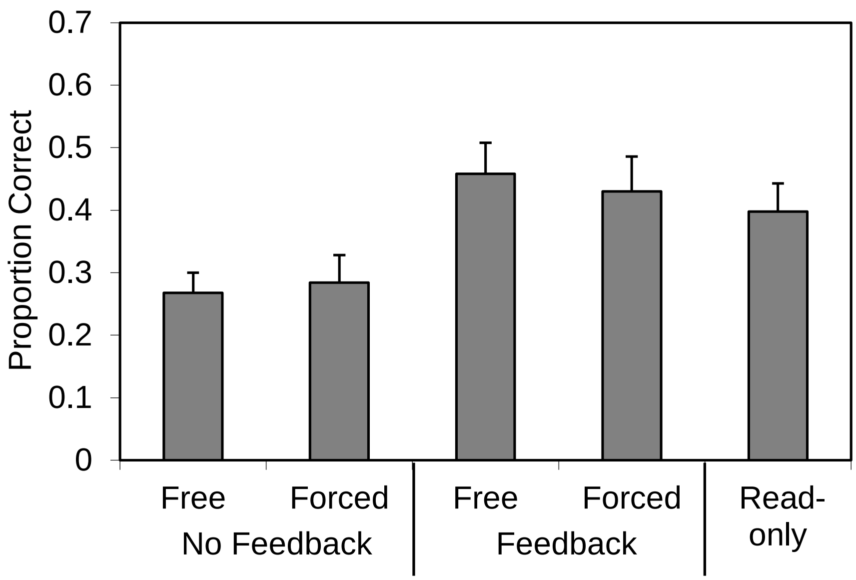


Figure 2. Proportion correct on the final test, as a function of study condition, in Experiment 2a and 2b (which were collapsed). The participants were university undergraduates.

There was a significant effect of feedback, $F(1, 76) = 278.05$, $p < .0001$, $\eta_p^2 = .79$ (Figure 2). There was no significant effect of making errors (i.e., Free- versus Forced-retrieval), $F(1, 76) = .38$, $p > .10$, and there was no interaction, $F(1, 76) = 1.01$, $p > .10$. Like Experiment 1, the instruction manipulation did result in more commission errors in the Forced-retrieval condition (62% and 63% in the Feedback and No feedback conditions, respectively) than the Free-retrieval condition (10% and 11%, respectively), $F(1, 76) = 413.94$, $p < .0001$, $\eta_p^2 = .84$.

Discussion

The results of Experiment 2 echoed those of Experiment 1. For both groups, feedback had a large positive effect. Making errors had no effect. Retrieving answers had a small positive effect, but the effect was not significant.

Combined analysis of Experiments 1 and 2 Because Experiments 1, 2a and 2b were methodologically similar, we analyzed combined data from all three experiments. There was a significant retrieval effect—the Free-retrieval condition resulted in more learning than the Read-only condition—although the effect was small, $F(1, 91) = 5.30$, $p < .05$, $\eta_p^2 = .05$. A 2x2x3 ANOVA (free vs. forced instructions x feedback x age group) showed that feedback had a significant effect, $F(1, 91) = 246.07$, $p < .0001$, $\eta_p^2 = .73$, but making errors did not, nor was there an interaction. Age group had no significant effects, which indicates that the three principles under investigation affected adults and middle-school students similarly.

In the remaining experiments we investigate the effects of retrieval and errors further. With respect to retrieval, as explained in the introduction, we examined four hypotheses: the mixed list hypothesis (i.e., that retrieval effects were masked by mixing retrieval and Read-only items in the same list); the definition hypothesis (i.e., that definitions are impervious to the benefits of retrieval); the activity hypothesis (i.e., that our Read-only conditions required study time decisions that created active, retrieval-like processing); and the retrieve-before-reading hypothesis (i.e., that participants took advantage—in the first two experiments as well as Experiment 3—of a short pause, in the Read-only condition, between the presentation of the question and answer to try to retrieve the answer, thereby turning the Read-only conditions into retrieval conditions). These questions, taken together, represent an examination of boundary conditions on the retrieval effect. To foreshadow, the data are most supportive of the retrieve-before-reading hypothesis.

Experiment 3a

The two most puzzling aspects of Experiments 1 and 2 were the fact that retrieving answers had small, if any, effects on memory, and that making errors had no apparent effects at all. In Experiment 3, we tested the hypothesis that the lack of retrieval effect in the first two experiments might have been due to the fact that retrieval and read-only items were mixed in the same list. Participants may have adopted a general tactic of testing themselves, and attempted to retrieve items even in the Read-only condition. Although this idea seemed plausible to us, it was not a foregone conclusion, insofar as Slamecka and Katsaiti (1987) have provided some evidence favoring generation effects even in mixed lists. Thus there were three retrieval conditions in Experiment 3: Forced-retrieval, Free-retrieval, and Read-only. Each of the three conditions was implemented in a different session. (Experiment 3 also tested the assumption that findings obtained in a within-participants design in Experiments 1 and 2 would also appear in a between-participants design Experiment 3).

We also tested a second hypothesis in Experiment 3, namely that the effect of retrieving answers might be larger when tested after a delay than when tested immediately, as previous results have shown (e.g., Roediger & Karpicke, 2006a). Thus half of the items were tested immediately, half were tested after a week delay. The participants were middle school children in Experiment 3a and college students in Experiment 3b.

Method

Participants. The participants were eight sixth and seventh grade students who participated as part of an after school program.

Materials. The materials were 90 of the definitions used in Experiment 1, for example, "A person who runs away from the law—Fugitive." Thirty definitions were introduced during each of the first three sessions, and items were assigned to session and condition randomly.

Design. The experiment was a 3 (Retrieval condition) x 2 (Delay) within-participants design. Retrieval condition had three levels: Free-retrieval, Forced-retrieval, and Read-only. To avoid confusion about the instructions, each retrieval condition was run in a different session. Delay had two levels: immediate and one week delay.

Procedure. The experiment took place over the course of four sessions. The first session consisted of two phases, study and test. Session two and three consisted of a delayed test on the previous week's words, followed by a study phase and a test phase on new words. Session four consisted of a delayed test on the words from session 3.

There was no initial learning phase in Experiment 3. The manipulation took place during the first phase. Each item was presented for study in one of three ways. In the Read-only condition, the definition was shown alone for .5 seconds, after which the answer appeared, and remained visible until the participant pressed a button to move on. In the Free-retrieval and Forced-retrieval conditions, the definition was shown alone for .5 seconds, and then an answer box appeared and participants were asked to type in the target word and press return. After they did so, the correct answer (i.e., feedback) was presented until the participant pressed a button to move on. The list of definitions was presented three times, each time in a different random order.

In the Free-retrieval condition, participants were instructed to answer only if they were sure. In the Forced-retrieval condition, participants were instructed to answer whether they knew the answer or not. If they gave an answer that did not contain a vowel, was shorter than 3 letters, or contained a word in the definition, they were told to revise their answer. Unlike the previous experiments, however, we decided not to force the participants to answer if they truly had no realistic guess; participants were allowed to leave the answer blank if they wanted to, although they were asked to confirm that they wanted to leave it blank each time they did so. The results confirm that despite being allowed to leave answers blank, participants still answered far more frequently in the Forced-retrieval condition than they did in the Free-retrieval condition.

The test phase followed the manipulation phase. Only the items in the immediate test condition were tested. Each definition was shown individually for .5 seconds, and then the answer box appeared and the participant was asked to type in the target word and press return. The delayed-test items were tested at the start of the next session, which occurred after a week delay. The delayed test was the same as the immediate test, with one exception (see next paragraph).

Like Experiments 1 and 2, during the manipulation and immediate test (but not during the delayed test), a recorded voice said “Hurry!” after 12 seconds, “Please respond now!” after 20 seconds, and “Next question” after 24 seconds, at which point the computer moved on to the next question.

Results

There was not a significant effect of retrieval condition, $F(2, 14) = 2.32$, $p > .10$, although numerically, performance was better in the Free- and Forced-retrieval conditions than it was in the Read-only condition. As expected, immediate test performance was better than week-delayed test performance, $F(1, 7) = 12.86$, $p < .01$, $\eta_p^2 = .65$ (Figure 3). There was no retrieval condition x delay interaction, $F(2, 14) = .35$, $p > .10$.

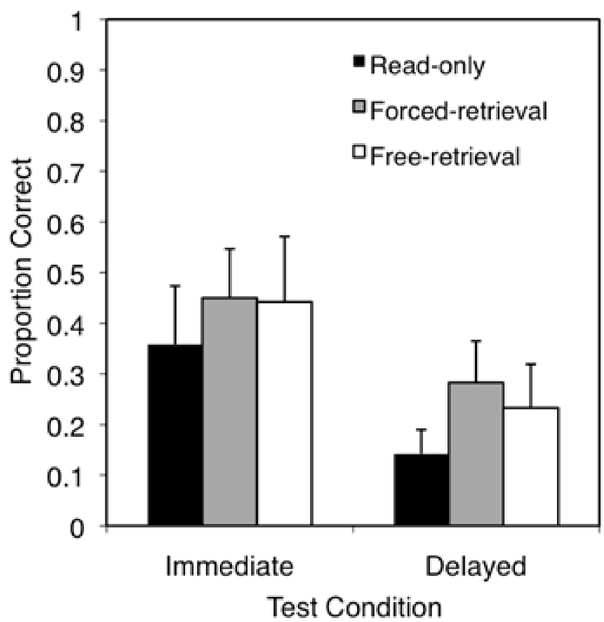


Figure 3. Proportion correct as a function of study condition and test delay in Experiment 3a. The participants were middle school children.

Making errors appears to have had virtually no effect on memory. The forced retrieval instructions did increase the rate of commission errors during study, which, averaged across the three sessions, was significantly greater in the Forced-retrieval condition (31% and 33% in the Immediate and Delayed conditions, respectively) than it was in the Free-retrieval condition (7% and 12% in the Immediate and Delayed conditions, respectively), $F(1, 7) = 29.07$, $p < .001$, $\eta_p^2 = .81$. Oddly, there was a significant effect of delay: There were more commission errors on items assigned to the Delayed than Immediate condition, $F(1, 7) = 7.43$, $p < .05$, $\eta_p^2 = .50$. However, given that test delay was a random variable that was not manipulated until after the commission errors had been made, this effect appears to have occurred by chance.

Discussion

Experiment 3a did not support the hypothesis that the lack of retrieval effect in Experiments 1 and 2 resulted from the use of a mixed-list design. Consistent with Experiments 1 and 2, making errors had no effect. There was little apparent benefit of retrieving answers from memory, although, as described above, it is possible that participants were attempting to retrieve in the Read-only condition, turning it into a retrieval-like condition.

Experiment 3b

Experiment 3b was similar to Experiment 3a, except that the participants were college students and the experiment took place during a single session, not over the course of four sessions.

Method

The participants in Experiment 3b were 12 Columbia University undergraduates. The materials were 90 of the 120 definitions from Experiment 2.

The experimental procedure was the same as the procedure in Experiment 3a with one exception. The one exception was that all four “sessions” took place within a single hour (as opposed to over the course of four separate days). There was a delay period of five minutes between sessions, during which participants were allowed to read funny and engaging newspaper columns by Dave Barry. Thus the delayed test took place after 5 minutes, not one week. Like Experiment 3a there were three retrieval conditions: Free-retrieval, Forced-retrieval, and Read-only.

Results

Neither retrieval nor errors had a significant effect in Experiment 3b: Comparing the Free-retrieval, Forced-retrieval, and Read-only conditions, there was no significant effect of retrieval condition, $F(2, 22) = 1.66$, $p > .10$ (Figure 4). Test performance was significantly worse after a five minute delay than on the immediate test, $F(1, 11) = 11.55$, $p < .01$, $\eta_p^2 = .51$, but there was no interaction between retrieval condition and delay, $F(2, 22) = .77$, $p > .1$. The manipulation worked, however, in the sense that there were more commission errors in the Forced-retrieval condition (43% and 50% in the Immediate and Delayed conditions, respectively) than in the Free-retrieval condition (17% and 19% in the Immediate and Delayed conditions, respectively), $F(1, 11) = 13.79$, $p < .01$, $\eta_p^2 = .56$. Like Experiment 3a, there were more commission errors on items assigned to the Delayed than Immediate condition, $F(1, 11) = 6.01$, $p < .05$, $\eta_p^2 = .36$, although paradoxically, test delay was not manipulated until after the commission errors had been made.

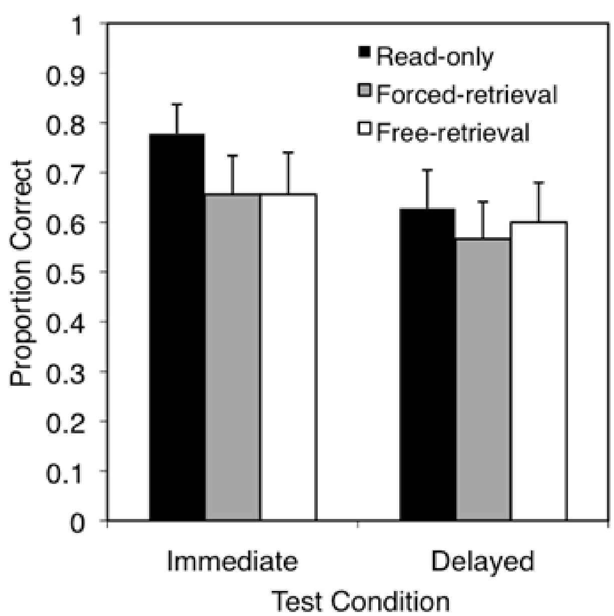


Figure 4. Proportion correct as a function of study condition and test delay in Experiment 3b. The participants were university undergraduates.

Discussion

The results of Experiments 3a and 3b were similar. Retrieval had little effect, and making errors appeared to have no effect whatsoever. As expected, recall performance was better on an immediate test than a delayed test. We had hypothesized that perhaps, in Experiments 1 and 2, mixing retrieved and Read-only items within a list had negated the retrieval effect. That hypothesis was not supported by the lack of retrieval effect in Experiment 3, in which the items were not mixed. There was no consistent retrieval effect, although again it is possible that our procedure effectively made the Read-only condition into a retrieval condition (a hypotheses that we explore further below).

Experiment 4

In the fourth experiment, we attempted one last time to uncover any possible effect of making errors on learning. College students participated in a one-session experiment. On some trials they were encouraged to answer every question, and on others they were asked to answer only if they were sure. We did not manipulate retrieval; because we wanted to maximize the chance of finding an effect of errors, participants retrieved information on every trial.

Method

Participants. The participants were 25 Columbia University students who participated for course credit or pay.

Materials. The materials were 213 general information questions taken from Nelson and Narens (1980), for example "What is the capital of Jamaica?—Kingston". The questions were assigned to subjects and conditions randomly.

Procedure. The experiment consisted of two phases, study and test. The study phase lasted 25 minutes, and thus the exact number of questions a participant was asked varied. The test phase lasted until participants had answered all of the questions that had been asked during the study phase, which took usually 20-25 minutes.

During the study phase, there were two conditions. In the Forced-retrieval condition, the screen was red, and participants were told to answer every question. They were told that if they did not know the answer, they should guess, and if that if they could not guess, they should enter a related answer. In the Free-retrieval condition, participants were told to answer only if they were sure. In both conditions, the correct answer was presented after the participant responded. Trials came in blocks of 6, so participants did 6 forced trials, then 6 free trials, and so on, which helped avoid confusion about the instructions.

During the test phase, all of the questions asked during the first phase were asked again, and participants were simply told to do their best.

Results

There was not a significant difference in final test accuracy between the Forced (M = .60, SD = .12) and Free (M = .59, SD = .13) conditions, $t(24) = .28$, $p > .10$. The manipulation was successful in making participants generate more commission errors during study in the Forced (M = .72, SD = .08) than Free (M = .12, SD = .13) condition, however, $t(24) = 20.12$, $p < .0001$.

Another way of looking at errors is to compare unforced errors of commission versus unforced errors of omission. To do so, we analyzed final test accuracy in the Free condition. Participants were more accurate on the final test on items on which they initially made commission errors (*M* = .56, *SD* = .28) versus omission errors (*M* = .45, *SD* = .16), $t(24) = 2.17$, $p < .05$. Based on this analysis, making errors actually had a positive effect on memory. However, this analysis is contaminated by item selection effects—for example, participants might be more likely to answer, and therefore to make commission errors, when they are familiar with a question's general topic area (e.g., American literature) than when they are not, and that familiarity might make learning the actual correct answer easier—which is precisely the problem we avoided by manipulating errors. Thus this analysis is more useful in illustrating the problem of item selection effects than it is in providing information about the effect of errors.

Discussion

The results of Experiment 4 are easily summarized: Making errors had no discernable effect on learning. This finding is consistent with the results of Experiments 1, 2 and 3.

Surveying the results of the first four experiments, some questions have clear answers: Feedback had universally, and powerfully, positive effects. Making errors, as long as they were followed by corrective feedback, made no difference. The nature of the retrieval effect remained a question, however, which we turned to in the last three experiments. We attempted to address two questions: Is there a retrieval effect in educationally realistic settings? And why was there no retrieval effect in the first three experiments?

Experiment 5

One possible explanation of why we did not obtain a retrieval effect in the first three experiments was the nature of the Read-only condition. In the Read-only conditions of the first three experiments, participants could control how much time they spent studying each item, which made them engage in a process of deciding how much time they should spend studying; that is, participants in the Read-only conditions were obliged to make study-time allocation decisions, presumably guided by metacognitive monitoring of how well they knew the answers to the questions. Engaging in metacognitive monitoring may have encouraged participants to become actively involved in their learning, and to process the to-be-learned materials effectively, even in the Read-only conditions. If so, it is possible that the control condition in the first three experiments bore some similarity to active retrieval conditions in previously published experiments. To test the hypothesis that active participation in the Read-only conditions negated the retrieval effect in Experiments 1-3, we compared three study conditions in Experiment 5: an active presentation condition, in which participants controlled how long they spent studying each item; a passive presentation condition, in which each item was presented at a fixed rate and participants observed it passively; and a retrieval condition in which participants actively attempted to retrieve information from memory.

Method

Middle school children studied synonyms in three study conditions—active presentation, passive presentation, and retrieval—and then took two tests, one immediately after the session ended, and one after a week delay.

Participants. The participants were 21 middle school children from the same school as the previous experiments.

Materials. Like the previous experiments, the materials were English vocabulary words, but in the current experiment they were synonyms (e.g., agree—concur), instead of definitions paired with words. There were 90 pairs, 30 of which were assigned randomly to each of the three study conditions (active presentation, passive presentation, and retrieval) on a participant-by-participant basis.

Procedure. The children participated in two sessions that took place one week apart. The first session consisted of three phases: initial learning, manipulation, and test. The second sessions consisted of a delayed retest on all of the items (all of which had also been tested during session 1).

Session 1 was broken into three separate blocks. The three blocks were the same except for the stimuli: Each contained 30 words and three phases—initial study, manipulation, and test—and participants completed an entire block before beginning the next one. The blocks were used to make the task of recalling the words easier, based on the concern that trying to learn 90 synonyms all at once would be too daunting. All three conditions (active presentation, passive presentation, and retrieval) were mixed within each block.

The first phase in a given block was initial study. The 30 synonym pairs were presented, individually, for 3 seconds each.

The second phase was the manipulation. The manipulation phase was split into three sets of 10 pairs—one for each of the three study conditions. There were instructions at the beginning of each set of 10 pairs explaining what the participant should do. In the retrieval condition, on each trial a cue was shown and the participant was asked to type in its synonym and press return. When they pressed return, the correct answer appeared and remained visible until they pressed return again to move on. In the active presentation condition, the cue and target were shown together, and remained visible until the participant pressed return to move on. In the passive presentation condition, the cue and target were shown for three seconds, and then the computer automatically moved on (with no input from the participant).

At the end of the second phase, as a distractor task, participants were shown a joke (e.g., "One day a sloth was out for a walk when he was mugged by four snails. A police officer asked him 'Can you describe the snails?' 'Well, not really,' replied the sloth, 'I mean, it all happened so fast!'"). A recording of the joke was also played over headphones.

During the test phase, each cue was shown individually, and participants were asked to type in its synonym and press return. Like the previous experiments, during the first session (but not during the week-delayed test) a recorded voice said "Hurry!" after 12 seconds, "Please respond now!" after 20 seconds, and "Next question" after 24 seconds, at which point the computer moved on to the next question.

A delayed test took place one week after the first session. The test was the same as the test phase during the first session, except that the prompts to hurry, etc., were removed. All 90 pairs were tested. All of the pairs had already been tested during the first session.

Results

On the immediate test, there was a significant effect of study condition, $F(2, 40) = 7.28$, $p < .01$, $\eta_p^2 = .27$ (Figure 5). Post-hoc Tukey tests showed that test performance was significantly better in the retrieval condition than either of the presentation conditions, which did not differ significantly. The benefit of retrieval occurred despite the fact that during the manipulation phase, retrieval attempts rarely met with success; participants answered correctly on just 11% of the trials during the manipulation phase. There was a floor effect on the week-delayed test—accuracy was below 3% in all three study conditions.

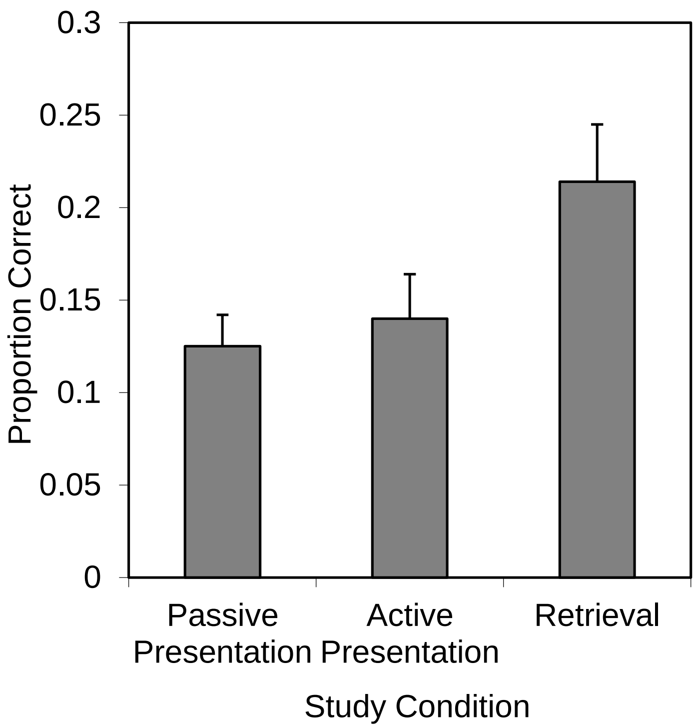


Figure 5. Proportion correct as a function of learning condition on the immediate test in Experiment 5. The participants were middle school children.

Discussion

For the first time in this series of experiments, we found robust evidence of a retrieval effect: Retrieving answers enhanced learning significantly more than either type of Read-only condition. The hypothesis that an active, self-paced Read-only condition negated the retrieval effect in the first three experiments was not supported. Test performance was not significantly better in the active than passive presentation condition. There was a significant retrieval effect, however.

The retrieval effect was elusive in Experiments 1-3. One explanation for that finding is to argue that retrieval attempts are primarily beneficial when they are successful, and that retrieval rates were too low to be beneficial during study in Experiments 1-3. The results of Experiment 5 argue against this explanation. In Experiment 1 and 2, adults and children answered correctly on .34 and .28 of the learning trials in the retrieval conditions; in Experiment 3, the proportions were .51 and .36. If retrieval is primarily effective when an item is actually retrieved successfully, then a minority of items were actually benefiting from being retrieved in the first three experiments. There is some evidence, however, that a retrieval attempt can be effective even if one does not answer successfully, as long as the attempt is followed by feedback (e.g., Izawa, 1970; Kornell, et al., 2009; Richland et al., 2009). Moreover, in Experiment 5, participants retrieved correct answers during study on only .11 of trials—a lower proportion than Experiments 1-3—and yet a retrieval effect occurred. Thus the moderate retrieval effects in Experiments 1-3 (which were significant only in the combined analysis of Experiments 1 and 2) does not appear to be attributable to low rates of successful retrieval during study.

Why was there a robust retrieval effect in Experiment 5, but not in the first three experiments? The most obvious difference was in the materials: Experiments 1-3 involved definitions paired with words; Experiment 5 involved two-word synonym pairs. Carroll and Nelson (1993) claim that general information questions are a boundary condition on the retrieval effect. deWinstanley (1995) has provided convincing evidence otherwise, however, by obtaining retrieval effects using general information questions, using Carroll and Nelson's methodology. The crucial factor was that deWinstanley did not ask participants to answer every question at the outset of the experiment, which may have effectively turned the read-only condition into a retrieval condition in Carroll and Nelson's study. Nevertheless, if general information questions produce relatively small retrieval effects, consistent with Carroll and Nelson's findings, the same might apply to definitions. With both types of stimuli, people often know the answer before they begin an experiment, even if they have forgotten it temporarily.

More importantly, it takes a relatively long time to read a general information question or definition, as compared to a single word. It might take long enough to allow people to try to retrieve the answer while reading the question, even in a Read-only condition. Doing so would turn a Read-only condition into a retrieval condition. An aspect of the procedure used in the first three experiments—but not in Experiment 5 (in Experiment 4 retrieval was not manipulated)—may have compounded this problem: The definition was presented alone, even in the Read-only condition, before the answer was presented (for 1 second in Experiments 1 and 2; for .5 seconds in Experiment 3). Seeing the answer appear alone may have further encouraged participants to try to recall the answer while reading the question. Consistent with the hypothesis that participants sometimes turn read-only conditions into retrieval conditions, deWinstanley and Bjork (2004) found that when participants experienced both generation and presentation conditions on a first list, the generation effect disappeared on a second list, presumably because the participants began generating answers spontaneously even in the read-only condition. Our participants may have similarly decided to retrieve in the read-only condition. In summary, we hypothesized that, in the first three experiments, the use of definitions as stimuli, and the presentation of the questions alone before the answer appeared in the Read-only condition, may have prompted participants to retrieve during the Read-only conditions, and thereby diminished the retrieval effect.

Experiment 6

In Experiment 6 we investigated the retrieval effect using simple laboratory materials, with college students, in a paradigm that did not require participants to make an overt response at all during the study phase (see Carpenter, Pashler, & Vul, 2006; Carrier & Pashler, 1992). Instead, participants were simply shown the question before the answer in the retrieval condition, or the question and answer together in the Read-only condition. We hypothesized that simply presenting the question before the answer—as we did in the Read-only conditions Experiment 1, 2, and 3—would be enough to create a retrieval effect.

Method

College students studied word pairs in one of four conditions: Read-only, in which the cue and target were shown together for 6 seconds; Retrieval-only, in which the cue was presented alone for 6 seconds;

Retrieve + Read, in which the cue was shown alone for 3 seconds, and then the cue and target were shown together for the next 3 seconds; and a Not-studied control condition.

Participants. The participants were 24 Columbia University students who participated for course credit or pay.

Materials. The materials were 6 letter words that were randomly paired together (e.g., Palace—Cattle).

Procedure. The experiment took place during a single session, which was split into 3 blocks of 24 pairs each. There were three phases during each block: initial study, manipulation, and test. The first two and last two pairs in each block served as buffers, and were not analyzed. The same pairs served as buffers during each of the three phases.

During the initial study phase, the pairs were presented individually for three seconds each. The entire list of pairs was presented twice, each time in random order (the order of the buffer items was not randomized). At the end of the initial study phase, participants were asked to count backwards by threes for 30 seconds.

During the manipulation phase, there were four conditions. In the Not-studied condition, the items were not presented or tested; in the Retrieval-only condition, the cue word was presented alone for 6 seconds; in the Read-only condition, the cue and target were presented together for 6 seconds; and in the Retrieve + Read condition, the cue was presented alone for 3 seconds, after which the cue and target were presented together for the next 3 seconds. All of the items were presented twice (except the Not-studied items), and were mixed in random order during study. At the end of the manipulation phase, participants were asked to count backwards by threes for 30 seconds.

During the test phase, each cue was presented in random order, and participants were asked to type in the target and press return.

Results

Retrieval enhanced learning in Experiment 6. There was a significant main effect of condition in which an item was presented, $F(3, 69) = 102.27$, $p < .0001$, $\eta_p^2 = .82$. As Figure 6 shows, performance was better when participants were tested and then shown the answer, in the Retrieve + Read condition, than in the Read-only condition. A planned comparison showed this difference to be significant, $t(23) = 3.93$, $p < .001$. Retrieval was also beneficial in the absence of feedback: The Retrieval-only condition resulted in more learning than the Not-studied condition, $t(23) = 4.22$, $p < .001$.

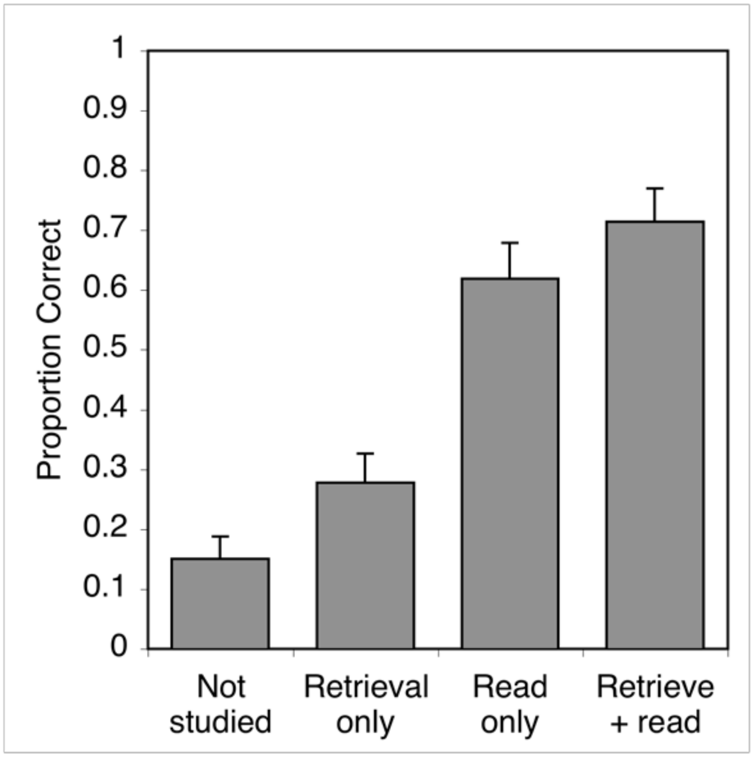


Figure 6. Proportion correct as a function of learning condition in Experiment 6. The participants were university undergraduates.

Discussion

In Experiment 6, a powerful retrieval effect was produced simply by presenting a question for 3 seconds before presenting the answer—even when the total time allowed for study in the two conditions was the same (for similar findings, see Carpenter et al, 2006; Carrier & Pashler, 1992). This finding suggests that in Experiments 1-3, by presenting the question before the answer in the Read-only condition, we may have effectively turned the Read-only condition into a retrieval condition, and thus diminished the retrieval effect (i.e., the difference between the Read-only and retrieval conditions). Instead of suggesting a weak retrieval effect, such a finding might be interpreted as just the opposite (as detailed in the general discussion): perhaps it is a testament to the power of retrieval that even a very brief opportunity to retrieve brought performance in the Read-only conditions up to the level of the retrieval conditions.

It is worth noting that each question was presented alone for 3 seconds in the current experiment, whereas it was presented alone for only 1 second in Experiments 1 and 2, and .5 seconds in Experiment 3. Even so, not showing the answer immediately may have encouraged participants to try to retrieve it. Indeed, the simple fact that the earlier experiments used definitions as cues may have allowed participants to test themselves while they were still in the midst of reading the question itself.

Experiment 7a

The purpose of Experiment 7 was to further test the robustness of the retrieval effect using realistic materials. The participants were middle school students in Experiment 7a, and college students in

Experiment 7b. Experiment 7a used synonyms, whereas Experiment 7b used definitions, allowing a test of the hypothesis that using definitions as stimuli—which we did in Experiments 1-3—can, by itself, eliminate the retrieval effect.

Method

Middle school students studied vocabulary synonyms in two conditions, Read-only and Retrieval. They were tested after a short delay, and the experiment took place within a single session.

Participants. The participants were eight middle school students from the same school as the previous experiments.

Materials. The stimuli were 120 synonyms (e.g., agree-concur). Many of the synonyms came from the set of 90 used in Experiment 5. Only 60 of the synonyms were used for any given participant. The assignment of stimuli to participants and conditions was random.

Procedure. The entire experiment took place in a single session. The session was split into four blocks. Each block was completed before the next block began. Two blocks were assigned to the Retrieval condition, and the other two were assigned to the Read-only condition. The assignment of blocks to conditions was random. Each block involved 15 items and three phases: initial study, manipulation, and test.

During the initial study phase, each of the 15 synonym pairs was presented individually for 3 seconds.

At the beginning of the manipulation phase, participants were reminded of the instructions for the block (i.e., either to retrieve the answers or to study the pairs). The list of pairs was studied 3 times, each time in a different random order. During a Read-only block, a cue and target were shown together for three seconds on every trial. During a retrieval block, a cue was shown on every trial, and the participant was asked to type in its synonym and press return. After the participant did so, the correct answer was presented until the participant pressed return to go on to the next item.

At the end of the manipulation phase, participants were shown a joke, which was also played over the headphones. Then the test began.

During the test phase, each cue was shown individually and, the participants were asked to type in the synonym and press return. At the end of the test phase, the next block began, or the experiment ended. During the manipulation and test phases, participants were given the same warnings to "Hurry!" etc., as in Experiments 1, 2, 3 and 5.

Results

Test performance was significantly better in the retrieval condition (M = .43, SD = .14) than the presentation condition (M = .28, SD = .16), $t(7) = 2.60$, $p < .05$. Five of the eight participants were run on a second session, which was the same as the first session but used different stimuli. The effect was also significant when the second session was included in the analysis, $t(7) = 2.75$, $p < .05$.

Experiment 7b

Experiment 7b was a replication of Experiment 7a using college students instead of middle school students as participants. The other major change was that instead of using synonyms as stimuli, which had produced a retrieval effect in Experiments 5 and 7a, we used the definitions from Experiments 2a, 2b, and 3b, which had not produced a retrieval effect in those experiments.

Method

The participants were 21 Columbia University students. The experiment was essentially the same as Experiment 7a, with the following exceptions. The stimuli were the definitions used with college students in Experiments 2a, 2b, and 3b. There were 25 definitions per block, instead of 15, for a total of 100 items. The items were presented for 2 seconds each instead of 3 during the initial presentation phase. Finally, the prompts to "Hurry!" etc., were not used.

Results

Test performance was significantly better in the Retrieval condition (M = .76, SD = .22) than the Read-only condition (M = .69, SD = .25), $t(20) = 2.16$, $p < .05$. The effect was not large, but given that it was obtained with definitions as the to-be-learned materials, it indicates that the materials alone cannot explain the lack of retrieval effect in Experiments 1, 2, and 3.

Discussion

Taken together, the results of Experiments 7a and 7b replicated the retrieval effect found in Experiment 5. Moreover, the results of Experiment 7b suggest that definitions are not a boundary condition on the retrieval effect, although the possibility remains that the retrieval matters less with definitions than it does with word pairs.

At the outset of Experiment 5, we had two questions: First, can a retrieval effect be obtained with educationally realistic materials? The answer appears to be yes, given that Experiments 5 and 7 both showed significant retrieval effects. Second, why did the first three experiments not produce retrieval effects? Experiment 5 indicates that it was not because the self-paced nature of the Read-only conditions in the first three experiments induced active participation on the participants' part. Experiment 7b indicates that it was not because definitions are a boundary condition on the retrieval effect (Carroll & Nelson, 1993). One possibility that still remains is that the lack of retrieval effect, in the first three experiments, occurred because the question was presented by itself before the answer appeared in the Read-only condition—which was enough to create a retrieval effect in Experiment 6—and perhaps participants took that short interval as a chance to try to retrieve the answer, even in the Read-only conditions.

General Discussion

In the current experiments, we examined the effects of three experiences that students encounter frequently while studying: retrieving answers from memory, making errors, and receiving feedback. Hypotheses derived mainly from laboratory experiments were tested in a real educational setting, with both middle school students and college students. Feedback had unambiguously large positive effects. Avoiding errors had virtually no measurable effects whatsoever. Retrieving answers, which previous studies have shown to be an effective way of learning, had surprisingly mixed results.

The Retrieval Effect

The expected benefit of retrieval did not materialize in all of the experiments: Retrieval had small effects Experiments 1 and 2, no significant effects in Experiments 3, and significant effects in Experiments 5, 6, and 7. From one perspective, the effect of retrieving answers was less robust than we expected based on previous research. From another perspective, however, retrieval might be interpreted as more effective than we expected. We have tentatively hypothesized that the retrieval effect was elusive in the first three experiments because participants retrieved even in the Read-only conditions (because of a brief—and unintentional—opportunity to do so). If so, then perhaps the Read-only conditions rose to the level of the Retrieval conditions, as opposed to the Retrieval condition lowering to the level of the Read-only conditions—and retrieval should be seen not as a weak effect that does not apply in real life, but instead quite the opposite; it is a strong effect that can occur with minimal provocation.

The methodological "flaw" in Experiments 1-3 was actually advantageous, because it suggests an intriguing possibility: Perhaps presenting a question without its answer for just one second, or even half a second, was enough to encourage students to attempt to retrieve answers from memory in the Read-only conditions. That would be consistent with Experiment 6 and prior research (Carrier & Pashler, 1992), although why intervals as short as a second or less were sufficient to create a retrieval effect is puzzling. One possible explanation is that using multi-word definitions, combined with the delay before the answer was presented, encouraged participants to try to retrieve while they were reading the definition. Moreover, students are often eager to test themselves while they study (Kornell & Bjork, 2007; Kornell & Son, 2009), and the delayed onset of the answer may have provided a window of opportunity for such testing.

Regardless of whether the retrieval effect is viewed as fragile or robust, it is clear that the effect did not translate directly from laboratory research to the somewhat more applied experiments presented here. In one sense, that is exactly what one should expect: Translating a phenomenon that is almost universally successful in the laboratory into a real world setting does not always result in the expected outcome. For educators, one way of seeing the results is that students should be asked to generate answers when they are learning; for example, teachers should pose questions to their classes, even if they do so briefly and without requiring the students to make a correct response, or any overt response at all.

Taken together, the current experiments establish an important set of boundary conditions on the retrieval effect. They include the following: that retrieval effects can be obtained even when levels of successful retrieval during study are low (on the order of 10% correct in Experiment 5); that retrieval can be effective whether or not study trials are self-paced; that retrieval effects can be obtained when the cues are single words or multi-word definitions; that overt responses are not necessary to create retrieval effects; and that a brief interval during which a question is presented without its answer may be enough, under some conditions, to produce generative processing and thereby enhance learning.

Attempting to recall information from memory can have three basic outcomes: successful retrieval of the correct answer; retrieval of an incorrect answer that one believes to be correct (i.e., a commission error); and a failure to retrieve an answer that one believes to be correct (i.e., an omission error). There is ample evidence for the benefits of successful retrieval (e.g., Allen, Mahler, & Estes, 1969; Bjork, 1988; Landauer & Bjork, 1978; Tulving, 1967), as well as theories to explain why successful retrieval is beneficial (such as the idea that retrieving answers for oneself results in a deeper encoding of the answer; see Bjork, 1975; Carpenter & DeLosh, 2006). When retrieval is not successful, the current results suggest that, apart from item differences (which make commissions appear more correctable than omissions), omissions and commissions are equally correctable (i.e., correct answers are equally learnable). It is when errors are made that role of feedback becomes important—feedback is of little benefit if one already knows the correct answer (Pashler et al., 2005).

The current results are consistent with a view of retrieval as an error correction process. Retrieval was, apparently, effective even when it was unsuccessful: In Experiment 5, 90% of participants' responses during learning were errors, and yet there was a retrieval effect. (More speculatively, in Experiments 1-3, the Read-only condition appears to have acted like a retrieval condition even though retrieval attempts were likely brief and presumably often short-circuited before they could succeed.) These findings are consistent with evidence that retrieval attempts potentiate encoding that happens during feedback, so that even unsuccessful retrieval attempts can enhance learning (e.g., Izawa, 1970; Kornell et al., 2009; Richland et al., 2009). Attempting to answer a question naturally brings to mind information that is related to a question's answer. Although some theories would suggest that thinking of an incorrect answer might have negative effects on future learning, another view is that bringing errors to mind is a very effective way of identifying and correcting those errors so that they do not occur in the future (c.f., Mozer et al., 2004; Pashler et al., 2003).

Errors and Feedback

Because the retrieval effect was elusive, a large amount of effort was expended investigating it. But that focus should not overshadow the fact that by far the most robust and powerful finding was that feedback greatly enhanced learning. There may be circumstances, such as in motor learning, in which too much feedback can be counterproductive. In verbal learning, however, if a student cannot answer a question, removing feedback altogether appears to be catastrophic. Different types of feedback can be differentially effective—providing hints that would have allowed participants to discover the answer might have been more effective than providing the answer outright, and simply indicating whether participants' responses were correct or incorrect would presumably be less effective than providing the answer—but in the current experiments, anything short of guiding participants to the correct answer essentially guaranteed that the answer would not be learned.

The effect of making errors was remarkably consistent: There was no significant effect of making errors in any of our experiments—indeed, errors made virtually no numerical difference at all. One might argue that some of the errors made in the forced retrieval conditions were not "real," in the sense that perhaps the participants did not really believe their erroneous answers. However, simply making a response can have powerful effects on memory. People who are asked to simply say (or silently mouth) the answer to a set of questions remember those answers better than a group asked to think of the answer without making an overt response (Hourihan & MacLeod, 2008). Thus the fact that participants were asked to produce answers in the forced condition is, by itself, enough to make those errors both real and memorable, and yet the errors had little or no cost when followed by feedback. Moreover, as Butterfield and Metcalfe (2001, 2006) have shown, errors that people firmly believe in (i.e., high confidence errors) are not the most dangerous kind—quite the contrary, high confidence errors, if corrected, are the most likely to be answered correctly on a delayed test. Low confidence errors, such as the errors that the Forced response condition probably produced, may be just as harmful (or harmless) as high confidence errors.

The fear that making errors might hamper learning appears to be unfounded. Educators would do well to introduce difficulty in study situations; the current evidence suggests that increased error rates will not hinder learning (see also Pashler et al., 2003). Moreover, the retrieval effect is one of a number of educationally relevant psychological effects (another is the spacing effect; see Dempster, 1988) that can be characterized as "desirable difficulties," that is, situations in which increasing difficulty during study enhances long-term learning (e.g., Bjork, 1999). Of course our findings were obtained in a set of verbal learning tasks, using middle school and college students, and they may not apply in other learning domains or with other populations.

Conclusion

The principles tested in the current experiments—memory retrieval, making errors, and feedback—are three among many principles with the potential to contribute to educational practice, as well as to students' own self-regulated learning when they do homework. The value of testing those principles in real educational settings is twofold; it validates laboratory research, and it also allows educators to design more adaptive teaching techniques—and therefore, to improve student learning.

References

Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior, 8,* 463-470.

Arnold, K. M., & McDermott, K. B. (2012). Test-Potentiated Learning: Distinguishing Between Direct and Indirect Effects of Tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 940–945. doi:10.1037/a0029199

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of educational research, 61,* 213-238.

Benjamin, A. S, Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127,* 55-68.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123-144). Hillsdale, NJ: Erlbaum.

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 396-401). New York: Wiley.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher and A. Koriat (Eds*.) Attention and performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). Cambridge, MA: MIT Press.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2012). Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology, 64*, 417–444. doi:10.1146/annurev-psych-113011-143823

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research, 65,* 245-281.

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1491–1494.

Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning, 1,* 69-84.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34,* 268-276.

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review, 13*, 826-830.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633–642.

Carroll, M., & Nelson, T. O. (1993). Failure to obtain a generation effect during naturalistic learning. *Memory & Cognition, 21,* 361-366.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist, 43*, 627-634.

deWinstanley, P. A., (1995). A generation effect can be found during naturalistic learning. *Psychonomic Bulletin & Review, 2,* 538-541.

deWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition, 32*, 945-955.

Dunlosky, J., Hertzog, C., Kennedy, M. R. F., & Thiede, K. W. (2005). The self-monitoring approach for effective learning. *International Journal of Cognitive Technology, 10,* 4-11

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20,* 374-380.

Glover, J.A. (1989). The ‘‘testing’’ phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81,* 392–399.

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition, 40*, 505–13. doi:10.3758/s13421-011-0174-0

Guthrie, E. (1952). *The psychology of learning (Rev. Ed.)*. New York: Harper.

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 39*, 290–6. doi:10.1037/a0028468

Hirshman, E. L., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14,* 484-494.

Hogan, R.M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 562–567

Hourihan, K. L., & MacLeod, C. M. (2008). Directed forgetting meets the production effect: Distinctive processing is resistant to intentional forgetting. *Canadian Journal of Experimental Psychology, 62,* 242-246.

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & cognition, 40*, 514–27. doi:10.3758/s13421-011-0167-z

Izawa, C. (1970.) Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology, 83*, 340-344.

Jacoby, L.L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17,* 649–667.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*, 966-968

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory & Language, 35,* 157-175.

Knight, J. B., Ball, B. H., Brewer, G. A., DeWitt, M. R., & Marsh, R. L. (2012). Testing unsuccessfully: A specification of the underlying mechanisms supporting its influence on retention. *Journal of Memory and Language, 66*, 731–746. doi:10.1016/j.jml.2011.12.008

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349-370.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133,* 643-656.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35*, 989-998.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*, 219-224.

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General, 138*, 449-468.

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17*, 493-501.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London: Academic Press.

Marsh, E. J., Roediger, H. L., III, Bjork, R. A., & Bjork, E. L (2007). Memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review, 14*, 194-199.

Metcalfe, J. (l998). Cognitive optimism: Self deception or memory-based processing heuristics? *Personality and Social Psychological Review, 2*, 100-110.

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors and feedback. *Psychonomic Bulletin & Review, 14*, 225-229.

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition, 37*, 1077-1087.

Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based program to enhance study efficacy in a high and low-risk setting. *European Journal of Cognitive Psychology, 19*, 743-768.

Mozer, M. C., Howe, M., & Pashler, H. (2004). Using testing to enhance learning: A comparison of two hypotheses. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 975-980). Hillsdale, NJ: Erlbaum Associates.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science, 2*, 267-270.

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning & Verbal Behavior, 19,* 338-368.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 3-8.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review, 13*, 187-193.

Pashler, H., Zarow, G., & Triplett, B. Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*, 1051-1057.

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*, 243-257.

Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249-255.

Roediger, H. L., & Karpicke, J. D. (2006b). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science, 1,* 181-210.

Roediger, H. L., III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31*, 1155-1159.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Skinner, B. F. (1958). Teaching machines. *Science, 128*, 969-977.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory, 4,* 592-604.

Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory & Language, 26,* 589-607.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6,* 175-184.

Terrace, H.S. (1963). Errorless transfers of a discrimination across two continua. *Journal of the Experimental Analysis of Behavior, 6*, 223–232.

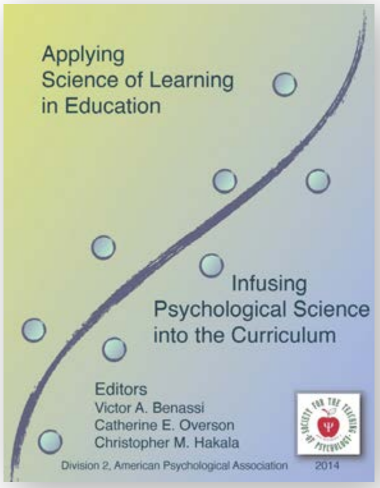
Vaughn, K. E., & Rawson, K. a. (2012). When is guessing incorrectly better than studying for enhancing memory? *Psychonomic Bulletin & Review, 19*, 899–905. doi:10.3758/s13423-012-0276-0

Author Note

Nate Kornell, Department of Psychology, Williams College, and Janet Metcalfe, Department of Psychology, Columbia University.

The research presented here was supported by CASL grant R305H060161 from the Institute of Educational Sciences, Department of Education, and grant 220020166 from the James S. McDonnell Foundation. The authors are solely responsible for the results and interpretations presented here. We are grateful to Lisa K. Son for her help.

Correspondence should be addressed to Nate Kornell, Department of Psychology, Williams College, 18 Hoxsey Street, Williamstown, MA 10267. Email: nkornell@gmail.com.



Applying Multimedia Principles to Slide Shows for Academic Presentation

Catherine E. Overson
University of New Hampshire

Introduction

Multimedia instructional materials are used with growing frequency in a variety of college and university courses and settings (e.g., Cudd, Lipscomb, & Tanner, 2003; Govaere Jan, de Kruif, & Valcke, 2012; Issa, et al., 2011; Kennedy, Pullen, Ely, & Cole, 2013;). Motivated by a desire to enhance student learning, faculty may be inspired to tap into one or more of the many available cutting-edge multimedia technologies to augment their courses. As we prepare material for our presentations, however, we should take care that our multimedia instruction is informed by what is known about how people learn, and in particular, by how people learn with words and pictures – that is, by the cognitive theory of multimedia learning.

Multimedia learning refers to factual, conceptual, or procedural understanding that students develop from both words (either written or spoken) and pictures (e.g., illustrations, diagrams, animation, video, etc.) (Mayer, 2005; 2009). The theory of multimedia learning is founded upon guiding principles in cognitive science relating to how people receive and process information. These principles assume that we receive visual and auditory information from two distinct channels, both of which have a limited capacity (see Mayer, this volume). Overloading either of these channels can interfere with learning. Multimedia instruction, based on how people learn (the science of learning), deals with effective and efficient methods to present material in ways that meet learning objectives and promote meaningful learning.

Much of what we know about the learning benefits associated with successful application of multimedia principles to learning materials comes to us through a substantial body of systematic, controlled, and well-documented experiments in which students were provided brief lessons (e.g., Harp & Mayer, 1997, 1998; Jamet & Le Bohec, 2007; Mayer, 2005; Moreno & Mayer, 1999). These evidence-based accounts on the powerful effects of the application of multimedia principles tell us that people learn more and are better able to apply what they have learned when they are instructed with both words and pictures than when they are instructed with words or pictures alone (Mayer, 2005; 2009).

Whether the learning benefits associated with multimedia principles generalize to authentic classroom settings is only recently being examined. In the following study, for example, two groups of third-year medical students listened to identical 50-minute lectures augmented with multimedia slides on the topic of shock (Issa, et al., 2011). One group viewed a set of slides from the prior year (traditional slides) and a second group viewed a set of slides with the same information; however, their set had been modified so that they were consistent with a combination of Mayer's principles of multimedia learning. One hour after the lecture, students took a test on information they had received. Students who viewed the modified slides performed significantly better on tests of retention than students who viewed the traditional slides. This ecologically sound study very nicely demonstrated learning benefits associated with the straightforward application of multimedia principles to the slideshow presentation. Expanding beyond this study takes into account the following considerations: This study examined medical students' learning on only one topic among many in a series of core curriculum lectures. Participants were graduate students; this gives rise to the question as to whether undergraduates with a variety of background knowledge, skills, and abilities would demonstrate similar learning benefits. In addition, all students were not exposed to both of the multimedia presentation conditions. Finally, evaluation of student learning was limited to short-term learning and retention. The following project addressed each of these considerations, and was designed to evaluate whether Mayer's multimedia principles (2009) hold up in authentic, ongoing undergraduate university courses, and demonstrate learning benefits to both immediate learning and retention of knowledge over an extended period of time.

This Project

The aim of this project was to assess the educational impact of applying the cognitive principles of multimedia learning to existing slide presentations (e.g., *PowerPoint*©, *Keynote*©) used by faculty in their courses. University of New Hampshire faculty who routinely use multimedia in their courses were invited to participate. Four faculty members representing four academic fields participated in this project. Each of the instructors had been using their multimedia slides to present course material for several semesters, and all were interested to determine the effectiveness of their slides in promoting student learning. Instructors in this project taught courses in introduction to justice studies (JS), science and nature of human stress (HS), introduction to child life (CL), and epidemiology and community (EPI).

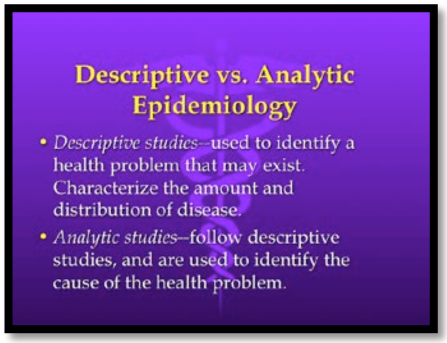
Prior to beginning the project, instructors made available to me their course syllabi, including the schedules for slide presentations. For all multimedia-related course materials leading up to a scheduled in-class exam, half of each of the instructor's original course daily slide presentations were randomly selected to present to students "as-is," and the other half were modified by me according to Mayer's (2009) cognitive principles of multimedia learning. This method of slide show randomization ensured that all students in the course experience both conditions of slide presentation. These modified slides were subsequently returned to the instructor for class presentation. Information content as set by the instructors was neither deleted nor expanded upon on the modified slides. Refer to Figure 1 for examples of original and modified slides used in this project.

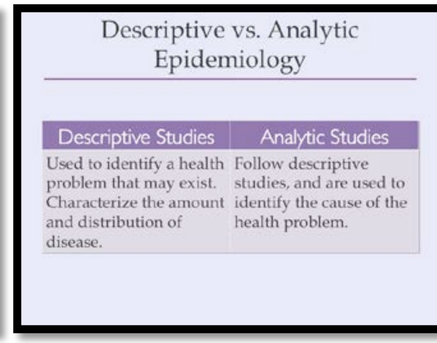
Figure 1: Examples of Original Slides and Slide Modifications Using Mayer's Multimedia Design Principles (Mayer, 2009)

Original | Modified

(1) *Multimedia Principle*: People learn better from words and pictures than from words alone (p. 223).

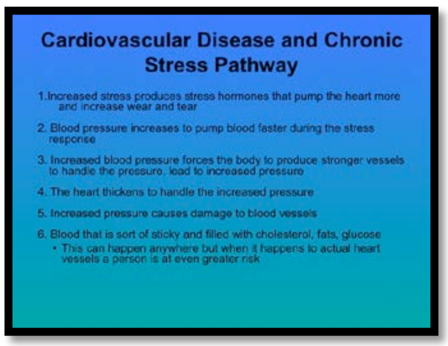
Coherence Principle: People learn better when extraneous material [words and pictures; sounds and music; unneeded words and symbols] is excluded rather than included (p.89).

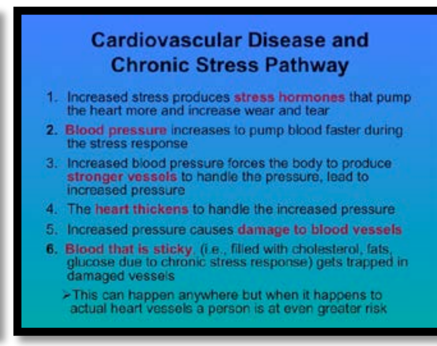




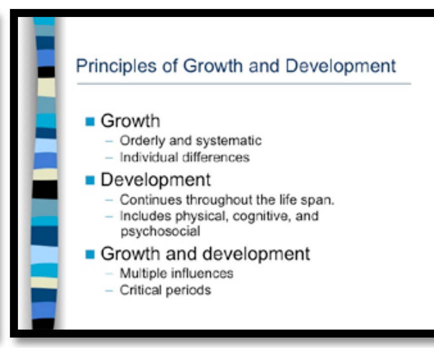
(2) Signaling Principle: People learn better when cues that highlight the organization of the essential material are added (p. 108).

Example 1

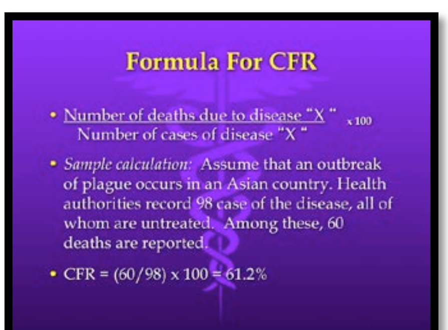


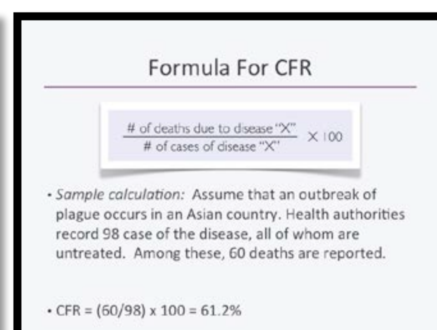


Example 2



(3) *Coherence Principle*: People learn better when extraneous material [words and pictures; sounds and music; unneeded words and symbols] is excluded rather than included (p.89). *Signaling Principle*: People learn better when cues that highlight the organization of the essential material are added (p.





108).

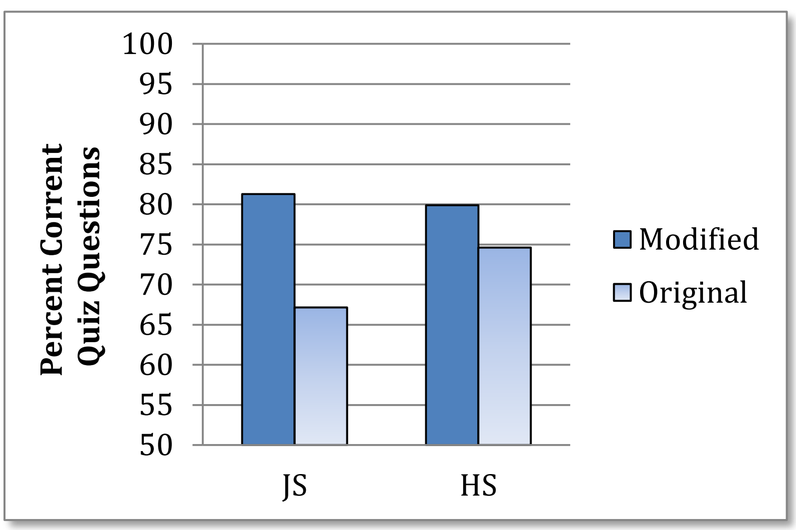
Study 1

A central goal of education is that students achieve long-term retention of course-related learned material. Proceeding from this premise, in order for students to retain multimedia-presented information, they must first acquire an understanding of that material. With this in mind, Study 1 focused on students' initial factual and conceptual understanding of multimedia-presented course material. Specifically, this study was designed to determine whether there would be an immediate learning benefit for undergraduate students in an authentic, on-going classroom setting related to course material on slides that were modified according to Mayer's (2009) multimedia principles compared to course material on instructors' original slides.

Faculty in two courses, JS (134 students) and HS (125 students) participated in this study. As noted above, for multimedia-presented information leading up to a scheduled course exam, a random half of the slide shows were presented as-is and the other half were presented in the modified format. In order to evaluate student learning, instructors presented daily "quizzing" on both original and modified slide material at the end of each class. All students enrolled in these courses – and who responded to the questions – were entered into the study. Quizzes comprised two to four questions in multiple-choice, true false, or short answer formats.

Students in the JS class performed better on the daily quizzes with the modified slide presentations (M = 81.28%; SD = .15) when compared to their quiz performance on the original slide presentations (M = 67.14%; SD = .22), t (133) = 6.28, p < .0001, d = 1.09 (Figure 2). In the HS class, students also performed better on the daily quizzes with the modified slide presentations (M = 79.88%; SD = .14) when compared to their quiz performance on the original slide presentations (M = 74.60%; SD = .19), t (124) = 2.37, p < .02, d = .44 (Figure 2).

Figure 2: Mean Percent Correct Responses to Questions Asked at End of Class During Which Modified or Original PowerPoint Slides Were Presented. (JS = introduction to justice studies; HS = science and nature of human stress).



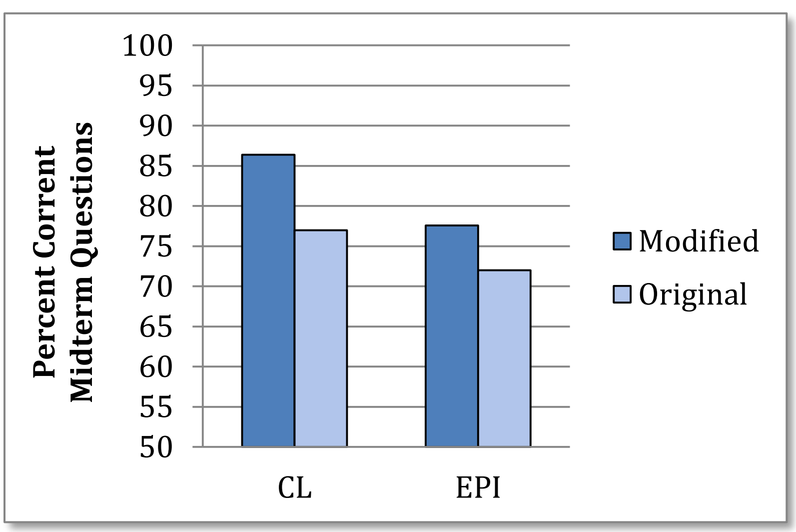
Results of this study clearly demonstrated initial student-learning benefits as measured by quiz performance when slide information was presented in a manner consistent with the cognitive principles of multimedia learning. Whether the impact of multimedia as a teaching tool can affect students' long-term retention of learned and conceptually understood material was the subject of Study 2. In particular, the following study was designed to determine whether there would be a benefit in learning retention as demonstrated by performance on regularly scheduled in-class exams of course-related material on slides that were modified according to Mayer's multimedia principles when compared to course material on instructors' original slides in an authentic, on-going classroom setting.

Study 2

Faculty in two courses, CL (21 students) and EPI (77 students) participated in Study 2, which was designed to evaluate students' retention of factual and conceptual understanding of multimedia-presented course material. As with Study 1, a random half of the slide shows were presented as-is and the other half were modified by me according to Mayer's (2009) multimedia principles. Faculty in these two courses composed for their major in-class exams questions that reflected the body of material presented on both original and modified course slide shows. For this study, in order to make appropriate analyses of student exam performance regarding multimedia presented material, it was necessary to select only those students who attended all classes in which multimedia-related course material was presented. Therefore, in CL, 15 students were included in the analysis, and for EPI, 57 students were included.

Students in the CL class performed better on the exam questions relating to the modified slide presentations (M = 86.41%; SD = .15) when compared to exam questions relating to the original slide presentations (M = 77.04%; SD = .10), t (14) = 3.69, p < .002, d = 1.71 (Figure 3). In the EPI class, students also performed better on the exam questions relating to the modified slide presentations (M = 77.56%; SD = .12) when compared to exam questions relating to the original slide presentations (M = 72.04%; SD = .12), t (56) = 3.50, p < .001, d = .94 (Figure 3).

Figure 3: Mean Percent Correct Responses to Exam Questions Based on Information in Modified or Original PowerPoint Slides. (CL = introduction to child life; EPI = epidemiology and community).



Results of Study 2 clearly demonstrated a superior benefit to student retention of learned material resulting from the application of cognitive principles of multimedia to presentation slides.

Conclusion

These two studies indicate that designing slides that are consistent with Mayer's (2009) cognitive principles of multimedia learning can produce benefits to both short-term learning and long-term retention of academic material. Designing multimedia slides based on the science of learning is not only easy to do, it is an effective and efficient way of promoting learning. See Mayer (this volume) for a review of 12 evidence-based principles detailing how we might design multimedia presentations in ways that promote meaningful learning.

Acknowledgement

This work is supported in part by a grant from the Davis Educational Foundation. The Foundation was established by Stanton and Elisabeth Davis after Mr. Davis's retirement as chairman of Shaw's Supermarkets, Inc.

References

Cudd, M., Lipscomb, M., & Tanner, J. (2003). Technology in the Classroom: An Assessment of Hardware and Software Use in Finance Instruction. *Journal of Education for Business, 78*(5), 244-248.

Govaere Jan, L. J., de Kruif, A., & Valcke, M. (2012). Differential Impact of Unguided versus Guided Use of a Multimedia Introduction to Equine Obstetrics in Veterinary Education. *Computers & Education, 58*, 1076-1084

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal Of Educational Psychology*, *90*(3), 414-434. doi:10.1037/0022-0663.90.3.414

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal Of Educational Psychology*, *89*(1), 92-102. doi:10.1037/0022-0663.89.1.92

Issa, N., Schuller, M., Santacaterina, S., Shapiro, M., Wang, E., Mayer, R., & DaRosa, D. (2011). Applying multimedia design principles enhances learning in medical education. *Medical Education, 45*(8), 818-826. doi:10.1111/j.1365-2923.2011.03988.x

Jamet, E., & Le Bohec, O. (2007). The effect of redundant text in multimedia instruction. *Contemporary Educational Psychology*, *32*, 588-598. doi:10.1016/j.cedpsych.2006.07.001

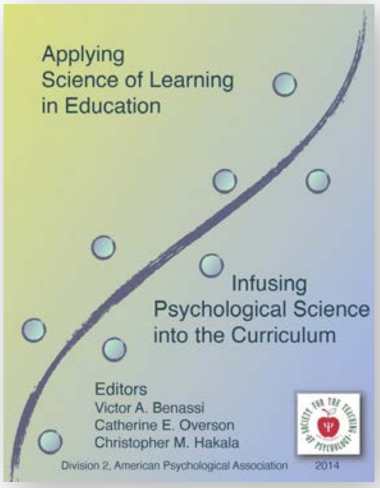
Kennedy, M. J., Driver, M. K., Pullen, P. C., Ely, E., & Cole, M. T. (2013). Improving teacher candidates' knowledge of phonological awareness: A multimedia approach. *Computers & Education*, 6442-51. doi:10.1016/j.compedu.2013.01.010

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. New York, NY US: Cambridge University Press.

Mayer, R. E. (2009). *Multimedia learning (2nd ed.)*. New York, NY US: Cambridge University Press.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43-52. doi:10.1207/S15326985EP3801_6

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal Of Educational Psychology*, *91*, 358-368. doi:10.1037/0022-0663.91.2.358



Student misconceptions: Where do they come from and what can we do?

Annette K. Taylor and Patricia Kowalski
University of San Diego

Overview

Inaccurate Prior Knowledge

Teachers are well aware that education is at least partly a matter of informing students that some of what they think they know just isn't so. From the belief that Columbus fought against those who claimed the earth was flat to students' certainty that increasing self-esteem causes better school performance, inaccurate prior knowledge exists in every domain. Not only are these beliefs pervasive, they can be particularly (and frustratingly) resistant to instruction. In physics (Hake, 1998; Kim & Pak, 2002), biochemistry (Morton, Doran & MacLaren, 2008), history (Leinhardt & Ravi, 2008), and psychology (Kowalski & Taylor, 2009) researchers and even the best instructors (e.g., McKeachie, 1960) find students enter class with misconceptions and leave with them intact. In addition, even when research shows short-term gains in correct knowledge, these gains often disappear over time (Liddy & Hughes, 2012).

This chapter briefly discusses the nature of inaccurate prior knowledge, its likely origins and reasons for its tenacity, as well as pathways for its reduction. We will then address some of our findings on student characteristics contributing to a reduction in misconceptions and describe a technique we have used in our classrooms. As our discussion of the nature of misconceptions will suggest, there are many ways that knowledge can be inaccurate and many explanations for why this knowledge may be tenacious. There are equally many terms used in the literature to denote inaccurate prior knowledge. The variety in terms can lead to much confusion. We will use the term inaccurate prior knowledge as the broadest term referring to students' initial knowledge, when it is inaccurate. Although one rarely knows the exact underlying nature of a student's inaccurate knowledge, we will also differentiate "alternative conceptions" or broad frameworks of mistaken knowledge from misconceptions, more specific inaccurate beliefs.

Traditional Teaching Doesn't Work

Studies documenting the prevalence of students' misconceptions show that even at the college level, the frequency of such beliefs is high; for example, Alparsian, Tekkaya, and Geban (2003) found 41% accuracy on a test of misconceptions in biology and Kowalski and Taylor (2009) found 30% accuracy on a test of misconceptions in psychology. Equally prevalent is the finding that after a traditional course there is little change, as little as 5% in some studies (Gutman, 1979). Clearly just "telling" students the authoritative view, doesn't work to change their misconceptions. Even if students pay attention to the new information they may just add the new claim on top of the old, or they may retain both pieces of information independently. For example, when young students who think of the earth as flat are told the earth is round with no additional explanation, the simple addition of this information can result in a hybrid earth that is flat and round like a pancake or an image of two earths: one which people live on and another which is "a planet" -- a two dimensional upright circle (Vosniadou & Brewer, 1989). In both case, the "earth" is still flat.

Why Do Students Have Inaccurate Prior Knowledge and Why is it So Tenacious?

The key to understanding the source and tenacity of students' inaccurate prior knowledge lies in understanding the nature of student learning. Cognitive science has moved us beyond a stimulus-response view of learning. Rather than seeing learning as a passive addition of new stimulus-response connections, we now understand that learning is an active process (Bransford, Brown, & Cocking, 2000; Ohlsson, 2009; Piaget, 1985). All new knowledge is integrated with, and therefore depends on existing knowledge. What we are able to learn and how we learn it involves our ability to use prior knowledge, to make sense out of new information. When prior knowledge supports or fits with new information it helps in learning. For example, if I am unfamiliar with the "smart phone" you show me, I can use what I know about my older cell phone and my computer to understand some of your device's capabilities.

However, when prior knowledge contradicts new information, as when students have misconceptions, it makes learning the new information more difficult than when students have no knowledge at all (Ambrose, Bridges, Lovett, DiPiero, & Norman, 2010). Using what one knows about human growth and respiration to understand plant growth can lead to false beliefs, such as "Plants use soil for food" and "Plants breathe CO_2." If teachers are unaware of how students are using their prior knowledge then students may incorporate new information into a faulty knowledge base and create additional inaccurate knowledge. This is a process that Ohlsson (2009) calls the "assimilation paradox" because it is seemingly impossible to store new, correct knowledge once a person possesses inaccurate prior knowledge.

Much of the research on conceptual change in science has examined concepts in physics (see Wandersee, Mintzes, & Novak, 1994). Misconceptions in physics are of interest because they are particularly resistant to instruction and because they seem to be based in naïve theories of nature. For example, a student's naïve theory that objects move only when a net force is exerted on them clearly contradicts scientific evidence. The understanding, however, is presumed to derive from the student's experiences within the natural world (e.g., seeing an object that has been pushed across a floor then come to a stop and not seeing friction as the dissipating force). This experience contributes to the development of deeply embedded, pre-instructional theories of natural objects. Altering students' pre-instructional theories is difficult as it requires teachers be both deliberate in identifying the roadblocks to correct conceptions and deliberate in scaffolding the construction of a new theory. Chi (2008), for example, notes that for students to recognize that they have misconceived of force as an entity (naïve theory) and to reconceptualize force as a process (scientific theory), they must understand that these are different categories—entities and processes—and understand what it actually means to conceive of force as a process. Chi claims that when misconceptions are based on such ontological (i.e., categorical) misclassification they are particularly robust and may take years to alter.

Domain Differences

Misconceptions can be tenacious for reasons other than their links with naïve theories (see Chi, 2008; Chinn & Samarapungavan, 2009, for a broader discussion of conceptual change in science). Psychological science, for example, has a long history of looking at student misconceptions (e.g., McKeachie, 1960). In contrast to the alternative conceptions students hold regarding physical phenomena, few of the misconceptions in psychology are embedded in larger theoretically-derived networks of how the world works (Hughes, Lyddy, & Lambe, 2013). Misconceptions in psychology are more often of the type that Chi (2008) and Carey (1986) characterize as “false beliefs,” distinguishing them from “concepts” that are part of stable and coherent explanatory frameworks. A similar argument can be made for many misconceptions that students hold in the fields of history (O’Hara & O’Hara, 2001), religion (Numbers, 2009) and even poetry (Luciarello, n.d.). These also often reflect smaller “beliefs” rather than theoretically-related “concepts.”

In addition, rather than being derived from interactions in a physical world, misconceptions in psychology and other social sciences are more likely derived from and reinforced by social sources. These social sources include folk wisdom, everyday versus domain specific language, personal social interactions, nonscientific sources (i.e., paranormal and religious beliefs), flawed instruction, and media misinterpretation (Landau & Bavaria, 2003; Lewandowsky, Ecker, Seifert, Schartz, & Cook, 2012). Compared with the physical sciences, behavioral science research is, perhaps incorrectly, perceived by most people as easily comprehended and applied. The validity of research findings is often judged in terms of consistency with intuitive ideas. Journalists, in their reporting, will reinterpret (misinterpret) behavioral science findings, making the claims more consistent with what is believed to be common knowledge about human behavior (Thompson & Nelson, 2001). Furthermore, media reports of behavioral science findings often overgeneralize research findings and overextend the reality of potential applications. Of note is the popular use of neuroimaging studies to overgeneralize mental behaviors based on neural activity and the tendency to overgeneralize from research on neuronal development to the education of young children (see Willingham, 2012 and Satel & Lilienfeld, 2013 for comprehensive discussions). The resulting media-created misconceptions are frequently repeated, often by varied and respected sources. The repetition contributes to familiarity and to the mistaken assumption that everyone “knows” (see Lewandowsky, et al., 2012). With the array of uninformed sources, it’s no wonder that students enter the psychology class with a great amount of inaccurate information.

The social nature of misconceptions in the behavioral sciences also contributes to the difficulty in their elimination. In the physical sciences, alternative conceptions tend to be connected in a coherent way, forming conceptual networks. In the behavioral sciences, however, misconceptions are more likely interconnected with other things the student believes as well as with a related affective network. These connections are likely to include attitudes, values, or spiritual beliefs, all of which may have personal and affective relevance or are associated with social groups with which one identifies. Often such misconceptions are strengthened by their connection with other misconceptions. For example the claim that one cannot simultaneously believe in God and in evolution is connected with the misconception that the major religions reject evolution (Numbers, 2009). The intertwining of beliefs results in students becoming deeply committed to their misconceptions and resistant to change (Sinatra & Taasoobshirazi, 2011). According to the literature in social psychology (Tavris & Aronson, 2007), when a person receives information that violates personal beliefs, that is inconsistent with values, or that is not supported by ones’ social network, then a feeling of uneasiness or dissonance occurs. One way to reduce the feeling of uneasiness is to discount the contradictory information. For example, when a teacher’s admired principal advocates teaching to students' learning styles, contradictory information about the lack of evidence for the need to match teaching styles to learning styles might be perceived as coming from an untrustworthy source, which would reduce dissonance. As a result, contradictory information, regardless of its scientific merit is dismissed and unlikely to result in any change in prior knowledge.

If Misconceptions are Connected to Some System of Beliefs for Students, What Conditions are Necessary for Students to Acquire New Knowledge Systems?

The challenge of helping students learn when learning involves changing misconceptions is the focus of research in a number of areas—not only in the cognitive psychology, science education, and reading research literatures, but also more recently in the literature on the teaching of psychology. Cognitive psychology and science education literatures in particular describe models for changing misconceptions, known as models for conceptual change. By understanding why conceptual change is hard, instructors can better guide student discussion and design activities that will promote change.

In an early view of conceptual change, Posner, Strike, Hewson and Gertzog (1982) described conditions necessary for students to successfully change their existing preconceptions in science. According to this view, in order to change from a prior nonscientific conception to a new scientific understanding, learners must experience dissatisfaction with their prior conceptions. To accept an alternative conception, however, the learner must comprehend the new concept, see it as plausible, and believe the new concept will be more useful than the prior conception. In a more recent description of conceptual change, Ohlsson (2009) emphasized that for change to occur, the new information must have greater cognitive utility—that is, students must see that the new information is more useful in everyday life, such as for making simple predictions. These conditions for change are important no matter the nature of the misconception. To take an example from psychology, the common assumption that "Opposites attract" contradicts the empirical finding that similarity underlies relationships. Knowledge about attraction in social situations becomes more useful when students understand why their relationships may or may not work out based on correct knowledge of the effects of interpersonal similarities versus interpersonal differences.

Although Posner et al. (1982) proposed conditions for change rather than pedagogies for change, the conditions provide guidelines for changing student misconceptions. Wandersee, et al. (1994) and Taber (2011) summarized a number of claims regarding students' alternative conceptions and described several pathways to change. These pathways focus on externalizing and monitoring learner knowledge. Empirically-derived pedagogies include concept mapping (Novak, 1996), in which students draw maps to illustrate how they visualize relations among concepts and "ConcepTests," which require students to evaluate alternatives (Chew, 2005). Many techniques emphasize the use of hands-on rather than lecture formats (Hake, 1998; Wieman & Perkins, 2005), specifically the Hypothesis-Experiment-Instruction method (Sokoloff & Thornton, 1997), and teacher led discussions of misconceptions (Alparsian, et al., 2003).

What these approaches have in common is that they provide opportunities for students to think about their preconceptions; this activation can help them attend to and notice the inadequacy of their misconception. It is important to note, however, that "activation" alone will not produce change. Students have to be aware of the misconceptions but they also have to be provided with and comprehend an acceptable alternative. The importance of providing students with an acceptable alternative suggests why student discussions alone don't work in reducing misconceptions. In cooperative learning groups, students may fail to notice key distinctions between the alternatives they are given or the most convincing and persuasive members may convince the others of their own erroneous conclusions. “Discovery learning,” where students alone must figure out the principles that underlie scientific phenomena, can also backfire (Klahr & Nigam, 2004). Researchers tend to find that students need specific guidance to direct their attention to scientific concepts that they are unlikely to develop on their own and to confront problems with their preconceptions. Guidance can come in the form of teacher direction (Marshall, 1989) or guided discussions that focus on the constructive argumentation where students are encouraged in their discussions to make clear assertions and provide evidence (Chinn & Samarapungavan, 2008). The point is, it is not just students who need to become aware of their misconceptions. Teachers must be aware of how students are misconceiving claims and how these misconceptions are affecting students’ learning in order to provide conditions for change.

Description and Discussion

Refutational Teaching

Because students’ active engagement in learning is key in altering misconceptions, hands-on participation dominates suggestions for teaching for change. However, most teachers cannot do activities for every misconception; activities are time consuming and class time is limited. Teachers must consider other ways to promote change, including the use of refutational pedagogies. Refutational pedagogies appear widely in the reading education literature (see Guzzetti, Snyder, Glass, & Gamas, 1993). These techniques include use of both readings and lectures that call students’ attention to their misconceptions and present the scientific evidence in a way that is designed to lower the status of the old view while raising the status of the new view. The refutation serves as a warning to students to block them from relying on their prior beliefs (Chinn & Samarapungavan, 2008). Several examples of refutation text appear in Lilienfeld, Lynn, Namy and Woolf’s (2010) introductory psychology text. For example, in regards to the myth that we only use 10% of our brains, the text begins by drawing students’ attention to the commonly made claim that humans use only 10% of the brain. The authors discuss the origins of the myth and how it acquired urban ledge status. They next show how this claim is in fact a misconception by detailing what the evidence suggests about brain activity.

In an experimental study, Bennett (2010) provided students with readings addressing eight different misconceptions. The readings were either refutational or standard text and students read them in enclosed cubicles. The students were aware that a test over the readings would follow both immediately after completing the readings and one week later. Pretest-posttest gain scores on a 16-item misconceptions test showed significant improvement with refutational text. Standard text showed some improvement, but did not reach statistical significance. Bennett concluded that if you can be sure that your students are actually reading, then refutational text alone may be sufficient to bring about change. Unfortunately, in the context of a real-world classroom, teachers can seldom be assured that all students have carefully completed all readings.

Despite its potential value, refutational text alone does not always dispel myths. Several studies have shown that some people may stop actively reading once they have activated their prior knowledge, regardless of its accuracy. Skurnik, Yoon, Park and Schwarz (2005) examined the use of refutational text in the real world context of providing information on flu vaccinations. They concluded that some readers find support for their nonscientific ideas and therefore fail to change their prior knowledge when it is inaccurate.

At other times the refutation is not sufficiently direct, leaving room for students to favor their misconceptions (Braasch, Goldman, & Wiley, 2013). Chinn and Brewer (1993) pointed out that even when an alternative is available and intelligible, a student is not likely to be persuaded if the idea does not seem plausible or seem to have any value. Confronted with contradictions, students may reject, ignore, or reinterpret the new information, allowing them to hang on to their prior knowledge, even when it is inaccurate (cf. Hakala & O'Brien, 1995). For example, Scarr (1997) found that people who have negative intuitions about day care continue to favor their personal intuitions rather than accept the data from scientific studies. This suggests that individuals rely on more than just scientific evidence—their day-to-day personal experiences seem more salient, important and immediate.

Learner Characteristics

Refuting misconceptions by identifying them and then providing the alternative is critical for students' belief change. Without specifically explaining that the information is false, the error may never occur to the student. Without a detailed description of an alternative, the student is unlikely to generate one independently. Whether the alternative view, however, is intelligible, plausible and useful will depend on both the information presented and on the student. Strike and Posner (1992) acknowledge this interaction in discussing the "ecological context" of conceptual change. This context not only includes prior and competing concepts, it also includes student goals and motivations. Views of conceptual change now consider such motivational and affective variables as "hot" or "warm." These variables, in addition to "cold" or rational cognitions, affect whether or not people will revise their knowledge, (Pintrich, Marx, & Boyle, 1993; Sinatra, 2005).

Much of the research on learner characteristics and belief change is correlational and should be cautiously interpreted. It consistently shows, however, that students' dispositions to evaluate claims and their motivation to evaluate claims predict whether they alter their inaccurate prior knowledge following instruction (Mason & Gava, 2007; Kowalski & Taylor, 2011b). We have examined a number of such student characteristics including openness to new experiences, a part of the Big-5 personality traits (Taylor, Kowalski, & Bennett, 2010) and critical thinking. Looking at whether students' thinking predicted change in their misconceptions, we found that at any level of ability, students who scored higher on a measure of critical thinking were more likely to alter mistaken belief following instruction (Kowalski & Taylor, 2004). We have also looked at student beliefs about the nature of knowledge (i.e. epistemic beliefs) in relation to change (Kowalski & Kowalski, 2013). Consistent with findings in other domains (e.g., Mason 2000), we found that students' belief that knowledge about psychology is complex and changing rather than simple and certain predicts whether they alter their beliefs following instruction.

Although students may have dispositions toward evaluating claims, engaging in evaluation involves effort. Not all students are motivated to exert the effort needed to critically evaluate and alter beliefs. Student learning goals can make a difference in whether they are motivated to engage in the task of evaluating claims and changing misconceptions (Sinatra & Mason, 2008). In our own classrooms we found that students' preference for surface learning strategies ("I try to memorize everything I think will be on the test") predicted their failure to change their misconception. Students' endorsement of mastery goals ("My goal is to master this material") and engagement in effort regulation ("I keep working even when it's hard") predicted their tendency to change and then retain correct knowledge even a semester following their course (Taylor, Kowalski, & Bennett, 2010). These relations are likely indirect, with goals influencing strategies, and strategies influencing performance (Elliot, McGregor, & Gable, 1999). The key is that changing mistaken beliefs is likely to involve work. The students who are motivated to do the work are more likely to change.

Finally, because changing beliefs involves comprehension of the alternative, student abilities must also be taken into account. Although mistaken beliefs appear initially among both high and low ability students, change is more likely for students of higher ability. Some of the poorer students may have reading strategies that are insufficient to notice contradictions between conceptions. They may be less likely to process the meaning of the text for their beliefs, and as a result, they may be unlikely to change their belief (Guzzetti, 2000). In our own classes, using SAT scores and GPA, we found that lower achieving students relative to higher achieving students were less likely to change inaccurate prior knowledge when only provided with refutational text. The lower achieving students needed the lecture, presumably to organize and help direct their thinking, in order to alter their misconceptions (see Kowalski & Taylor, 2011a).

Beyond Refutational Teaching

Given that learner characteristics are likely to contribute to whether students change inaccurate prior knowledge, can teachers encourage knowledge change by addressing student skills? We recently addressed the question of whether improvement in students' knowledge of research methods in psychology can increase the likelihood that they will change their inaccurate prior knowledge (Lessard, Taylor, & Kowalski, 2013). Students in our introductory psychology courses completed pre- and post-test assessments of research methods knowledge and a misconceptions questionnaire. During the semester we used refutational text and teaching to address misconceptions and classroom activities to promote achievement of APA Guidelines for the Undergraduate Psychology Major Goal 2: Research Methods in Psychology to "understand and apply basic research methods in psychology, including research design, data analysis, and interpretation." Using gain scores we found that even after controlling for ability with student SAT scores, increasing knowledge of research methods predicted changes in misconceptions. That is, students who showed the most gain in knowledge of research methods also showed the greatest reduction in misconceptions.

The instructional implications are that learner characteristics such as goal orientation, predisposition to evaluate new information, ability and goal orientation all make a difference and must be taken into account in attempting to alter students' mistake beliefs. In addition to the activation and identification of a misconception, we need to know how the student evaluates claims and whether the student is motivated to evaluate claims. It is likely that developing students' ability to critically evaluate claims will contribute to their ability to relinquish misconceptions following instruction. This claim is subject of continuing research (e.g., Nussbaum, Sinatra, & Poliquin, 2008).

Advice

Given that students are likely to enter class with many inaccurate beliefs and that traditional teaching methods are unlikely to alter these beliefs, what can teachers actually do in their classrooms to address student beliefs? Can the theories and research on changing false beliefs be effectively applied in the classroom? In our research we have looked at whether the claims from the cognitive, science, and reading education literatures are supported when addressing change in psychological science misconceptions. Our findings provide direct information on how our teaching affects whether our students change their misconceptions following instruction.

Our Research

Much of our research takes place in real classrooms with real students and not in the laboratory. We sacrifice the control of the laboratory for the ecological validity we gain from real world contexts. In the real world students don't always do what we would expect in an ideal world, such as carefully read all materials before class and come prepared to discuss, critique and analyze.

In our research we have identified misconceptions, the degree of confidence students have in them, and their probable source. We have also investigated predictors of change in such knowledge including pedagogy and student characteristics. Across a number of controlled studies, our method always begins with an assessment of what students think they know about psychology. In our earlier work we used a standard true/false format in which students read a statement, most often stated as a misconception, such as "Subliminal messages can induce people to buy things that they would not ordinarily purchase," and then noted whether they believed the item to be true or false (Kowalski & Taylor, 2009). This format suffers from several psychometric problems; therefore, more recently we have adopted a 2-option forced-choice format we call the A/B format (see Taylor & Kowalski, 2012 for a discussion and assessment instrument). The corresponding sample item for the subliminal message example reads,

> "Which is most true about subliminal messages, such as suggestions that the viewer purchase some product?
>
> A. Subliminal messages have little effect on people's consumer behaviors.
> B. Subliminal messages can induce people to buy things that they would not ordinarily purchase."

In this format students must endorse which of the two statements they believe is correct, with the two statements including the common misconception and a correct statement based on evidence. This format allows students to think about each item fully whereas the true/false format allowed students to perhaps endorse a misconception without fully thinking through the item. Compared with the true/false format, the A/B format appears to be a more accurate assessment of students' level of misconceptions.

Following completion of the pretest we address items using five different classroom pedagogies. Four of these five include pedagogies that addressed the misconceptions, but in different ways; the fifth is a control condition. We begin with a standard text for introductory psychology and supplement with readings. Some of the text readings and some of the additional readings provide correct information to the students without directly addressing any misconceptions. This is the way most texts are written. Other text and readings address the misconception in a refutational format, activating the misconception and then providing the scientific evidence that fails to support this inaccurate knowledge or supports a more scientific conclusion. The textbook we generally rely on contains some but not a predominant amount of refutational text, allowing us to target misconceptions either refutationally or with standard teaching. Supplemental refutation and standard readings come from a variety of sources including periodicals, supplemental texts, and websites. Lilienfeld, Lynn, Ruscio, & Beyerstein's (2010) 50 Great myths of popular psychology, for example, provides refutational text for a number of misconceptions.

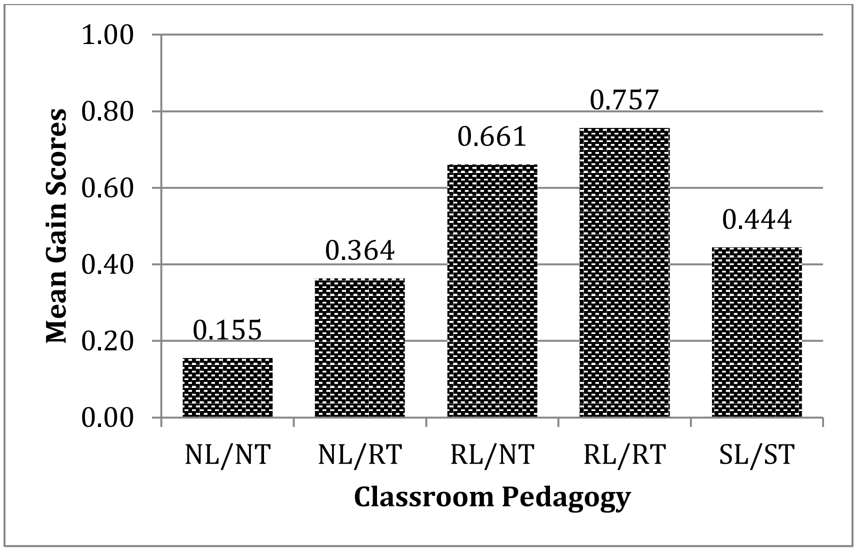
In one pedagogical condition, we do not address the misconception at all, either in lecture or by providing readings. We call this condition NL/NT (no lecture/no text) as a baseline for comparison to the other four conditions. In a second condition we provide students with both a standard classroom lecture and standard text. In this approach students read and hear about the correct information, however, we do not address the misconception. We merely provide current evidence-based information. This is the standard or traditional way of teaching in which most instructors engage. So, for example, when addressing subliminal messages in the SL/ST condition (standard lecture/standard text) we discuss the empirical findings of studies on subliminal perception but not the failure of studies to replicate alleged findings of the influence of subliminal messages on consumer behavior. In other conditions, as noted below, we do so.

The next three conditions involve a combination of refutational text and refutational lecture combined with either no readings or no lecture on specific topics. For example, for the subliminal message example we provide readings regarding the history of the alleged Vicary studies on popcorn and soda sales in movie theaters (see Pratkanis, 1992). In class we have a discussion of the limits of subliminal perception findings and go over the details of the Vicary story. This is the refutational lecture/refutational text (RL/RT) condition. When appropriate, depending on the specific topic, we talk about the common biases in thinking that might support confident belief in misconceptions, such as confirmation bias (sugar causes hyperactivity in children) or the availability heuristic (many criminals are acquitted based on the insanity defense) or using simple logic of what else must be true if the inaccurate prior knowledge were to be true (listening to Mozart's music increases intelligence in infants). In the subliminal message example we talk about how findings make the news because of their shock value, but that retractions seldom make the news—and when they do they are disconnected over time. The important point of these lectures is that they involve some type of stoppage of "telling" for students to interactively think about and evaluate the information. We use explicit slides with embedded thinking points and simple pair-and-share which only takes a couple of minutes. Many instructors fear that if they must address each misconception individually then it reduces their time to cover core information; our experience is that this need not be the case at all. We can cover any given misconception in a refutational manner with just a few minutes of class time. Finally, in the refutational lecture/no text (RL/NT) condition we only have classroom coverage and in the no lecture/refutational text (NL/RT) we provide refutational readings but have no classroom coverage.

Students complete a posttest at the end of the semester. We analyze our data using gain scores which allow us to determine each student's change using their own pretest performance as a baseline for comparison. We calculate gain as (posttest score – pretest score) / (maximum score – pretest score). The score is a ratio of how much any given student has improved relative to how much he or she could have improved. A simple posttest minus pretest score does not take into account the different levels of individual prior knowledge. The variability is quite great as some students come in scoring zero on the pretest, whereas we have had students score above 70% on the pretest. This means the range of improvement could run from 30% to 100% if we did not take account of this initial variability.

Our findings

We have consistently found, both using the older true/false format and the new A/B format, that students show the greatest change when we use both refutational readings and refutational lectures. They show the least improvement for those items which we fail to address or to provide readings for. Using standard approaches is significantly better than no coverage; however, it results in significantly less change than using combined refutational readings with refutational lecture and significantly less than having no readings but having a refutational lecture. Finally, having refutational readings without support in the classroom produces about the same amount of change as do the standard coverage conditions. Our most recent results can be seen in Figure 1.



These data have not yet been reported elsewhere. A repeated-measures analysis of variance was significant with $F(1,104) = 108.98$, $p < .001$, $\eta2 = 0.51$. Paired t-tests with a Simes correction showed that all conditions were significantly different from one another except for NL/RT (no lecture, refutational text) with SL/ST (standard lecture and text). This pattern of results is similar to what we found in Kowalski and Taylor (2009) using the true/false question format, however now we have a psychometrically better measure. The end result shows that refutation works and it works best in the real world when we have readings supported by lectures that require students to stop and think.

Summary

1. We have reviewed the evidence showing that students, across a range of disciplines, come into their introductory courses with a number of misconceptions. We advise all instructors in all disciplines to be aware of common misconceptions in their area and assess the degree to which students endorse these in their classes.This is important not only in introductory level courses; we have found widespread misconceptions in upper division courses as well (c.f., Kowalski & Kowalski, 2013). Teachers must assess students' prior knowledge and they must directly address the misconceptions. Otherwise students are likely to leave the classroom with their misconceptions intact, interfering with the learning of new, accurate knowledge.

2. These misconceptions matter because all new knowledge is interpreted, analyzed and stored relative to existing knowledge. When the prior knowledge is inaccurate, then the new knowledge can be misinterpreted or can be incompletely or inaccurately stored as well, resulting in an even larger inaccurate knowledge base. We have reviewed the evidence, much of which we have replicated in our own classroom-based studies, that these misconceptions do NOT change if not directly, repeatedly and actively addressed. Simply providing readings in the context of a real world classroom or simply providing correct information about some phenomenon in class is not sufficient.

3. Changing these misconceptions is very difficult; both the correct and the inaccurate conceptions can co-exist for years before the correct information comes to dominate. This is why instructors must continue to address these in advanced classes. Over time, students are apt to revert to their prior misconceptions, if they ever truly gave them up. We currently have an on-going longitudinal investigation of this very question of what happens over the long term.

4. It is not just the cognitive nature of the information that determines the ease or difficulty of changing misconceptions into correct knowledge. Various characteristics of the learner, such as the disposition and ability for critical thinking, the disposition and ability for scientific thinking, students' epistemological beliefs and their overall intellectual ability as well as their learning goals and strategies applied to reach those goals all interact with the probability of whether or not such change can and will occur. It is important to be particularly aware that many students may have shallow learning strategies as they pursue performance-oriented goals; these strategies may have served the students well in their prior education, especially if high-stakes testing was involved.

5. There are many pathways to change which reflect the findings that there are many types of misconceptions—whether these be single beliefs, organized networks of beliefs or networks of beliefs organized around a larger theoretical structure. We have had success with, and advocate the use of refutational teaching—the combined use of refutational readings which are supported by refutational lecture. Although refutational readings alone may be sufficient for some students, others require instructor lead lecture and discussion. Refutational lecture benefits from interactive components but requires that the instructor have control of the direction of learning in these aspects of learning that require changing prior knowledge. Other methods should be explored, depending on the discipline. The point of any teaching should be to get students to actively evaluate their prior beliefs and knowledge and then to enlarge their knowledge base with accurate information.

References

Alparsian, C., Tekkaya, C., & Geban, O. (2003) Using conceptual change instruction to improve teaching. *Journal of Biological Education, 37*(3), 133-137.

Ambrose, S. A., Bridges, M. W. B., Lovett, M. C., DiPiero, M., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching.* San Francisco, CA: Jossey-Bass.

Bennett, S. (April, 2010). *Using refutational reading to enhance conceptual change.* Poster presented at the 90th Annual Convention of the Western Psychological Association, Cancun, MX.

Braasch, J. L. G., Goldman, S. R., & Wiley, J. (2013). The influences of text and reader characteristics on learning from refutations in science texts. *Journal of Educational Psychology, 17*, doi:10.1037/a0032627

Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Carey, S. (1986). Cognitive science and science education. *American Psychologist, 41*, 1123-1130. doi.org/10.1037/0003-066X.41.10.1123

Chew, S. L. (2005). Seldom in doubt but often wrong: Addressing tenacious student misconceptions. In Dunn, D. S., & Chew, S. L. (Eds.). *Best practices in teaching general psychology* (pp. 211-223). Mahwah, NJ: Erlbaum.

Chi, M. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.) *International handbook of research on conceptual change* (pp.61-82). New York: Routledge.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition" A theoretical framework and implications for science instruction. *Review of Educational Research, 63*, 1-49. doi.org/10.3102/00346543063001001

Chinn, C. A., & Samarapungavan, A. (2008). Learning to use scientific models: Multiple dimensions of conceptual change. In R. A. Duschl & R. E. Grandy (Eds.), *Teaching Scientific Inquiry: Recommendations for Research and Implementation*, Rotterdam: Sense Publishers.

Chinn, C. A., & Samarapungavan, A. (2009). Conceptual change—multiple routes, multiple mechanisms: A commentary on Ohlsson (2009). *Educational Psychologist, 44*(1), 48-57. doi.org/10.1080/00461520802616291

Elliott, A. J., McGregor, H. A., & Gable, S. (1999). Achievement goals, study strategies and exam performance: A mediational analysis. *Journal of Educational Psychology, 91*(3), 549-563. doi.org/10.1037//0022-0663.91.3.549

Gutman, A. (1979). Misconceptions of psychology and performance in the introductory course. *Teaching of Psychology, 6*, 159-161. doi.org/10.1207/s15328023top0603_9

Guzzetti, B. J. (2000). Learning counter-intuitive science concepts: What have we learned from over a decade of research? *Reading & Writing Quarterly, 16*(2), 89-98. doi.org/10.1080/105735600277971

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly, 28*, 116-159. doi.org/10.2307/747886

Hakala, C.M., & O'Brien, E.J. (1995). Strategies for resolving coherence breaks in reading. *Discourse Processes*, 20(2), 167-185.

Hake, R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64-74. doi.org/10.1119/1.18809

Hughes, S., Lyddy, F., & Lambe, S. (2013). Misconceptions about psychological science: A review. *Psychology Learning and Teaching, 12(1)*, 20-31. doi.org/10.2304/plat.2013.12.1.20

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15,* 661-667. doi: 10.1111/j.0956-7976.2004.00737.x

Kim, E., & Pak, S-J. (2002). Students do not overcome conceptual difficulties after solving 1000 traditional problems. *American Journal of Physics, 70(7)*, 759-765. doi.org/10.1119/1.1484151

Kowalski, P., & Taylor A. (2004). Ability and critical thinking as predictors of change in students' psychological misconceptions. *Journal of Instructional Psychology, 31*(4), 297-303.

Kowalski, P., & Taylor, A. (2009). The effect of refuting misconceptions in the introductory psychology classroom. *Teaching of Psychology*, *36*(3), 153-159. doi.org/10.1080/00986280902959986.

Kowalski, P., & Taylor, A. (2011a). Effectiveness of refutational teaching for high- and low-achieving students. *Journal of the Scholarship of Teaching and Learning, 11* (1), 79-90.

Kowalski, P., & Taylor, A. (2011b). *Beliefs, strategies, and motivation as predictors of change in misconceptions.* Poster presented at the 91st Annual Convention of the Western Psychological Association, Los Angeles, CA.

Kowalski, P., & Kowalski, M. (2013). *Predicting change in students' misconceptions about developmental psychology.* Poster presented at the 93rd Annual Convention of the Western Psychological Association, Reno, NV.

Landau, J. D., & Bavaria, A. J. (2003). Does deliberate source monitoring reduce students' misconceptions about psychology? *Teaching of Psychology, 30*, 311-314.

Leinhardt, G., & Ravi, A. K. (2008). Changing historical conceptions of history. In S. Vosniadou (Ed.), *International handbook of research conceptual* change (pp. 328-341). New York: Routledge.

Lessard, L., Taylor, A., & Kowalski, P. (2013). *Gain in knowledge of research methods influences understanding of misconceptions.* Poster presented at the 25th Annual Convention of the Association for Psychological Science, Washington, DC.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Sciences in the Public Interest, 13(3),* 106-131. doi.org/10.1177/1529100612451018

Luciarello, J. (n.d.). *How do I get my students over their alternative conceptions (misconceptions) for learning.* Retrieved on July 12, 2013 from http://www.apa.org/education/k12/misconceptions.aspx?item=1

Lyddy, F., & Hughes, S. (2012). Attitudes towards psychology as a science and the persistence of psychological misconceptions in psychology undergraduates. In V. Karandashev & S. McCarthy (Eds.), *Teaching around the world* (Vol. 3). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Lilienfeld, S. O., Lynn, S. J., Namy, L. L., & Woolf, N. J. (2010). *Psychology: A framework for everyday living*. Boston: Allyn and Bacon.

Lilienfeld, S. O., Lynn, S. J., Ruscio, J., & Beyerstein, B. (2010). *50 great myths of popular psychology: Shattering widespread misconceptions about human behavior*. New York: Wiley-Blackwell.

Marshall, N. (1989). Overcoming problems with incorrect prior knowledge: An instructional study. *National Reading Conference Yearbook*, *38*, 323-330.

Mason, L. (2000). Role of anomalous data and epistemological beliefs in middle students' theory change on two controversial topics. *European Journal of Psychology and Education, 15*, 329-346.

Mason, L, & Gava, M. (2007). Effects of epistemological beliefs and learning text structure on conceptual change. In S. Vosniadou, A. Baltas and X. Vamavakoussi (Eds.), *Re-framing the conceptual change approach in learning and instruction* (pp. 165-196). New York: Elsevier

McKeachie. W. (1960). Changes in scores on the Northwestern Misconceptions Test in six elementary psychology courses. *Journal of Educational Psychology, 51*, 240-244. doi.org/10.1037/h0048569

Morton, J. P., Doran, D. A., & MacLaren, D. P. (2008). Common student misconceptions in exercise physiology and biochemistry. *Advances in Physiology Education, 32*, 142-146. doi.org/10.1152/advan.00095.2007

Novak, J. D. (1996). Concept mapping: A tool for improving science teaching and learning. In D. F. Treagust (Ed.), *Improving teaching and learning in science and mathematics* (pp. 32-43). New York: Teacher's College Press.

Nussbaum, E. M., Sinatra, G. M., Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. *International Journal of Science Education, 30(15*), 1977-1999.

Numbers, R. (2009). *Galileo goes to jail: And other myths about science and religion*. Cambridge, MA: Harvard University Press.

O'Hara, M., & O'Hara, L. (2001). *Teaching history 3-11*. New York: Continuum.

Ohlsson, S. (2009). Resubsumption: A possible mechanism for conceptual change and belief revision. *Educational Psychologist, 44*(1), 20-40. doi.org/10.1080/00461520802616267

Piaget, J. (1985). The equilibration of cognitive structures. Chicago: University of Chicago Press.

Pintrich, P., Marx, R. W., & Boyle, R. A. (1993) Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research, 63*, 167-190. doi.org/10.3102/00346543063002167

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*, 211-227. doi.org/10.1002/sce.3730660207

Pratkanis, A. R. (1992). The cargo-cult science of subliminal persuasion. *Skeptical Inquirer, 16*. Retrieved on June 25, 2013 from http://www.csicop.org/si/show/cargo-cult_science_of_subliminal_persuasion/

Satel, S., & Lilienfeld, S. O. (2013). *Brainwashed: The seductive appeal of mindless neuroscience*. New York: Basic Books.

Scarr, S. (1997). Rules of evidence: A larger context for the statistical debate. *Psychological Science, 8*(1), 16-17. doi.org/10.1111/j.1467-9280.1997.tb00537.x

Sinatra, G. M. (2005). The "warming trend" in conceptual change research: The legacy of Paul R. Pintrich. *Educational Psychologist, 40*(2), 107-115. doi.org/10.1207/s15326985ep4002_5

Sinatra, G. M., & Mason, L. (2008). Beyond knowledge: Learner characteristics influencing conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change*, New York: Routledge.

Sinatra, G. M. & Taasoobshirazi, G. (2011). Intentional conceptual change: the self-regulation of science learning. In B. Zimmerman, D. Shunk (Eds.), *Handbook of self-regulation of learning and performance*, Routledge, New York, NY (2011), pp. 203–216.

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research, 31*, 713-724. doi.org/10.1086/426605

Sokoloff, D. R., & Thornton, R. K. (1997). Using interactive lecture demonstrations to create an active learning environment, *The Physics Teacher, 35*(6), 340-347.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. A. Duschl & R. J. Hamilton (Eds.), *Philosophy of science, cognitive psychology, and educational theory and practice* (pp. 147-176). Albany: State University of New York Press.

Taber, K. S. (2011). Stella Vosniadou (Ed.): International handbook of research on conceptual change. *Science and Education, 20*, 563-576. doi.org/10.1007/s11191-010-9283-6

Tavris, C. & Aronson, E. (2007). *Mistakes were made (but not by me): Why we justify foolish beliefs, bad decisions, and hurtful acts.* New York: Harcourt, Inc.

Taylor, A., & Kowalski, P. (2012). Students' misconceptions in psychology: How you ask matters…Sometimes," *Journal of the Scholarship of Teaching and Learning, 12*(3).

Taylor, A., Kowalski, P., & Bennett, S. (2010). *Goal orientation and learning strategies predict conceptual change over time.* Poster presented at the 90th Annual Convention of the Western Psychological Association, Cancun, MX.

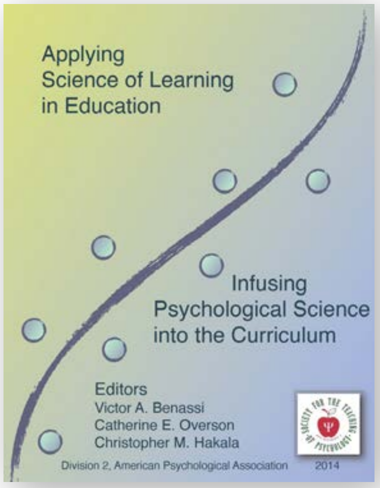
Thompson, R. A.. & Nelson, C. A. (2001). Developmental science and the media. *American Psychologist*, 56(1), 5-15. doi.org/10.1037/003-066X.56.1.5

Vosniadou, S., & Brewer, W. F. (1989). Theories of knowledge restructuring in development. *Review of Educational Research, 57*, 51-67. doi.org/10.3102/00346543057001051

Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative conceptions in science. In D. L. Gabel (Ed.) *Handbook of research on science teaching and learning project* (pp. 177-210). New York: MacMillan Publishing Company.

Wieman, C. & K. Perkins (2005). Transforming physics education, *Physics Today, 58*(11). Retrieved on June 25, 2013 from http://www.colorado.edu/physics/EducationIssues/papers/PhysicsTodayFinal.pdf doi.org/10.1063/1.2155756

Willingham, D. T. (2012). *When can you trust the experts: How to tell good science from bad in education*. San Francisco, CA: Jossey-Bass.



Examining the Benefits of Cumulative Exams

Natalie K. Lawrence
James Madison University

“Will the final exam be cumulative?” This is one of the first questions students ask about a college course. The answer to this single question may determine whether students enroll—or stay enrolled—in a course. Clearly, many students dread taking cumulative final exams.

I recently asked a large group of introductory psychology teachers about their attitudes toward and use of cumulative final exams (Lawrence, 2012). Fifty-seven percent of them believed that students remember more of the course material after taking a cumulative final exam. The same 57% reported that they gave at least one cumulative exam in their most recent introductory psychology course. Of those teachers, 11% gave more than one cumulative exam.

It is interesting that 20% of the teachers I surveyed believed that cumulative exams do *not* enhance long-term retention. There is some research in the field of cognitive psychology that suggests the opposite. For example, Szpunar, McDermott, and Roediger (2007) asked participants to learn a series of word lists and tested them after each list. Then, they administered a cumulative “final exam” 30 minutes after the initial testing. Some participants were aware of the final test; others were not. Not surprisingly, participants who expected the final exam performed better than those who had no expectation. Szpunar et al. argued that the final test encouraged continued processing of the material, which made the words more accessible at the time of the final. This research suggests that the expectation of a cumulative final exam in a college course may change the way students process the material – which could affect their long-term retention of that material.

Szpunar et al.’s (2007) work suggests that the expectation of a final exam may enhance students’ long-term retention. However, taking a “final exam” in the laboratory is very different than taking one in the classroom. Do students really benefit from taking an exam on 16 weeks of material? Given how much students seem to dislike cumulative exams, teachers should make evidence-based decisions about their use. The problem is that there is very little evidence. When I reviewed the literature in 2011, there were no direct, empirical tests of the benefits of cumulative exams in the classroom. For this reason, I decided to investigate the effects of cumulative exams in one of my own courses[1].

I was assigned to teach two equivalent sections of introductory psychology during a single semester. This allowed to me to manipulate the type of exam administered in each section to determine the impact on students' learning. A straightforward manipulation could have involved assigning one section to take a cumulative final exam and one section to take a noncumulative final exam. I could then determine whether taking a cumulative final exam enhanced students' retention of the course material. Although it is important to determine whether Szpunar et al.'s (2007) results generalize to the classroom, I wanted to tackle a slightly different research question. The research participants in Szpunar et al.'s study kept the material accessible because they knew they would be tested on it later. I doubt that this happens for most students taking a final exam in a college course. Students are likely to "cram" for this type of exam rather than continually process the material over an entire semester. Indeed, a single cumulative final exam at the end of a long semester may not be the best strategy to boost long-term retention. A better strategy might be to have several cumulative exams spaced throughout the semester. This notion is consistent with research in cognitive psychology on the spacing effect (see Carpenter, this volume, for a review; Cepeda, Pashley, Vul, Wixted, & Rohrer, 2006) and the testing effect (see Pyc, Agarwal, & Roediger, this volume, for a review; Karpicke & Roediger, 2008).

The purpose of my study, then, was to determine if students who take cumulative exams throughout the semester would have better long-term retention than students who take a single cumulative final exam. Students were 59 women and 46 men enrolled in one of two sections of an introductory psychology course. The breakdown was 53% first-years, 14% sophomores, 26% juniors, and 7% seniors. The two sections were equivalent in every possible way, except for the type of exams they took. Students in one section (the noncumulative section) took three noncumulative exams plus a cumulative final exam. Students in the other section (the cumulative section) took four cumulative exams. If I was able to teach a third section of the course, I would have been able to have a control group in which students took four noncumulative exams. Unfortunately, this was not possible. Teachers who want to conduct research in their classes must do so within the constraints of the "real world" just as other applied researchers must do.

It would have been possible for me to randomly assign students to conditions (instead of randomly assigning sections of students). The random assignment would have increased the internal validity of the study[2] but it would have caused some problems. First, it would have been cumbersome to pass out the exams (wasting precious time on exam days). Second, I believe it would have increased students' sense of injustice. Students realize that different sections are sometimes taught differently but they expect to be treated the same as other students in their class. I don't have any data to support this belief, but it is easy to imagine that some students would be upset if they had to take cumulative exams when the person sitting next to them did not have to. Of course, the students in the cumulative section of my course may have been upset that the students in the other section did not have to. However, I explained the purpose of my research – and its importance – on the first day of class. I told students that taking cumulative exams may improve their final grade in the course. I also told them that if one section did better on the final exam than the other section, I would raise the grades of the students in the lower-scoring section to restore equity. When you conduct research in your own classes, it is important to consider the beliefs and feelings of your students. Students need to know that you have their best interests in mind.

In my study, the first exam was the same for both sections but the second and third exams were different. In the noncumulative section, the exams included 50 multiple-choice questions based on the most recently covered material. In the cumulative section, the exams included 40 questions based on the most recent material and 10 questions based on previously tested material. The final exam was the same for both sections and included 40 questions on new material and 50 questions on earlier material.

Two months after the course ended, I gave students a follow-up test that measured their long-term retention. This test was optional but I gave students three points of extra credit on the final exam if they promised to take the follow-up test. Students who completed the follow-up test were entered into a raffle to win some prizes. I've learned that, if you want to collect data on students after the course ends, it's almost essential that you provide several incentives for doing so.

Previous research shows that low- and high-performing students are differentially affected by classroom interventions (see Lee & Kalyuga, this volume, for a review; Forsyth, Lawrence, Burnett, & Baumeister, 2007). For this reason, I performed a median split on the scores for the first exam and created two groups of students: low-scorers and high-scorers. I then analyzed students' scores on the cumulative part of the final exam (see Table 1). The analysis revealed that students in the cumulative section did better than students in the noncumulative section – regardless of the type of student.

Table 1. Mean percentage correct on cumulative section of the final exam as a function of type of student (low scorers vs. high scorers) and section (cumulative vs. noncumulative).

	Cumulative Section		**Noncumulative Section**	
	M	*SD*	*M*	*SD*
Low-scorers	75.56	9.39	70.96	8.17
High-scorers	84.83	6.88	82.85	6.20

When I analyzed total course grades, I found that high-scorers' course grades were unaffected by section. However, low-scorers in the cumulative section earned a higher average than low-scorers in the noncumulative section (see Table 2). These results suggest that students may benefit more from taking multiple cumulative exams rather than taking a single cumulative exam at the end of the semester, especially those who do not get off to a great start.

Table 2. Mean final course grade as a function of type of student (low scorers vs. high scorers) and section (cumulative vs. noncumulative).

	Cumulative Section		**Noncumulative Section**	
	M	SD	M	SD
Low-scorers	82.66	6.67	77.78	9.39
High-scorers	89.23	6.43	89.08	4.42

What about the long-term benefits of cumulative exams? When I analyzed the scores on the follow-up test, I found an interaction between the type of student described above (low scorers vs. high scorers) and the type of exam they took (see Table 3). High-scorers did well regardless of which exams they took. However, the low-scorers remembered more of the course material after taking multiple cumulative exams. One explanation for this result is that having multiple cumulative exams motivates low-scoring students to engage in behaviors that improve their performance and long-term retention (like spacing out their study and practicing retrieval). The good students do not benefit as much from the intervention because they are already doing the things they need to succeed.

Table 3. Mean percentage correct on the follow-up test as a function of type of student (low scorers vs. high scorers) and section (cumulative vs. noncumulative).

	Cumulative Section		**Noncumulative Section**	
	M	*SD*	*M*	*SD*
Low-scorers	73.38	12.58	65.09	13.40
High-scorers	75.30	11.17	78.56	5.77

I have often questioned whether the benefits of cumulative exams outweigh the costs (especially students' disdain of them). After collecting data in my course, I have the answer. Of course, more applied research is needed to replicate and extend my findings. One recent study found that introductory students benefitted more from a cumulative final exam than students in upper-level courses (Khanna, Badura Brack & Finken, 2013). It is unknown whether students in upper-level course would benefit from taking *multiple* cumulative exams. My hope is that other teachers will conduct experiments in their own courses so that we might better understand the testing conditions that maximize students' learning.

References

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D., (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354-380. doi: 10.1037/0033-2909.132.3.354

Forsyth, D. R., Lawrence, N. K., Burnette, J. L, & Baumeister, R. F. (2007). Attempting to improve the academic performance of struggling college students by bolstering their self-esteem: An intervention that backfired. *Journal of Social and Clinical Psychology, 26(4),* 447-459.

Karpicke, J.D. & Roediger, H.L. (2008). The critical importance of retrieval for learning. Science, 319, 966-968. doi: 10.1080/09541440601056620.

Khanna, M.M., Badura Brack, A.S., & Finken, L.L. (2013). Short-and long-term effects of cumulative finals on student learning. Teaching of Psychology, 40(3), 175-182.

Lawrence, N. K. (2012). [Survey of teachers' attitudes toward and use of cumulative exams.] Unpublished raw data.

Lawrence, N.K. (2013). Cumulative exams in the introductory psychology course. *Teaching of Psychology, 40*(1), 15-19.

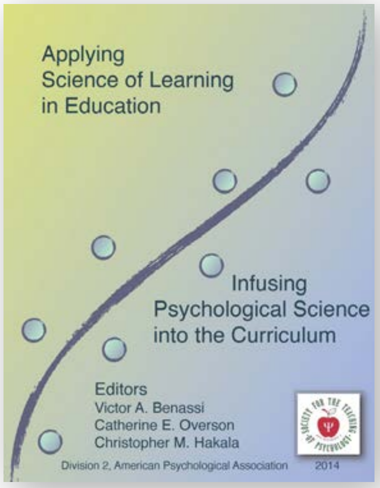
Szpunar, K. K., McDermott, K. B., & Roediger III, H. L. (2007). Expectation of a final cumulative test enhances long-term retention. *Memory & Cognition, 35,* 1007-1013.

Footnote

[1] The full citation for this work is:

Lawrence, N.K. (2013). Cumulative exams in the introductory psychology course. *Teaching of Psychology, 40*(1), 15-19.

[2] To ensure that there were no important differences between the two sections, I compared their scores on the first exam and their demographic information. There were no significant differences on any of these measures. Thus, I have confidence that my results are due to the experimental manipulation – and not to preexisting differences between the sections.



Intervening on Behalf of Low-Skilled Comprehenders in a University General Chemistry Course

Samuel Pazicni and Daniel T. Pyburn
Department of Chemistry
University of New Hampshire

Lana's Story

Imagine a first-year university student—let her name be Lana. Lana is a student in a one-semester accelerated general chemistry course designed for engineering majors. This course placement wasn't her choice, nor was it because she displayed an advanced aptitude for chemistry. Lana was placed in the course because of her intended major. Early on, she recognizes that her high school preparation for such a course was especially limited and begins to regularly visit the instructor's office hours. Lana is clearly a student cognizant of her own weaknesses and ready to work to resolve them—for what more could an instructor ask?

In preparation for the first midterm examination, the instructor works with Lana very diligently. He recognizes that Lana's prior preparation is indeed poor, and that her performance on this first exam is likely going to be far from the top of the class. Still, the instructor is confident that, given the quality of their office hour interactions, Lana would at least pass the first exam and perhaps gain a few study skills that will help her through the rest of the course. The instructor notes Lana's score while entering scores on the first midterm—39%. Failing. He is very puzzled by this and pledges to help Lana improve on the next exam. Together, they discuss study strategies and work practice problems in subsequent office hours. By the time the second exam is to be given, the instructor is quite confident that Lana will improve substantially. She is displaying a relatively solid ability to solve problems without his help as well as explaining the chemical concepts at play in each of the problems. He, again, knows that she isn't going to achieve the top score in the class, but another failing grade is simply out of the question. Lana's second exam grade: 29%. The instructor is now convinced that Lana suffers from an issue unrelated to chemistry, perhaps testing anxiety. He recommends that she seek help from academic services, advice that Lana accepts without protest. Lana and the instructor continue their work during office hours—discussing problem solving strategies, exam-taking strategies, and working to help Lana achieve a more solid conceptual understanding of the material. Lana's third exam score: 47%. Lana visits the instructor only once between the third exam and the final exam. During this time the instructor, dejected, admits that he is out of ideas. Lana assures him that she appreciates everything he had tried and that she enjoyed their time together—perhaps chemistry just isn't her "thing". Lana and the instructor part ways amicably; she scores a 26% on the final exam and receives an "F" in the course.

During the first week of the following semester, Lana stops by to visit her former chemistry instructor. She relays that, following her rather disastrous first semester (her troubles were not isolated to chemistry), she made an appointment with academic services to be tested for a learning disability. To her (and perhaps the instructor's) surprise, Lana didn't have a learning disability, per se. *Lana reads at a ninth-grade level.*

Is Lana an isolated case?

With a keen interest in examining the reading comprehension abilities of our students, we administered a measure of reading comprehension to students at the beginning of the accelerated general chemistry course for engineering majors. The Gates-MacGinitie Reading Test (GMRT) was used (MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000a). The GMRT consists 48 multiple-choice questions designed to assess student comprehension on several short text passages. These passages are taken from published books and periodicals; the content was selected by the test's creators to reflect the types of material that students are required to read as academic work and choose to read for recreation. GMRT scores can be compared to national norms to establish percentile rankings and grade equivalents (MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000b)—useful information when attempting to survey the relative comprehension abilities of students in a particular class. Figure 1 provides the distribution of comprehension abilities (as national percentile rankings) for 550 students enrolled in this course over four semesters. Indeed, the comprehension abilities of students in this course were quite variable, ranging from the 1st to the 99th percentile of graduating high school seniors. The mean GMRT score of this population (32.5/48) placed the average student between the 56th and 59th percentile; that is, on average, students in the course scored higher than 56-59% of graduating high school seniors. This mean score equated to a post-high school grade level, as expected of students enrolled in a college-level chemistry course. However, approximately 22.2% of students scored below the 35th percentile of graduating high school seniors. This percentile equated to comprehension abilities below the national average score for 9th graders. "Lana" is certainly not alone. She is one of many students whose comprehension skills lag well below the level expected of a college engineering student.

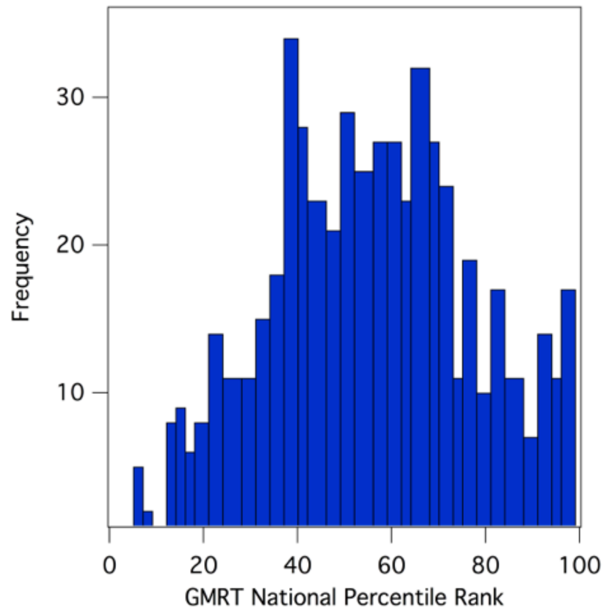


Figure **1**. Distribution of comprehension abilities for students in the accelerated general chemistry course for engineering majors.

Language and Learning Chemistry

A recent study of ours investigated the quantitative relationship between comprehension ability and performance in general chemistry courses (Pyburn, Pazicni, Benassi, & Tappin, 2013). Our work examined this relationship in multiple courses, using multiple measures of comprehension ability and multiple measures of course performance. We concluded that a robust and generalizable relationship exists between comprehension ability and course performance. Granted, like many courses, performance in general chemistry is dependent on a range of factors (Tai, Sadler, & Loehr, 2005). But why should general chemistry performance be dependent on comprehension ability? And why should instructors care about this relationship?

For over thirty years, chemistry education researchers and practitioners alike have documented issues associated with language and learning chemistry. Formative work by Cassels and Johnstone (1980) with pre-college students demonstrated non-technical words associated with chemistry were often a cause of alternative conceptions. At the college level, Jasien and Oberem (2008) and Jasien (2010, 2011) documented students' confusion regarding the terms *dense*, *energy*, *neutral* and *strong*, while Hamori and Muldrey (1984) mused on the ambiguity of the term *spontaneous* in the context of teaching thermodynamics. Cassels and Johnstone (1985) recognized that language is a contributor to information overload and, consequently, limits a student's ability to solve problems. Potential issues beyond lexical ambiguity include unfamiliar/misleading vocabulary, use of "high-sounding" language, and the use of multiple negatives. Many of these language issues are further amplified in students for whom English is a foreign language (Childs & O'Farrell, 2003; Johnstone & Selepeng, 2001).

Gabel (1999) echoed much of the aforementioned work by noting that the difficulties students have with chemistry might not be related to the subject matter itself but to how chemistry knowledge is linguistically expressed. Middlecamp and Kean (1988) lamented it "is not uncommon for 20 or more concepts to be introduced in the course of a one-hour lecture. Small wonder that some students perceive chemistry to be a language course!" (p. 54). Ver Beek and Louters (1991) concluded from their study, in which they divorced chemical language from common language and math ability, that "the difficulties experienced by our beginning college chemistry students appear to be largely precipitated by a lack of chemical language skill rather than by a lack of native reasoning and/or mathematical skills" (p. 391). Indeed, language is fundamental to the teaching and learning of chemistry, as it is to all sciences. The relationship is so profound that Bulman (1985, pp 1) proposed any school science department should work towards developing and improving language skills in its students. Toward this end, Herron (1996, pp. 161-182) devoted an entire chapter in his book *The Chemistry Classroom* to discussing this topic.

The Hallmarks of Low-skilled Comprehenders

In attempting to aid a disadvantaged subgroup of students, designing potential strategies to help low-skilled comprehenders is surprisingly straightforward. The theoretical basis for comprehension presents clear guidelines for the development of pedagogical strategies. There are several prominent models of comprehension ability, as reviewed by McNamara and Magliano (2009). We have found two models especially useful for identifying the hallmarks of low-skilled comprehenders: *Construction-Integration* and *Structure Building*.

According to *Construction-Integration*, when a text is read, the reader builds a mental representation comprised of three levels: the *surface structure*, the *propositional textbase*, and the *situation model* (Kintsch, 1988, 1998; van Dijk & Kintsh, 1983). The *surface structure* encompasses the words in the text and their grammatical relations, but is generally assumed to have little effect on comprehension (McNamara & Magliano, 2009, p. 309). The *propositional textbase* is the superficial meaning of the text. The *situational model* is the deepest level of representation and is constructed by combining prior knowledge with information in the textbase, i.e. by making inferences. McNamara and Magliano (2009, p. 366) argued that that the most critical difference between low-skilled and high-skilled comprehenders is the ability and tendency to generate these inferences. Beyond this model, which considers elements associated with the text only, researchers have begun to consider how readers' behavior affects comprehension. Chief among these extratextual dimensions is reading strategy use. Deep comprehension of content is presumed to emerge from strategies that prompt the learner to generate inferences connecting what is being learned to prior knowledge. These strategies include asking questions, answering questions, evaluating the quality of answers to questions, generating explanations, solving problems, and reflecting on the success of such strategies (Bransford, Brown, & Cocking, 2000, p. 67; Graesser, McNamara, & Van Lehn, 2005; McNamara, 2010). In other words, skilled comprehenders are better able to strategically use prior knowledge to fill in the conceptual gaps encountered when attempting to comprehend new information. Low-skilled comprehenders often fail to employ strategies necessary to improve comprehension at a deeper level (Lenski & Nierstheimer, 2002).

Structure Building describes how new information, both linguistic and non-linguistic, is incorporated into one's existing knowledge base (see also Nguyen & McDaniel, this volume; Gernsbacher, 1990). Briefly, Structure Building describes an individual's prior knowledge as having produced a *foundation* to which new information can be linked. As new information relevant to this foundation is encountered, a new *substructure* is built—a connection between new information and relevant prior knowledge. If new information irrelevant to an active substructure is encountered, the corresponding prior knowledge base is suppressed as another prior knowledge base is activated, allowing a new substructure to be built. As one substructure is suppressed in favor of building a new one, the information encoded in the suppressed substructure becomes less accessible. Students with low comprehension ability are at a distinct disadvantage as they have inefficient suppression mechanisms and are unable to inhibit irrelevant information (Gernsbacher & Faust, 1991). As a result, poor comprehenders attempt to accommodate information that will not fit into existing substructures and regularly *shift* to build new substructures. These students readily lose access to recently encoded information and build less coherent structures of new information.

Designing a Classroom Intervention for Low-skilled Comprehenders

It is well recognized that reading strategy instruction improves comprehension and is one of the most effective means of helping students to overcome deficits in comprehension ability (e.g. Bereiter & Bird, 1985; King & Rosenshine, 1993; McNamara, 2007; Ozgungor & Guthrie, 2004; Palinscar & Brown, 1984). The main goal of most strategy instruction is for less-skilled students to learn strategies that mimic those exhibited by skilled students or that compensate for processes exhibited by skilled students (McNamara, 2009). Consistent with the models described above, direct strategy instruction has been shown to be particularly effective for students with low prior knowledge or low comprehension ability (McNamara, 2007; O'Reilly & McNamara, 2007). However, it is rather unreasonable to expect university instructors to provide deliberate strategy instruction. Instructors in STEM disciplines, for example, are typically neither trained in strategy instruction, nor able to sacrifice classroom time for such endeavors, their success notwithstanding. As an alternative, however, we conjecture that purposefully scaffolding the use of strategies onto learning activities may be beneficial. Low-skilled comprehenders tend not to employ the range and frequency of strategy use exhibited by high-skilled comprehenders; thus, domain-centered learning activities that promote successful reading strategies should help to close the achievement gap between low- and high-skilled comprehenders. Designing instructional interventions that demand students make use of a strategy or strategies (i.e. mimic the behavior of skilled comprehenders) may the most efficient way an instructor can aid low-skilled comprehenders.

In designing our intervention, we exploited the fact that high-skilled comprehenders tend to answer questions and problem-solve—a strategy believed to help inference making, connecting what is being learned to prior knowledge. Consequently, our intervention took the form of "pre-testing." Traditionally, testing is used to only *assess* student learning. However, a body of literature supports the idea that testing can be used to *facilitate* learning (see Pyc, Agarwal, & Roediger, this volume, for a review; Bartlett, 1977; Darley & Murdoch, 1971; Hanawalt & Tarr, 1961; Hogan & Kintsch, 1971; Masson & McDaniel, 1981; McDaniel, Kowitz, & Dunay, 1989; McDaniel & Masson, 1985; McDaniel, Anderson, Derbish, & Morrisette, 2007; Whitten & Bjork, 1977). This phenomenon of test-enhanced learning has been termed the *testing effect*. It is hypothesized that recalling information while taking a quiz or exam promotes long-term retention of that information, and that this has a greater benefit than repeated studying (Pyc et al., this volume; Callender & McDaniel, 2009; McDaniel et al., 2007). A handful of recent studies (e.g. McDaniel et al., 2007; Roediger, Agarwal, McDaniel, & McDermott, 2011; Rohrer, Taylor, & Sholar, 2010) has demonstrated the benefits of the testing effect in real classroom settings.

We also considered what types of questions would be appropriate for our pre-testing intervention. An attractive option, from a theoretical standpoint, was elaborative interrogation (EI) questions. EI questions are simply "why" questions (e.g., Atomic radii decrease going from left to right on the periodic table. Why?). Studies have demonstrated that including EI questions with reading assignments results in positive effects on recalling information or making inferences on subsequent tests (Boudreau, Wood, Willoughby, & Specht, 1999; Callender & McDaniel, 2007; McDaniel & Donnelly, 1996; Ozungor & Guthrie, 2004; Pressley, McDaniel, Turnure, Wood, & Ahmed, 1987; Ramsay, 2010; Smith, Holliday, &

Austin, 2010). It is believed that EI questions activate prior knowledge and help the reader relate new information to existing knowledge (Woloshyn, Pressley, & Schneider, 1992; Woloshyn, Wood, & Willoughby, 1994). Using EI questions in our pre-testing intervention is thus consistent with our guiding models of comprehension, as EI questions are thought to aid students in building inferences, a cornerstone of *Construction-Integration*. In addition, because poor comprehenders experience difficulty in parsing out irrelevant information (resulting in a high amount of shifting, according to *Structure Building*), EI questions may prove beneficial because such questions anchor students to relevant prior knowledge.

In the context of being used in a real classroom, however, EI questions have one substantial drawback — time intensiveness. To motivate students to complete the multiple-testing intervention, the intervention should be “high-stakes”; that is, all of the testing associated with the intervention should count toward students’ overall course grades. The course instructional staff would need to assess student responses to all of the EI questions posed during the course of the intervention—no small feat, especially in a large-lecture general chemistry course. Thus, in addition to EI, we considered another question type that demanded far less time: multiple-choice (MC) questions. McDaniel, Agarwal, Huelser, McDermott, and Roediger (2011) have shown that repeated MC testing produces significant learning gains in a middle school science classroom, while Little, Bjork, Bjork, and Angello (2012) recently demonstrated that MC questions not only fostered retention of information, but also facilitated recall of information pertaining to incorrect alternatives. Thus, using MC questions to aid low-skilled comprehenders is also consistent with our guiding models of comprehension, as knowledge of both correct and incorrect responses to MC questions may help low-skilled comprehenders filter irrelevant information and build a more coherent structure of a concept. It addition, should MC questions elicit performance gains at least as large as EI questions, MC questions would be a much more “instructor-friendly” intervention for a large course.

Implementing and Evaluating the Multiple-testing Intervention

Implementation and evaluation of our multiple-testing intervention strategy took place over two semesters of the same accelerated general chemistry course. Data were gathered for 103 students and 127 students during these two semesters, respectively. We chose a within-students design for our assessment that allowed all students in the course to be treated equally. Our intervention was thus performed at the level of course content—a learning goal centered design (Krajcik, McNeill, & Reiser, 2008). Prior to the beginning of our two-semester study, the content of the accelerated general chemistry course was “unpacked” into five sets of learning goals, one set for each course exam, including the final exam. We randomly chose eighteen learning goals from each set to be tested on course exams; the same learning goals were tested during both semesters of the study. These sets of learning goals were also distributed to students as study aids. Students were informed that eighteen goals would be tested on each exam, but they did not know which goals specifically would be assessed. Exam items that assessed the mastery of learning goals were open-ended in nature and typically required calculations and/or explanations. These exams items required written problem solving and short answer essays graded with partial credit; no multiple-choice questions were given.

To test for the effects of repeated testing and question type on student performance (as measured by course exams), the eighteen learning goals chosen for each exam were randomly partitioned among three possible pre-exam testing treatments. Six of the learning goals were tested twice with MC questions, another six goals were tested twice with EI questions, and the remaining six goals were not pre-tested and therefore served as a control. If a learning goal received a pre-exam testing treatment, the first test was administered as a part of an online “reading check” that students completed (presumably) following the reading of an assigned section of text, while the second test was administered as part of a weekly content summary quiz. For pre-tested learning goals, different questions were used for each of the tests comprising the treatment. Examples of learning goals, pre-exam MC (Table 1) and EI (Table 2) questions keyed to those goals, and exam questions are provided below. Each type of test contained both MC and EI questions and feedback was provided for each pre-testing event. MC questions were automatically scored by the online system and students were immediately informed of their performance. EI questions were scored manually by the course teaching staff and feedback was provided one week following the pre-testing event, at most.

Table 1: An example of a learning goal that received a pre-exam treatment with MC questions

Learning Goal	Apply your knowledge regarding how energy can be dispersed in the motions of atoms and molecules to predict the sign of the *entropy change* for a chemical reaction.
MC Question 1	Which of the following processes has a $\Delta S > 0$? a. $CH_4(g) + H_2O(g) \rightarrow CO(g) + 3\ H_2(g)$ b. $CH_3OH(l) \rightarrow CH_3OH(s)$ c. $N_2(g) + 3\ H_2(g) \rightarrow 2\ NH_3(g)$ d. $Na_2CO_3(s) + H_2O(g) + CO_2(g) \rightarrow 2\ NaHCO_3(s)$
MC Question 2	For each of the following processes, indicate whether you expect the entropy change of the system to be positive or negative. a. butter melting b. carbon dioxide gas dissolving in soda c. water vaporizing when boiling d. water vapor in clouds precipitating as snow
Exam Question	Which reaction has a negative change in entropy? (Only <u>one</u> reaction will have a negative change in entropy.) Support your response with a discussion of how energy can be dispersed in the motions of matter and how this system's potential to disperse energy decreases as a result of the reaction. **Do <u>not</u> support your response with a calculation.**

Reaction	Balanced chemical equation	$\Delta H°$ (kJ)	$\Delta G°$ (kJ)
I	$KCl(s) \rightarrow K^+(aq) + Cl^-(aq)$	18.04	–4.25
II	$N_2(g) + 3\ H_2(g) \rightarrow 2\ NH_3(g)$	–161.7	53.4
III	$Mg(OH)_2(s) \rightarrow MgO(s) + H_2O(g)$	81.1	28.0

Table 2: An example of a learning goal that received a pre-exam treatment with EI questions.

Learning Goal	Use the periodic table and *aufbau rules* to predict the full and valence electron configurations of elements and ions to *Z* = 56 (including the aufbau "exceptions" Cr and Cu).
EI Question 1	When constructing an electron configuration, it is better to "place" electrons in separate orbitals (within the same subshell) before pairing them. Why?
EI Question 2	A friend of yours incorrectly believes that the electron configuration of Mn^{2+} is [Ar]$4s^2 3d^3$. Why is your friend mistaken?
Exam Question	Write the electron configuration for Sn. You may use the noble gas shorthand.

First, we confirmed that the testing effect holds true. On average, students scored better on exam items keyed to pre-tested learning goals (69.5%) than on those exam items keyed to non-tested learning goals (66.7%). This testing effect differentially helped low-skilled comprehenders (Figure 2). We split the student sample into thirds based on GMRT scores and designated the lower third to be low-skilled comprehenders. The upper third was designated to be high-skilled comprehenders. The 11.2% achievement gap between low- and high skilled comprehenders on control learning goals was abated somewhat by testing. The high-skilled comprehenders performed 7.9% better than low-skilled comprehenders on pre-tested learning goals. This testing effect was also amplified for MC questions in comparison to EI questions. On average, students scored 7.3% better on exam items keyed to learning goals that had been pre-tested with MC questions (74.0%) than on those exam items keyed to control learning goals (66.7%). Surprisingly, students scored 1.8% *lower* on exam items keyed to learning goals that had been pre-tested with EI questions than on those exam items keyed to control learning goals.

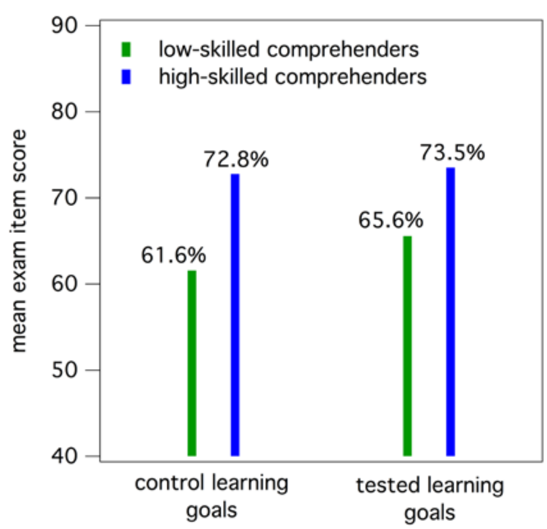


Figure **2.** Comparing the effect of testing on course performance for students of high and low comprehension.

Second, we demonstrated that the effect of MC testing was more pronounced for low-skilled comprehenders (again defined as those possessing GMRT scores in the lowest third of the sample) than high-skilled comprehenders. Figure 3 compares means of exam item performance for low- and high-skilled comprehenders over each of the testing treatments (control vs. EI vs. MC). While the achievement gap between low- and high-skilled comprehenders (illustrated by the control condition in Figure 3) was not completely abolished by repeated MC pre-testing, the gap was significantly closed. Low-skilled comprehenders scored 7.9% better than control when learning goals were pre-tested with MC questions prior to examination; high-skilled comprehenders scored 5.2% better than control. Thus, the achievement gap between these two groups was closed by ~3% when learning goals were pre-tested with MC questions prior to examination. Pre-testing learning goals with EI questions prior to examination did not improve performance for either group; as before, mean exam item scores for both groups were lower than control. Thus, our results suggest that repeated testing with MC questions is a strong instructional strategy for intervening on behalf of low-skilled comprehenders.

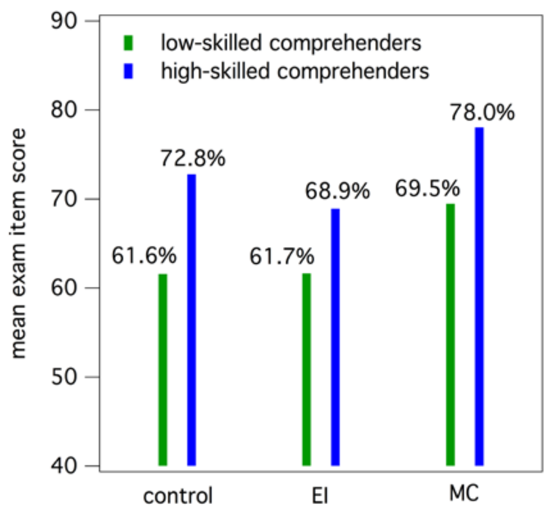


Figure 3. Comparing the effects of EI and MC testing on course performance versus control for low- and high-skilled comprehenders.

We were initially puzzled by the results for the EI testing condition. As mentioned above, a wealth of literature exists venerating the effects of EI questions on low-skilled comprehenders. However, in our research scenario, EI questions had a small negative effect. One possible explanation is that MC questions, by virtue of being scored automatically by the online course system, provided immediate performance feedback to students. For EI questions, feedback was delayed by up to a week; students may have not even bothered to review this feedback after it was provided. This speaks to the relative ease of employing MC testing in large classrooms in comparison to EI questions. EI questions may have elicited performance gains if students were provided immediate feedback; such a scenario, however, was not possible in the context of the course.

The overall achievement gain for low-skilled comprehenders due to repeated testing with MC questions when compared to the control was substantial (~8%); this gain could be on the order of a letter grade, depending on how instructors choose to build this construct. This gain estimated for low-skilled comprehenders was nearly 33% higher than that estimated for high-skilled comprehenders. Thus, our multiple pre-testing intervention using MC questions on behalf of low-skilled comprehenders was a success when both effectiveness and equity are considered. Not only was this intervention associated with improved performance for all involved, but also it differentially aided low-skilled comprehenders, promoting equity in our general chemistry classroom.

Summary and Recommendations

We were able to use the theoretical underpinnings of comprehension and, in general, the science of learning, to build a successful intervention strategy for low-skilled comprehenders in a general chemistry course. We designed a multiple pre-testing intervention strategy in which "instructor-friendly" multiple-choice questions appear to work best. Repeated testing with multiple-choice questions not only aided all students, but also *differentially* helped low-skilled comprehenders. It is likely that low-skilled comprehenders will be a disadvantaged population in any classroom. Thus, we consider our intervention strategy to be widely applicable, especially to other STEM courses where content can take the form of specific learning goals. Indeed, we believe a key to setting up this intervention strategy was specific instructor-generated learning goals around which multiple-choice quiz and examination questions could be built. It is enticing to speculate that these sets of learning goals may have themselves supported the low-skilled comprehenders, as a means for filtering irrelevant information.

We have two general recommendations for those instructors who may wish intervene on behalf of a disadvantaged population in their course. First, identify the particulars of a disadvantaged population with an instrument that possesses a solid basis in a theory of learning or cognition. No matter how an instrument discriminates between students or how well student data fits a particular trend, if the instrument is not backed by theory suggesting an intervening pedagogical approach, instructors can do little to help students. For example, many have touted that mathematics ability must be part of any instrument intending to assess potential general chemistry performance (e.g. Wagner, Sasser, & DiBiase, 2002). However, student performance on mathematics diagnostics is mostly likely related to prior instruction in mathematics, and not to any theory of learning. Thus, knowledge of a student's "math ability" isn't likely to be of much use to a chemistry instructor, as this construct provides little information concerning potential instructional strategies to aid those with poor math ability. Conversely, we demonstrated here that identifying disadvantaged students using an instrument that *is* tied to theory could provide powerful insight into helping those students.

Second, we recommend that the evaluation and implementation of the intervention be intentionally linked. Evaluating the effectiveness of an intervention is a crucial aspect of developing the intervention. One should consider the evaluation of an intervention in concert with its implementation (i.e. implement the intervention in such a way as to enable the collection of well-regulated and easily-analyzed data). This may mean that during the implementation/evaluation phase, the intervention (or the course!) does not take on precisely the same format as when it was conceived. For example, our multiple-testing intervention currently uses solely multiple-choice questions that are keyed to every learning goal for the course. The control set of learning goals and question of whether MC or EI would better suit the intervention were aspects of evaluation. We have also moved away from the rather rigid exam structure used in the evaluation. After all, there will be plenty of opportunities to adjust and hone our pre-testing intervention, now that its main premise has proven effective

Acknowledgement

This work is supported in part by a grant from the Davis Educational Foundation. The Foundation was established by Stanton and Elisabeth Davis after Mr. Davis's retirement as chairman of Shaw's Supermarkets, Inc.

References

Bartlett, J. C. (1977). Effects of immediate testing on delayed retrieval: Search and recovery operations with four types of cue. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 719-732. doi: 10.1037/0278-7393.3.6.719.

Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction*, *2*, 131-156. doi: 10.1207/s1532690xci0202_2

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Boudreau, R. L., Wood, E., Willoughby, T., & Specht, J. (1999). Evaluating the efficacy of elaborative strategies for remembering expository text. *The Alberta Journal of Educational Research*, *45*, 170-183.

Bulman, L. (1985). *Teaching language and study skills in secondary science*; London: Heinemann Educational Book Ltd.

Callender, A. A., & McDaniel, M. A. (2007). The benefits of embedded question adjuncts for low and high structure builders. *Journal of Educational Psychology*, *99*, 339-348. doi: 10.1037/0022-0663.99.2.339.

Cassels, J. R. T., & Johnstone, A. (1980). Understanding of non-technical words in science. London: Chemical Society.

Cassels, J. R. T., & Johnstone, A. H. (1985). Words that matter in science. London: Royal Society of Chemistry.

Childs, P. E., & O'Farrell, J. O. (2003). Learning science through English: An investigation of the vocabulary skills of native and non-native English speakers in international schools. *Chemistry Education Research and Practice*, *4*, 233-247. doi: 10.1039/B3RP90015K.

Darley, C.F., & Murdock, B.B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, *91*, 66-73. doi: 10.1037/h0031836.

Gabel, D. (1999). Improving teaching and learning through chemistry education research: A look to the future. *Journal of Chemical Education*, *76*, 548-554. doi: 10.1021/ed076p548.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 245-262. doi: 10.1037/0278-7393.17.2.245.

Graesser, A. C., McNamara, D. S., & Van Lehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist*, *40*, 225-234. doi: 10.1207/s15326985ep4004_4.

Hamori, E., & Muldrey, J. E. (1984). Use of the word eager instead of spontaneous for the description of exergonic reactions. *Journal of Chemical Education*, *61*, 710. doi: 10.1021/ed061p710.

Hanawalt, N.G., & Tarr, A.G. (1961). The effect of recall upon recognition. *Journal of Experimental Psychology*, *62*, 361-367. doi: 10.1037/h0041917.

Herron, J. D. (1996). *The chemistry classroom*. Washington, DC: American Chemical Society.

Hogan, R.M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 562-567. doi: 10.1016/S0022-5371(71)80029-4.

Jasien, P. G. (2010). You said "neutral", but what do you mean? *Journal of Chemical Education*, *87*, 33-34. doi: 10.1021/ed8000182.

Jasien, P. G. (2011). What do you mean that "strong" doesn't mean "powerful"? *Journal of Chemical Education*, *88*, 1247-1249. doi: 10.1021/ed100744x.

Jasien, P. G., & Oberem, G. E. (2008). Student contextual understanding of the terms sense and energy: A qualitative study. *The Chemical Educator*, *13*, 46-53. doi: 10.1333/s00897082117a.

Johnstone, A. H., & Selepeng, D. (2001). A language problem revisited. *Chemistry Education Research and Practice*, *2*, 19-29. doi: 10.1039/B0RP90028A.

King, A., & Rosenshine, B. (1993). Effects of guided cooperative-questioning on children's knowledge construction. *Journal of Experimental Education*, *6*, 127-148. doi: 10.1080/00220973.1993.9943857.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163-182. doi: 10.1037/0033-295X.95.2.163.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Krajcik, J., McNeill, K. L., & Reiser, B. (2008). Learning goals-driven design model: Curriculum materials that align with national standards and incorporate project-based pedagogy. *Science Education*, *92*, 1-32. doi: 10.1002/sce.20240.

Lenski, S. D., & Nierstheimer, S. L. (2002). Strategy instruction from a sociocognitive perspective. *Reading Psychology*, *23*, 127-143. doi: 10.1080/027027102760351034.

Little, J.L., Bjork, E.L., Bjork, R.A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science*, *23*, 1337-1344. doi: 10.1177/0956797612443370.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2000a). *Gates MacGinitie Reading Test Fourth Edition Forms S and T – Paper-pencil*. Chicago: Riverside.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2000b). *Gates MacGinitie Reading Tests Fourth Edition Levels 7/9 & 10/12 Forms S&T: Manual for Scoring and Interpretation*. Chicago: Riverside.

Masson, M. E. J., & McDaniel, M. A. (1981). The role of organization processes in long-term retention. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 100-110. doi: 10.1037/0278-7393.7.2.100.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*, 399-414. doi: 10.1037/a0021782.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*, 494-513. doi: 10.1080/09541440701326154.

McDaniel, M. A., & Donnelly, C. M. (1996). Learning with analogy and elaborative interrogation. *Journal of Educational Psychology*, *88*, 508-519. doi: 10.1037/0022-0663.88.3.508.

McDaniel, M. A., Kowitz, M. D., & Dunay, P. K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory and Cognition*, *17*, 423-434. doi: 10.3758/BF03202614.

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 371-385. doi: 10.1037/0278-7393.11.2.371.

McNamara, D. S. (Ed.). (2007). *Reading comprehension strategies: Theory, interventions, and technologies*. Mahwah, NJ: Erlbaum.

McNamara, D. S. (2009). The importance of teaching reading strategies. *Perspectives on Language and Literacy*, *35*, 34-40.

McNamara, D. S. (2010). Computational methods to extract meaning from text and advance theories of human cognition. *Topics in Cognitive Science*, *2*, 1-15. doi: 10.1111/j.1756-8765.2010.01117.x.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. In Brian Ross (Ed.) *The psychology of learning and motivation* (pp. 298-384). New York: Academic Press. doi: 10.1016/S0079-7421(09)51009-2.

Middlecamp, C., & Kean, E. (1988). Problems and “That Other Stuff”: Types of chemical content. *Journal of Chemical Education*, *65*, 53-56. doi: 10.1021/ed065p53.

O’Reilly, T., McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional “high stakes” measures of high school students’ science achievement. *American Educational Research Journal*, *44*, 161-196. doi: 10.3102/0002831206298171.

Ozgungor, S., & Guthrie, J. T. (2004). Interactions among elaborative interrogation, knowledge, and interest in the process of constructing knowledge from text. *Journal of Educational Psychology*, *96*, 437-443. doi: 10.1037/0022-0663.96.3.437.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering ad comprehension monitoring activities. *Cognition and Instruction*, *1*, 117-175. doi: 10.1207/s1532690xci0102_1.

Pressley, M., McDaniel, M. A., Turnure, J. E., Wood, E., & Ahmed, M. (1987). Generation and precision of elaboration: Effects on intentional and incidental learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 291-300. doi: 10.1037/0278-7393.13.2.291.

Pyburn, D. T., Pazicni, S., Benassi, V. A., & Reilly, E. (2013). Assessing the relation between language comprehension and performance in general chemistry. *Chemistry Education Research and Practice*, doi: 10.1039/C3RP00014A.

Ramsay, C. M., Sperling, R. A., & Dornisch, M. M. (2010). A comparison of the effects of students’ expository text comprehension strategies. *Instructional Science*, *38*, 551-570. doi: 10.1007/s11251-008-9081-6.

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. Journal of Experimental Psychology: Applied, 17, 382-395. doi: 10.1037/a0026252.

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 233-239. doi: 10.1037/a0017678.

Smith, B. L., Holliday, W. G., & Austin, H. W. (2010). Students' comprehension of science textbooks using a question-based reading strategy. *Journal of Research in Science Teaching*, *47*, 363-379. doi: 10.1002/tea.20378.

Tai, R. H., Sadler, P. H., & Loehr, J. F. (2005). Factors influencing success in introductory college chemistry. *Journal of Research in Science Teaching*, *42*, 987-1012. doi: 10.1002/tea.20082.

Van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

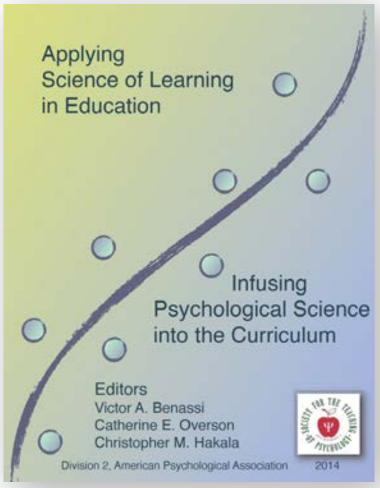
Ver Beek, K., & Louters, L. (1991). Chemical language skills: investigating the deficit. *Journal of Chemical Eudcation*, *68*, 389-392. doi: 10.1021/ed068p389.

Wagner, E. P., Sasser, H., & DiBiase, W. J. (2002). Predicting students at risk in general chemistry using pre-semester assessments and demographic information. *Journal of Chemical Education*, 79, 749-755.

Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 465-478.

Woloshyn, V. E., Pressley, M., and Schneider, W. (1992). Elaborative interrogation and prior knowledge effects on learning of facts. *Journal of Experimental Psychology*, *84*, 115-124. doi: 10.1037//0022-0663.84.1.115.

Woloshyn, V. E., Wood, E., & Willoughby, T. (1994). Considering prior knowledge when using elaborative interrogation. *Applied Cognitive Psychology*, *8*, 25-36. doi: 10.1002/acp.2350080104.



The Influence of Guiding Questions on Skilled- and Less-Skilled Readers' Understanding of Written Discourse

Jennifer J. Stiegler-Balfour
University of New England

Victor A. Benassi
University of New Hampshire

Heather Tatsak and Andrea Taatjes
University of New England

Although individualized learner-tailored instruction has long been viewed as an ideal scenario, such an objective remains elusive and reaching students of all skill levels equally remains a challenge for many educators. Most introductory-level classes can be taught under the assumption that all learners are novices; however, students might still have different academic skill levels, which can pose additional challenges. For example, a person's general reading comprehension skill can have a significant influence on their ability to acquire new knowledge (Hamilton, 1986). Studies by the National Assessment of Educational Progress (NAEP, 2009) indicate that the inability to comprehend text is one of the primary reasons students perform poorly in school. The goal of the current study was to explore how educators can increase the comprehension level of course material for less-skilled readers without slowing down class pace or hindering highly skilled readers' learning.

Research has shown that background knowledge in a certain subject plays a significant role in how students acquire new information. For example, according to the expertise reversal effect (see Lee & Kalyuga, this volume, for review), instructional design should be based on learners' background knowledge and skills. To construct a coherent representation of relevant material, novice learners depend heavily on guidance from instructors because they lack existing prior knowledge structures (also called schemas) in which to connect the new information. Because novice learners often lack background knowledge relationships between concepts, educators must spell out the information explicitly to ensure their academic success. That said, individuals with strong background knowledge on a topic benefit from reduced guidance as they have sufficient knowledge to construct their own coherent representation of the material. Interference on the part of the instructor with this process can sometimes hinder the learning process. Because many educators must teach students of varying academic skill levels and background knowledge of a subject, it is crucial to identify ways in which to provide extra resources for struggling readers while not holding back skilled students.

In addition to the importance of having background knowledge in a certain domain, research has shown that having poor reading comprehension skill can also contribute to poor academic performance (Gernsbacher, Varner, & Faust, 1990). According to research conducted by the National Center for Educational Statistics (NCES, 2012), 23% of U.S. adults fail to meet basic reading proficiency levels. In addition, many students who struggle with reading comprehension also tend to read less, which creates additional disadvantages as they also tend to lack sufficient background knowledge. Reading deficits can be linked to differences in basic cognitive skills. There are a variety of factors that influence reading skill, such as working memory capacity (Miyake, Just, & Carpenter, 1994), domain knowledge (McNamara & O'Reilly, 2002), use of effective reading strategies (Magliano, Trabasso, & Graesser, 1999), and the ability to differentiate between relevant and irrelevant information (Conway & Engle, 1994). Yet, to date, it remains unknown as to which of these factors carries the greatest significance in determining comprehension levels.

Several studies have addressed the cognitive differences that differentiate skilled- from less-skilled readers. A study conducted by Long and Chong (2001), for example, examined whether skilled and less-skilled readers differed in their ability to activate and integrate relevant text information. The results showed that skilled- and less-skilled readers were able to activate the relevant text information; however, only skilled readers were able to integrate the information. A plausible explanation for this finding is that less-skilled readers have difficulty differentiating between relevant and irrelevant information, and attempt to remember as much information as possible—regardless of relevance—thereby quickly overloading their working memory capacity. As a result, less-skilled readers' comprehension levels suffer because they are unable to comprehend the overall message of the text.

The current research expanded on previous laboratory research by applying knowledge gained from previous studies to the classroom. Our goal was to investigate whether providing readers with guiding questions, which they had to answer as they were completing the course readings, would help them distinguish between important and less important information, thereby allowing them to focus mostly on the relevant parts of the text. We hypothesized that focusing on the most relevant aspects of the text would reduce memory load and thus help to enhance less-skilled readers' understanding of the most important aspects of the text.

Participants consisted of 42 undergraduate students from the University of New England who were enrolled in the upper-level psychology of consciousness course. Fifty-five percent of students were seniors, 43 percent were juniors, and 2 percent were sophomores. Of the 42 students who participated in the study, 79 percent were members of the Psychology Department (55 percent Psychology, 14 percent Animal Behavior, and percent Neuroscience) and 21 percent were members of other departments at the University of New England (e.g., Medical Biology, Nursing, Occupational Studies, and Education). Because the class enrollment for this course is small, the data was collected over the course of two semesters in order to increase the sample size. The data of three students was not included in any of the analyses reported because they either withdrew from the course during the semester and did not complete all aspects of the study or scored in the middle range (47^{th} to 68^{th} percentile) on the Gates MacGinitie Reading Test (MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000).

At the beginning of each semester, each student was required to complete the reading comprehension section of the Gates MacGinitie Reading Test (MacGinitie et al., 2000) as a measure for assessing their reading comprehension skill. Similar to previous research, students who scored at or below the 46^{th} percentile were designated less-skilled readers, and those who scored at or above the 69^{th} percentile were designated as skilled readers (Ozuru, Dempsey, & McNamara, 2009). Less-skilled readers' received an average score of 28.44 out of 48 (*SD* = 4.34) on the test while skilled readers' average score was a 41.05 out of 48 (*SD* = 3.87). Of the 39 students included in the analysis, 18 students fell into the category of less-skilled readers and 21 students were considered skilled readers. Because scores on the Gates-MacGinitie Reading Comprehension Test are related to SAT Verbal scores, we also assessed students' verbal scores on the SAT (Williams & Takaku, 2011). Less-skilled readers had an average score of 511 (*SD* =65.48) and skilled readers had an average score of 584 (*SD* = 87.26) on the verbal portion of the SATs. The current study did not assess students' background knowledge of topics covered in this course but assumed students had roughly the same background knowledge as all students had completed an Introductory Psychology and Statistics course with a passing grade. Each semester, half of the skilled and less-skilled readers were then randomly assigned to groups A or B, and instructed to complete a guiding questions assignment alternating each week between groups A and B. Although students only had to submit the questions every other week, they were required to complete all the readings for every class meeting.

The guiding questions consisted of five to 10 questions that students answered based on the assigned textbook readings and journal articles. Students were asked to read the guiding questions prior to reading the assigned course materials and to complete each question once they had read the appropriate section in the text. The guiding questions consisted of questions developed to assist students with understanding the "bigger picture" and processing the material at a deeper level (Hamaker, 1986). For example, the questions encouraged students to not only summarize the text but also think about how different sections in the text were related (e.g., "What potential problems are there with the idea that consciousness causes our actions?" or "What kinds of evidence suggest a dissociation between vasomotor control and visual perception?"). All guiding questions were written to emphasize the most relevant aspects of the text as it pertained to the class. Responses to the guiding questions had to be submitted via the course management system (*Blackboard*©) prior to each class meeting and were evaluated for correctness (0-5 scale, 5 being the best). Students received instructor feedback on their responses.

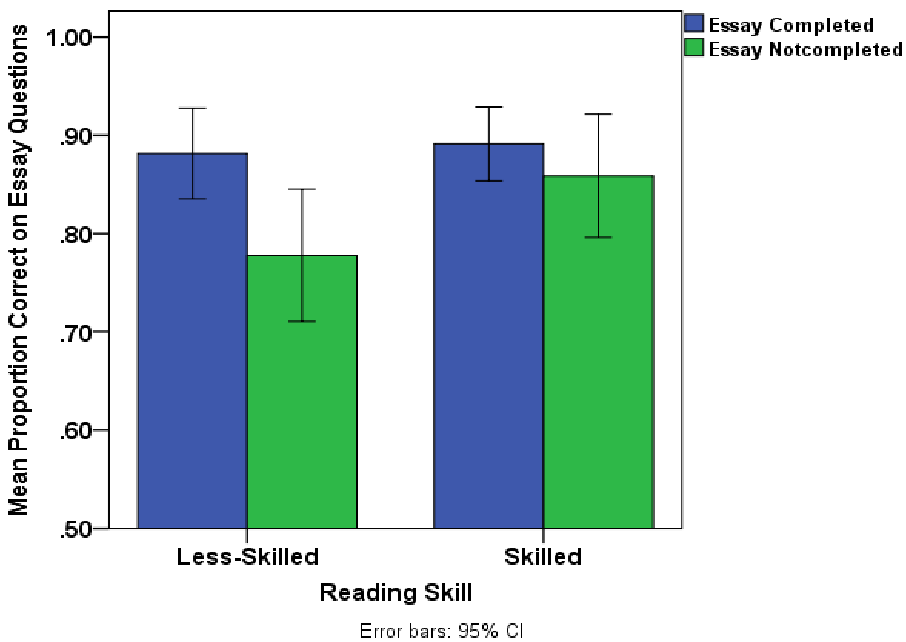
In addition to collecting data from the guiding questions responses, data were also obtained from the essay sections of the three examinations given throughout the semester. The three exams typically consisted of 14 to 25 multiple choice questions and five to eight essay questions. For example, essay questions included the following: "How does non-conscious biases guide behavior in a gambling situation before conscious knowledge does?," "Based on the readings, is it absolutely necessary to pay attention to something in order to be conscious of it? Explain your reasoning and be sure to mention top-down versus. bottom-up processing, selective attention, and how these findings relate to the real world." Last, students completed a 17-item survey at the end of the semester which assessed their perception of the effectiveness of the guiding questions, the level of effort put into the assignments, and their perceived reading skill and reading habits. For the purpose of the current paper, we will only report the findings pertaining to the essay exam questions. The complete discussion of the results of the multiple choice exam questions and survey will be available in a later publication.

Students designated as less-skilled readers performed significantly better on exam questions for which they had completed the guiding questions versus questions for which they had not completed the guiding questions (see Table 1). Interestingly, skilled readers did not benefit from completing the guiding questions (see Chee & Kalyuga, this volume). It is likely that skilled readers naturally engage in the deep processing of material even when they are not prompted to do so by an assignment. As a result, they did not show a significant improvement when they completed the questions compared to when they did not. As mentioned earlier, data were collected for two offerings of the course and combined for analysis (the patterns of results for both semesters was the same; Table 1). Figure 1 shows the combined mean essay exam scores on the essay exam questions for the two semesters.

Table 1: Mean performance on essay portion of exams for skilled and less-skilled readers when they completed the guiding questions versus when they did not.

	Semester 1					Semester 2				
	Completed			Not Completed		Completed			Not Completed	
Skilled	*n*	*M*	*SD*	*M*	*SD*	*n*	*M*	*SD*	*M*	*SD*
Reader	8	.89	.09	.87	.07	13	.89	.08	.85	.17
Less-Skilled Reader	8	.87	.12	.76	.17	10	.89	.07	.80	.10

Figure 1. Mean performance on essay portion of exams for skilled and less-skilled readers when they completed the guiding questions versus when they did not.



Although students were only required to complete the guiding questions for assigned readings on an alternating basis, nothing prevented them from working on questions during their "off weeks." To eliminate the possibility that skilled readers performed equally well on essay exam questions for which they did not have to complete the questions because they completed the guiding questions even on "off weeks," any student who had responded on the survey stating that they frequently (a score of 3 or higher on a 1 to 5 scale, never to always, respectively) completed the guiding questions on "off weeks" was eliminated from the analysis. The mean ratings for the less-skilled and skilled readers were 2.63 and 2.25, respectively (non-significant difference, $p > .51$). When students were eliminated from the analysis – if they had a score of three of greater on the question – the pattern of results remained the same as in the table.

Skilled readers were significantly more likely than less-skilled readers to report the frequent use of reading strategies when they completed reading assignments for their classes even when not prompted to do so (*M* = 3.875 , *SD* = .641 and *M* = 2.875, *SD* = .641, 1 to 5 scale, never to always, respectively, significant difference, $p > .50$). Interestingly, less-skilled readers not only failed to routinely employ reading strategies but more than half rated themselves as highly skilled readers (4 or 5 on a 1 to 5 scale, very low to very high reading comprehension skill, respectively) indicating that they were unaware of their reading comprehension difficulties.

These findings demonstrate that less-skilled readers benefit from being provided with guiding questions prior to completing course readings. Providing the question may have enabled them to enhance the reading process and capture important information from the assigned readings. As a result, their working memory was not overloaded with irrelevant information – providing students with the opportunity to focus on the most important aspects of the text, which in turn increased their understanding of the course material.

These results are in line with previous findings which showed that less-skilled readers require more guidance as to what information is and is not relevant so that they can suppress information that is currently not needed (McNamara & McDaniel, 2004). The apparent difference in effectiveness of the guiding questions for readers of different comprehension skill is not without precedent, as previous findings have suggested that skilled readers automatically elaborate on the text to fill in the gaps and/or suppress irrelevant text information to allow for more efficient processing of the most relevant information (McNamara & Kintsch, 1996). The results of the current study suggest that providing less-skilled readers with guiding questions for course readings significantly enhanced their understanding of the text and thus increase the likelihood that they will better learn and retain course information.

Teaching students of various skill levels can be difficult as instructors must balance supporting struggling students while also challenging skilled students' in order to maintain their interest. Based on the present findings, providing students with guiding questions for their readings would be one method instructors can utilize in their classes to increase the comprehension levels for course material while at the same time helping students to focus their attention on the most relevant aspects of the course readings. Because the guiding questions were completed outside of the classroom they did not consume class time and thus did not slow down the class pace. Because many of the less-skilled readers in this study were unaware of their inability to comprehend text at a deeper level, this intervention proved particularly useful because it can be applied to students of all skill levels without hindering their natural reading and learning strategies.

Based on the results of this study, the use of guiding questions would be equally appropriate for lower and upper-level courses, textbook readings and primary literature. Future research should assess whether implementing this type of intervention over time actually leads to an improvement of overall reading comprehension skills as it models what students should be doing as they are reading expository text.

Acknowledgement

This work is supported in part by a grant from the Davis Educational Foundation. The Foundation was established by Stanton and Elisabeth Davis after Mr. Davis's retirement as chairman of Shaw's Supermarkets, Inc.

References

Conway, A.A., & Engle, R.W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123*, 354-373. doi:10.1037/0096-3445.123.4.354

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 430-445.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, *56*, 212-242. doi:10.2307/1170376

Hamilton, R. J. (1986). Role of adjunct questions and subject ability levels on the learning of concepts from prose. *American Educational Research Journal*, *23*, 87-94. doi:10.2307/1163044

Long, D.L., & Chong, J.L. (2001). Comprehension skill and global coherence; A paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1424-1429

MacGinitie, W.H., MacGinitie, R.K., Maria, K., Dreyer, L.G., & Hughes, K.E. (2000). *Gates-MacGinitie reading tests.* Itasca, IL: Riverside Publishing Company.

Miyake, A., Just, M., & Carpenter, P.A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language*, 33, 175-202. doi:10.1006/jmla.1994.1009

McNamara, D. S., & McDaniel, M. A. (2004). Suppressing Irrelevant Information: Knowledge Activation or Inhibition?. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*,*30*, 465-482. doi:10.1037/0278-7393.30.2.465

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247-288. doi:10.1080/01638539609544975

McNamara, D.S., & O'Reilly, T. (2002). Learning: Knowledge acquisition, representation, and organization. J.W. Guthrie et al. (Eds.), *The Encyclopedia of education*. New York: MacMillan Reference.

Magliano, J.P., Trabasso, T., & Graesser, A.C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91*, 615-629. doi:10.1037/0022-0663.91.4.615

Ozuru, Y., Dempsey, K., & McNamara, D.S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction, 19*, 228-242.

U.S. Department of Education, National Center for Education Statistics. (2012). *Digest of Education Statistics, 2011* (NCES 2012-001)

Williams, J. D., & Takaku, S. (2011). Gender, Writing Self-Efficacy, and Help Seeking. *International Journal of Business, Humanities and Technology*, *1*, 46-54.