St Leonard’s College

Professional Readings

2018
## Contents

### STL Professional Readings 2018

<table>
<thead>
<tr>
<th>Section 1: Conceptual Understanding</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2 Learning from Mistakes - Geoff Masters (2017)</td>
<td>13</td>
</tr>
<tr>
<td>1.4 Improving Students’ Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology - Dunlosky, Rawson, Marsh, Nathan &amp; Willingham (2013)</td>
<td>20</td>
</tr>
<tr>
<td>1.5 Learning Styles: Concepts and Evidence - Pashler, McDaniel, Rohrer and Bjork (2009)</td>
<td>75</td>
</tr>
<tr>
<td>1.6 Progressive Mastery Through Deliberate Practice: A Promising Approach for Improving Writing - Nell Scharff Panero (2016)</td>
<td>91</td>
</tr>
<tr>
<td>1.7 Getting the Big Idea: Concept-Based Teaching and Learning – “Transforming Learning Environments through Global and STEM Education” (August 13, 2013)</td>
<td>108</td>
</tr>
<tr>
<td>1.9 Rethinking Instructional Delivery for Diverse Student Populations: Serving All Learners with Concept-Based Instruction – Jan D.McCoy and Leanne R. Ketterlin-Geller (2004)</td>
<td>119</td>
</tr>
<tr>
<td>1.10 Concept-Based Teaching and Learning, IB Position Paper – H Lynn Erickson (2012)</td>
<td>128</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 2: Agency</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 The Influence of Teaching. Beyond Standardized Test Scores: Engagement, Mindsets, and Agency. A Study of 16,000 Sixth through Ninth Grade Classrooms - A report from the Achievement Gap Initiative at Harvard University (2015). Note: There is a detailed report of 100 pages available but only the executive summary has been included here.</td>
<td>143</td>
</tr>
<tr>
<td>2.2 From Sage on the Stage to Guide on the Side - Alison King (1993)</td>
<td>162</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 3: Developing a Growth Mindset</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2 DNA is Not Destiny - Article from The Age (February 19, 2012)</td>
<td>180</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 4: Assessment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Using Assessments to Improve Teaching and Learning (2007) Thomas R. Guskey</td>
<td>189</td>
</tr>
<tr>
<td>4.2 Assessing Student Learning: Why Reform is Overdue - Geoff Masters, ACER (2011)</td>
<td>205</td>
</tr>
<tr>
<td>3.3 Towards a Growth Mindset in Assessment – Geoff Masters, ACER (2013)</td>
<td>210</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 5: Creativity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Artful Thinking – various articles</td>
<td>218</td>
</tr>
<tr>
<td>5.2 How to Promote Creativity in the Classroom – The Conversation Journal, Dan Davies (2015)</td>
<td>247</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 6: Teaching and Pedagogy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 The Role of the Teacher in the Thinking Curriculum</td>
<td>253</td>
</tr>
<tr>
<td>6.2 Let’s Get Serious About Teacher Quality: The Need For a New Career Architecture for Australia’s Teachers – Dean’s Lecture, Professor Stephen Dinham OAM (27 September 2011)</td>
<td>265</td>
</tr>
</tbody>
</table>
Section 1:

Conceptual Understanding
Critical Thinking

Why Is It So Hard to Teach?

By Daniel T. Willingham

Virtually everyone would agree that a primary, yet insufficiently met, goal of schooling is to enable students to think critically. In layperson's terms, critical thinking consists of seeing both sides of an issue, being open to new evidence that disconfirms your ideas, reasoning dispassionately, demanding that claims be backed by evidence, deducing and inferring conclusions from available facts, solving problems, and so forth. Then too, there are specific types of critical thinking that are characteristic of different subject matter: That's what we mean when we refer to "thinking like a scientist" or "thinking like a historian."

This proper and commonsensical goal has very often been translated into calls to teach "critical thinking skills" and "higher-order thinking skills"—and into generic calls for teaching students to make better judgments, reason more logically, and so forth. In a recent survey of human resource officials and in testimony delivered just a few months ago before the Senate Finance Committee, business leaders have repeatedly exhorted schools to do a better job of teaching students to think critically. And they are not alone. Organizations and initiatives involved in education reform, such as the National Center on Education and the Economy, the American Diploma Project, and the Aspen Institute, have pointed out the need for students to think and/or reason critically. The College Board recently revamped the SAT to better assess students' critical thinking. And ACT, Inc. offers a test of critical thinking for college students.

These calls are not new. In 1983, A Nation At Risk, a report by the National Commission on Excellence in Education, found that many 17-year-olds did not possess the "higher-order intellectual skills" this country needed. It claimed that nearly 40 percent could not draw inferences from written material and only one-fifth could write a persuasive essay.

Following the release of A Nation At Risk, programs designed to teach students to think critically across the curriculum became extremely popular. By 1990, most states had initiatives designed to encourage educators to teach critical thinking, and one of the most widely used programs, Tactics for Thinking, sold 70,000 teacher guides. But, for reasons I'll explain, the programs were not very effective—and today we still lament students' lack of critical thinking.

After more than 20 years of lamentation, exhortation, and little improvement, maybe it's time to ask a fundamental question: Can critical thinking actually be taught? Decades of cognitive research point to a disappointing answer: not really. People who have sought to teach critical thinking have assumed that it is a skill, like riding a bicycle, and that, like other skills, once you learn it, you can apply it in any situation. Research from cognitive science shows that thinking is not that sort of skill. The processes of thinking are intertwined with the content of thought (that is, domain knowledge). Thus, if you remind a student to "look at an issue from multiple perspectives" often enough, he will learn that he ought to do so, but if he doesn't know much about...
Critical thinking is not a set of skills that can be deployed at any time, in any context. It is a type of thought that even 3-year-olds can engage in—and even trained scientists can fail in.

In this article, I will describe the nature of critical thinking, explain why it is so hard to do and to teach, and explore how students acquire a specific type of critical thinking: thinking scientifically. Along the way, we'll see that critical thinking is not a set of skills that can be deployed at any time, in any context. It is a type of thought that even 3-year-olds can engage in—and even trained scientists can fail in. And it is very much dependent on domain knowledge and practice.

**Thinking Tends to Focus on a Problem's "Surface Structure"**

To understand why the surface structure of a problem is so distracting and, as a result, why it's so hard to apply familiar solutions to problems that appear new, let's first consider how you understand what's being asked when you are given a problem. Anything you hear or read is automatically interpreted in light of what you already know about similar subjects. For example, suppose you read these two sentences: “After years of pressure from the film and television industry, the President has filed a formal complaint with China over what U.S. firms say is copyright infringement. These firms assert that the Chinese government sets stringent trade restrictions for U.S. entertainment products, even as it turns a blind eye to Chinese companies that copy American movies and television shows and sell them on the black market.” Background knowledge not only allows you to comprehend the sentences, it also has a powerful effect as you continue to read because it narrows the interpretations of new text that you will entertain. For example, if you later read the word "Bush," it would not make you think of a small shrub, nor would you wonder whether it referred to the former President Bush, the rock band, or a term for rural hinterlands. If you read "piracy," you would not think of eye-patched swabbies shouting "shiver me timbers!" The cognitive system gambles that incoming information will be related to what you've just been thinking about. Thus, it significantly narrows the scope of possible interpretations of words, sentences, and ideas. The benefit is that comprehension proceeds faster and more smoothly; the cost is that the deep structure of a problem is harder to recognize.

The narrowing of ideas that occurs while you read (or
How Do Cognitive Scientists Define Critical Thinking?

From the cognitive scientist's point of view, the mental activities that are typically called critical thinking are actually a subset of three types of thinking: reasoning, making judgments and decisions, and problem solving. I say that critical thinking is a subset of these because we think in these ways all the time, but only sometimes in a critical way.Deciding to read this article, for example, is not critical thinking. But carefully weighing the evidence it presents in order to decide whether or not to believe what it says is. Critical reasoning, decision making, and problem solving—which, for brevity's sake, I will refer to as critical thinking—have three key features: effectiveness, novelty, and self-direction. Critical thinking is effective in that it avoids common pitfalls, such as seeing only one side of an issue, discounting new evidence that disconfirms your ideas, reasoning from passion rather than logic, failing to support statements with evidence, and so on. Critical thinking is novel in that you don't simply remember a solution or a situation that is similar enough to guide you. For example, solving a complex but familiar physics problem by applying a multi-step algorithm isn't critical thinking because you are really drawing on memory to solve the problem. But devising a new algorithm is critical thinking. Critical thinking is self-directed in that the thinker must be calling the shots: We wouldn't give a student much credit for critical thinking if the teacher were prompting each step he took.

—D.W.

listen) means that you tend to focus on the surface structure, rather than on the underlying structure of the problem. For example, in one experiment,1 subjects saw a problem like this one:

Members of the West High School Band were hard at work practicing for the annual Homecoming Parade. First they tried marching in rows of 12, but Andrew was left by himself to bring up the rear. Then the director told the band members to march in columns of eight, but Andrew was still left to march alone. Even when the band marched in rows of three, Andrew was left out. Finally, in exasperation, Andrew told the band director that they should march in rows of five in order to have all the rows filled. He was right. Given that there were at least 45 musicians on the field but fewer than 200 musicians, how many students were there in the West High School Band?

Earlier in the experiment, subjects had read four problems along with detailed explanations of how to solve each one, ostensibly to rate them for the clarity of the writing. One of the four problems concerned the number of vegetables to buy for a garden, and it relied on the same type of solution necessary for the band problem—calculation of the least common multiple. Yet, few subjects—just 19 percent—saw that the band problem was similar and that they could use the garden problem solution. Why?

When a student reads a word problem, her mind interprets the problem in light of her prior knowledge, as happened when you read the two sentences about copyrights and China. The difficulty is that the knowledge that seems relevant relates to the surface structure—in this problem, the reader dredges up knowledge about bands, high school, musicians, and so forth. The student is unlikely to read the problem and think of it in terms of its deep structure—using the least common multiple. The surface structure of the problem is overt, but the deep structure of the problem is not. Thus, people fail to use the first problem to help them solve the second: In their minds, the first was about vegetables in a garden and the second was about rows of band marchers.

With Deep Knowledge, Thinking Can Penetrate Beyond Surface Structure

If knowledge of how to solve a problem never transferred to problems with new surface structures, schooling would be inefficient or even futile—but of course, such transfer does occur. When and why is complex,2 but two factors are especially relevant for educators: familiarity with a problem's deep structure and the knowledge that one should look for a deep structure. I'll address each in turn.

When one is very familiar with a problem's deep-structure, knowledge about how to solve it transfers well. That familiarity can come from long-term, repeated experience with one problem, or with various manifestations of one type of problem (i.e., many problems that have different surface structures, but the same deep structure). After repeated exposure to either one both, the subject simply perceives the deep structure as part of the problem description. Here's an example:

A treasure hunter is going to explore a cave up on a hill near a beach. He suspected there might be many paths inside the cave so he was afraid he might get lost. Obviously, he did not have a map of the cave; all he had with him were some common items such as a flashlight and a bag. What could he do to make sure he did not get lost trying to get back out of the cave later?

The solution is to carry some sand with you in the bag, and leave a trail as you go, so you can trace your path back when you're ready to leave the cave. About 75 percent of American college students thought of this solution—but only 25 percent of Chinese students solved it.3 The experimenters suggested that Americans solved it because most grew up hearing the story of Hansel and Gre-

SUMMER 2007
tel, which includes the idea of leaving a trail as you travel to an unknown place in order to find your way back. The experimenters also gave subjects another puzzle based on a common Chinese folk tale, and the percentage of solvers from each culture reversed. (To read the puzzle based on the Chinese folk tale, and the tale itself, go to www.aft.org/pubs-reports/american_educator/index.htm.)

It takes a good deal of practice with problem type before students know it well enough to immediately recognize its deep structure, irrespective of the surface structure, as Americans did for the Hansel and Gretel problem. American subjects didn’t think of the problem in terms of sand, caves, and treasure; they thought of it in terms of finding something with which to leave a trail. The deep structure of the problem is so well represented in their memory, that they immediately saw that structure when they read the problem.

Looking for a Deep Structure Helps, but It Only Takes You So Far

Now let’s turn to the second factor that aids in transfer despite distracting differences in surface structure—knowing to look for a deep structure. Consider what would happen if I said to a student working on the band problem, “this one is similar to the garden problem.” The student would understand that the problems must share a deep structure and would try to figure out what it is. Students can do something similar without the hint. A student might think “I’m seeing this problem in a math class, so there must be a math formula that will solve this problem.” Then he could scan his memory (or textbook) for candidates, and see if one of them helps. This is an example of what psychologists call metacognition, or regulating one’s thoughts. In the introduction, I mentioned that you can teach students maxims about how they ought to think.

Critical Thinking Programs: Lots of Time, Modest Benefit

Since the ability to think critically is a primary goal of education, it’s no surprise that people have tried to develop programs that could directly teach students to think critically without immersing them in any particular academic content. But the evidence shows that such programs primarily improve students’ thinking with the sort of problems they practiced in the program—not with other types of problems. More generally, it’s doubtful that a program that effectively teaches students to think critically in a variety of situations will ever be developed.

As the main article explains, the ability to think critically depends on having adequate content knowledge; you can’t think critically about topics you know little about or solve problems that you don’t know well enough to recognize and execute the type of solutions they call for.

Nonetheless, these programs do help us better understand what can be taught, so they are worth reviewing briefly.

A large number of programs designed to make students better thinkers are available, and they have some features in common. They are premised on the idea that there is a set of critical thinking skills that can be applied and practiced across content domains. They are designed to supplement regular curricula, not to replace them, and so they are not tied to particular content areas such as language arts, science, or social studies. Many programs are intended to last about three years, with several hours of instruction (delivered in one or two lessons) per week. The programs vary in how they deliver this instruction and practice. Some use abstract problems such as finding patterns in meaningless figures (Reuven Feuerstein’s Instrumental Enrichment), some use mystery stories (Martin Covington’s Productive Thinking), some use group discussion of interesting problems that one might encounter in daily life (Edward de Bono’s Cognitive Research Trust, or CoRT), and so on. However it is implemented, each program introduces students to examples of critical thinking and then requires that the students practice such thinking themselves.

How well do these programs work? Many researchers have tried to answer that question, but their studies tend to have methodological problems.

Four limitations of these studies are especially typical, and they make any effects suspect:

1) students are evaluated just once after the program, so it’s not known whether any observed effects are enduring;
2) there is not a control group, leaving it unclear whether gains are due to the thinking program, to other aspects of schooling, or to experiences outside the classroom;
3) the control group does not have a comparison intervention, so any positive effects found may be due, for example, to the teacher’s enthusiasm for something new, not the program itself; and
4) there is no measure of whether or not students can transfer their new thinking ability to materials that differ from those used in the program. In addition, only a small fraction of the studies have undergone peer review (meaning that they have been impartially evaluated by independent experts). Peer review is crucial because it is known that researchers unconsciously bias the design and analysis of their research to favor the conclusions they hope to see.
Cognitive scientists refer to these maxims as metacognitive strategies. They are little chunks of knowledge—like "look for a problem's deep structure" or "consider both sides of an issue"—that students can learn and then use to steer their thoughts in more productive directions.

Helping students become better at regulating their thoughts was one of the goals of the critical thinking programs that were popular 20 years ago. As the sidebar below explains, these programs are not very effective. Their modest benefit is likely due to teaching students to effectively use metacognitive strategies. Students learn to avoid biases that most of us are prey to when we think, such as settling on the first conclusion that seems reasonable, only seeking evidence that confirms one's beliefs, ignoring countervailing evidence, overconfidence, and others. Thus, a student who has been encouraged many times to see both sides of an issue, for example, is probably more likely to spontaneously think "I should look at both sides of this issue" when working on a problem.

Unfortunately, metacognitive strategies can only take you so far. Although they suggest what you ought to do, they don’t provide the knowledge necessary to implement the strategy. For example, when experimenters told subjects working on the band problem that it was similar to the garden problem, more subjects solved the problem (35 percent compared to 19 percent without the hint), but most subjects, even when told what to do, weren't able to do it. Likewise, you may know that you ought not accept the first reasonable-sounding solution to a problem, but that doesn't mean you know how to come up with alternative solutions or weigh how reasonable each one is. That requires domain knowledge and practice in putting that knowledge to work.

Since critical thinking relies so heavily on domain

---

Studies of the Philosophy for Children program may be taken as typical. Two researchers identified eight studies that evaluated academic outcomes and met minimal research-design criteria. (Of these eight, only one had been subjected to peer review.) Still, they concluded that three of the eight had identifiable problems that clouded the researchers' conclusions. Among the remaining five studies, three measured reading ability, and one of these reported a significant gain. Three studies measured reasoning ability, and two reported significant gains. And, two studies took more impressionistic measures of student's participation in class (e.g., generating ideas, providing reasons), and both reported a positive effect.

Despite the difficulties and general lack of rigor in evaluation, most researchers reviewing the literature conclude that some critical thinking programs do have some positive effect. But these reviewers offer two important caveats. First, as with almost any educational endeavor, the success of the program depends on the skill of the teacher. Second, thinking programs look good when the outcome measure is quite similar to the material in the program. As one tests for transfer to more and more dissimilar material, the apparent effectiveness of the program rapidly drops.

Both the conclusion and the caveats make sense from the cognitive scientist's point of view. It is not surprising that the success of the program depends on the skill of the teacher. The developers of the programs cannot anticipate all of the ideas—right or wrong—that students will generate as they practice thinking critically, so it is up to the teacher to provide the all-important feedback to the students.

It is also reasonable that the programs should lead to gains in abilities that are measured with materials similar to those used in the program.

Knowing that one should think critically is not the same as being able to do so. That requires domain knowledge and practice.

The programs that include puzzles like those found on IQ tests, for instance, report gains in IQ scores. In an earlier column, I described a bedrock principle of memory: You remember what you think about. The same goes for critical thinking: You learn to think critically in the ways in which you practice thinking critically. If you practice logic puzzles with an effective teacher, you are likely to get better at solving logic puzzles. But substantial improvement requires a great deal of practice. Unfortunately, because critical thinking curricula include many different types of problems, students typically don't get enough practice with any one type of problem. As explained in the main article, the modest benefits that these programs seem to produce are likely due to teaching students metacognitive strategies—like "look at both sides of an issue"—that cue them to try to think critically. But knowing that one should think critically is not the same as being able to do so. That requires domain knowledge and practice.

---

(Endnotes on page 19)
Teaching students to think critically probably lies in large part in enabling them to deploy the right type of thinking at the right time.

knowledge, educators may wonder if thinking critically in a particular domain is easier to learn. The quick answer is yes, it's a little easier. To understand why, let's focus on one domain, science, and examine the development of scientific thinking.

Is Thinking Like a Scientist Easier?
Teaching science has been the focus of intensive study for decades, and the research can be usefully categorized into two strands. The first examines how children acquire scientific concepts; for example, how they come to forgo naive conceptions of motion and replace them with an understanding of physics. The second strand is what we would call thinking scientifically, that is, the mental procedures by which science is conducted: developing a model, deriving a hypothesis from the model, designing an experiment to test the hypothesis, gathering data from the experiment, interpreting the data in light of the model, and so forth.¹

Most researchers believe that scientific thinking is really a subset of reasoning that is not different in kind from other types of reasoning that children and adults do.² What makes it scientific thinking is knowing when to engage in such reasoning, and having accumulated enough relevant knowledge and spent enough time practicing to do so.

Recognizing when to engage in scientific reasoning is so important because the evidence shows that being able to reason is not enough; children and adults use and fail to use the proper reasoning processes on problems that seem similar. For example, consider a type of reasoning about cause and effect that is very important in science: conditional probabilities. If two things go together, it's possible that one causes the other. Suppose you start a new medicine and notice that you seem to be getting headaches more often than usual. You would infer that the medication influenced your chances of getting a headache. But it could also be that the medication increases your chances of getting a headache only in certain circumstances or conditions. In conditional probability, the relationship between two things (e.g., medication and headaches) is dependent on a third factor. For example, the medication might increase the probability of a headache only when you've had a cup of coffee. The relationship of the medication and headaches is conditional on the presence of coffee.

Understanding and using conditional probabilities is essential to scientific thinking because it is so important in reasoning about what causes what. But people's success in thinking this way depends on the particulars of how the question is presented. Studies show that adults sometimes use conditional probabilities successfully, but fail to do so with many problems that call for it.³ Even trained scientists are open to pitfalls in reasoning about conditional probabilities (as well as other types of reasoning). Physicians are known to discount or misinterpret new patient data that conflict with a diagnosis they have in mind,⁴ and Ph.D.-level scientists are prey to faulty reasoning when faced with a problem embedded in an unfamiliar context.⁵

And yet, young children are sometimes able to reason about conditional probabilities. In one experiment,⁶ the researchers showed 3-year-olds a box and told them it was a "blicket detector" that would play music if a blicket were placed on top. The child then saw one of the two sequences shown below in which blocks are placed on the blicket detector. At the end of the sequence, the child was asked whether each block was a blicket. In other words, the child was to use conditional reasoning to infer which block caused the music to play.

Note that the relationship between each individual block (yellow cube and blue cylinder) and the music is the same in sequences 1 and 2. In either sequence, the child sees the yellow cube associated with music three times, and the blue cylinder associated with the absence of music once and the presence of music twice. What differs between the first and second sequence is the relationship between the blue and yellow blocks, and therefore, the conditional probability of each block being a blicket. Three-year-olds understood the importance of conditional probabilities.

¹ These two strands are the most often studied, but these two approaches—content and process of science—are incomplete. Underemphasized in U.S. classrooms are the many methods of scientific study, and the role of theories and models in advancing scientific thought.

² Although this is not highly relevant for K-12 teachers, it is important to note that for people with extensive training, such as Ph.D.-level scientists, critical thinking does have some skill-like characteristics. In particular, they are better able to deploy critical reasoning with a wide variety of content, even that with which they are not very familiar. But, of course, this does not mean that they will never make mistakes.
"Teaching content alone is not likely to lead to proficiency in science, nor is engaging in inquiry experiences devoid of meaningful science content."

—National Research Council

For sequence 1, they said the yellow cube was a blicket, but the blue cylinder was not; for sequence 2, they chose equally between the two blocks.

This body of studies has been summarized simply: Children are not as dumb as you might think, and adults (even trained scientists) are not as smart as you might think.

What's going on? One issue is that the common conception of critical thinking or scientific thinking (or historical thinking) as a set of skills is not accurate. Critical thinking does not have certain characteristics normally associated with skills—in particular, being able to use that skill at any time. If I told you that I learned to read music, for example, you would expect, correctly, that I could use my new skill (i.e., read music) whenever I wanted. But critical thinking is very different. As we saw in the discussion of conditional probabilities, people can engage in some types of critical thinking without training, but even with extensive training, they will sometimes fail to think critically. This understanding that critical thinking is not a skill is vital. It tells us that teaching students to think critically probably lies in small part in showing them new ways of thinking, and in large part in enabling them to deploy the right type of thinking at the right time.

Returning to our focus on science, we're ready to address a key question: Can students be taught when to engage in scientific thinking? Sort of. It is easier than trying to teach general critical thinking, but not as easy as we would like. Recall that when we were discussing problem solving, we found that students can learn metacognitive strategies that help them look past the surface structure of a problem and identify its deep structure, thereby getting them a step closer to figuring out a solution. Essentially the same thing can happen with scientific thinking. Students can learn certain metacognitive strategies that will cue them to think scientifically. But, as with problem solving, the metacognitive strategies only tell the students what they should do—they do not provide the knowledge that students need to actually do it. The good news is that within a content area like science, students have more context cues to help them figure out which metacognitive strategy to use, and teachers have a clearer idea of what

domain knowledge they must teach to enable students to do what the strategy calls for.

For example, two researchers taught second-, third-, and fourth-graders the scientific concept behind controlling variables; that is, of keeping everything in two comparison conditions the same, except for the one variable that is the focus of investigation. The experimenters gave explicit instruction about this strategy for conducting experiments and then had students practice with a set of materials (e.g., springs) to answer a specific question (e.g., which of these factors determine how far a spring will stretch: length, coil diameter, wire diameter, or weight?). The experimenters found that students not only understood the concept of controlling variables, they were able to apply it seven months later with different materials and a different experimenter, although the older children showed more robust transfer than the younger children. In this case, the students recognized that they were designing an experiment and that cued them to recall the metacognitive strategy, "When I design experiments, I should try to control variables." Of course, succeeding in controlling all of the relevant variables is another matter—that depends on knowing which variables may matter and how they could vary.

**Why Scientific Thinking Depends on Scientific Knowledge**

Experts in teaching science recommend that scientific reasoning be taught in the context of rich subject matter knowledge. A committee of prominent science educators brought together by the National Research Council put it plainly: "Teaching content alone is not likely to lead to proficiency in science, nor is engaging in inquiry experiences devoid of meaningful science content."

The committee drew this conclusion based on evidence that background knowledge is necessary to engage in scientific thinking. For example, knowing that one needs a control group in an experiment is important. Like having two comparison conditions, having a control group in addition to an experimental group helps you focus on the variable you want to study. But knowing that you need a control group is not the same as being able to create one. Since it's not always possible to have two groups that are exactly alike, knowing which factors can vary between groups and which must not vary is one example of necessary background knowledge. In experiments measuring how quickly subjects can respond, for example, control groups must be matched for age, because age affects response speed, but they need not be perfectly matched for gender.

More formal experimental work verifies that background knowledge is necessary to reason scientifically. For example, consider devising a research hypothesis. One could generate multiple hypotheses for any given situation. Suppose you know that car A gets better gas mileage than car B and you'd like to know why. There are many differences between the cars, so which will you investigate first? Engine size? Tire pressure? A key determinant of the hypothesis you select is plausibility. You won't choose to investigate a difference between cars A and B that you think is unlikely to contribute to gas mileage (e.g., paint color), but if someone provides a reason to make this factor more plausible (e.g., the way your teenage son's driving habits changed after he painted his car red), you are more likely to say that this new plausible factor should be investigated. One's judgment about the plausibility of a factor being important is based on one's knowledge of the domain.

Other data indicate that familiarity with the domain makes it easier to juggle different factors simultaneously, which in turn allows you to construct experiments that simultaneously control for more factors. For example, in one experiment, eighth-graders completed two tasks. In one, they were to manipulate conditions in a com-

---

**Did Sherlock Holmes Take a Course in Critical Thinking?**

No one better exemplifies the power of broad, deep knowledge in driving critical thinking than Sherlock Holmes. In his famous first encounter with Dr. Watson, Holmes greets him with this observation: "You have been in Afghanistan, I perceive." Watson is astonished—how could Holmes have known? Eventually Holmes explains his insight, which turns not on incredible intelligence or creativity or wild guessing, but on having relevant knowledge. Holmes is told that Watson is a doctor; everything else he deduces by drawing on his knowledge of, among other things, the military, geography, how injuries heal, and current events. Here's how Holmes explains his thought process:

> I knew you came from Afghanistan. From long habit the train of thoughts ran so swiftly through my mind, that I arrived at the conclusion without being conscious of intermediate steps. There were such steps, however. The train of reasoning ran, "Here is a gentleman of a medical type, but with the air of a military man. Clearly an army doctor, then. He has just come from the tropics, for his face is dark, and that is not the natural tint of his skin, for his wrists are fair. He has undergone hardship and sickness, as his haggard face says clearly. His left arm has been injured. He holds it in a stiff and unnatural manner. Where in the tropics could an English army doctor have seen much hardship and got his arm wounded? Clearly in Afghanistan."
> "The whole train of thought did not occupy a second. I then remarked that you came from Afghanistan, and you were astonished."

*—EDITORS*

Source: A Study in Scarlet by Sir Arthur Conan Doyle.

---

16 AMERICAN EDUCATOR  SUMMER
Subjects who started with more and better integrated knowledge planned more informative experiments and made better use of experimental outcomes.

They tell you that your understanding is incomplete, and they guide the development of new hypotheses. But you could only recognize the outcome of an experiment as anomalous if you had some expectation of how it would turn out. And that expectation would be based on domain knowledge, as would your ability to create a new hypothesis that takes the anomalous outcome into account.

The idea that scientific thinking must be taught hand in hand with scientific content is further supported by research on scientific problem solving; that is, when students calculate an answer to a textbook-like problem, rather than design their own experiment. A meta-analysis of 40 experiments investigating methods for teaching scientific problem solving showed that effective approaches were those that focused on building complex integrated knowledge bases as part of problem solving, for example by including exercises like concept mapping. Ineffective approaches focused exclusively on the strategies to be used in problem solving while ignoring the knowledge necessary for the solution.

What do all these studies boil down to? First, critical thinking (as well as scientific thinking and other domain-based thinking) is not a skill. There is not a set of critical thinking skills that can be acquired and deployed regardless of context. Second, there are metacognitive strategies that, once learned, make critical thinking more likely. Third, the ability to think critically (to actually do what the metacognitive strategies call for) depends on domain knowledge and practice. For teachers, the situation is not hopeless, but no one should underestimate the difficulty of teaching students to think critically.

Endnotes

9 Spellman, B. A. (1996). “Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes,”
Teaching Critical Thinking

Teaching students to think critically is high on any teacher’s to-do list. So what strategies are consistent with the research?

- **Special programs aren’t worth it.** In the sidebar on page 12, I’ve mentioned a few of the better known programs. Despite their widespread availability, the evidence that these programs succeed in teaching students to think critically, especially in novel situations, is very limited. The modest boost that such programs may provide should be viewed, as should all claims of educational effectiveness, in light of their opportunity costs. Every hour students spend on the program is an hour they won’t be learning something else.

- **Thinking critically should be taught in the context of subject matter.** The foregoing does not mean that teachers shouldn’t teach students to think critically—it means that critical thinking shouldn’t be taught on its own. People do not spontaneously examine assumptions that underlie their thinking, try to consider all sides of an issue, question what they know, etc. These things must be modeled for students, and students must be given opportunities to practice—preferably in the context of normal classroom activity. This is true not only for science (as discussed in the main article), but for other subject matter. For example, an important part of thinking like a historian is considering the source of a document—who wrote it, when, and why. But teaching students to ask that question, independent of subject matter knowledge, won’t do much good. Knowing that a letter was written by a Confederate private to his wife in New Orleans just after the Battle of Vicksburg won’t help the student interpret the letter unless he knows something of Civil War history.

- **Student experiences offer entree to complex concepts.** Although critical thinking needs to be nested in subject matter, when students don’t have much subject matter knowledge, introducing a concept by drawing on student experiences can help. For example, the importance of a source in evaluating a historical document is familiar to even young children; deepening their understanding is a matter of asking questions that they have the knowledge to grapple with. Elementary school teachers could ask: Would a letter to a newspaper editor that criticized the abolishment of recess be viewed differently if written by a school principal versus a third-grader? Various concepts that are central to scientific thinking can also be taught with examples that draw on students’ everyday knowledge and experience. For example, “correlation does not imply causation” is often illustrated by the robust association between the consumption of ice cream and the number of crimes committed on a given day. With a little prodding, students soon realize that ice cream consumption doesn’t cause crime, but high temperatures might cause increases in both.

- **To teach critical thinking strategies, make them explicit and practice them.** Critical thinking strategies are abstractions. A plausible approach to teaching them is to make them explicit, and to proceed in stages. The first time (or several times) the concept is introduced, explain it with at least two different examples (possibly examples based on students’ experiences, as discussed above), label it so as to identify it as a strategy that can be applied in various contexts, and show how it applies to the course content at hand. In future instances, try naming the appropriate critical thinking strategy to see if students remember it and can figure out how it applies to the material under discussion. With still more practice, students may see which strategy applies without a cue from you.

Knowing that a letter was written by a Confederate private to his wife in New Orleans just after the Battle of Vicksburg won’t help the student interpret the letter—unless he knows something of Civil War history.

—D.W.


Sidebar Endnotes (p. 12)


Geoff Masters (/columnists/geoff-masters)
Chief Executive of the Australian Council for Educational Research

Learning from mistakes

In his book *Black Box Thinking*, Matthew Syed argues that some organisations and fields of endeavour are better than others at learning from mistakes.

Syed names health care as an area that, historically, has been slow to learn from mistakes. Too often, he argues, healthcare mistakes are stigmatised, doctors are expected to be infallible, and systems are set up to ignore and deny errors rather than to investigate and learn from them. Data are not always collected and analysed in ways that allow learning; poor practices and mistakes go undetected; and, in the English-speaking world alone, tens of thousands of patients die each year as a direct result of medical errors.

On the other hand, Syed argues, the aviation industry has developed an admirable willingness and tenacity to investigate the reasons for failure. It has created systems and cultures that enable the entire industry to learn from individual mistakes.

The title of Syed’s book is a reference to the ‘black box’ flight recorder. He draws a parallel with the way science progresses through a process of self-correction that depends on testing ideas, subjecting them to failure and profession-wide learning.

Syed concludes that learning from mistakes depends on:

- a willingness to learn from failure, and
• investigative skills and systems to extract and promulgate key lessons from mistakes.

**Willingness to learn**

A willingness to acknowledge and learn from failure is essential for all progress. Obstacles to acknowledging mistakes include the desire to protect personal reputations and deeply held beliefs. However, when evidence is avoided, hidden or ignored, learning opportunities are lost. This is true at the level of individuals and is equally true at the level of organisations, industries and professions.

Learning from failure is facilitated by a culture that accepts mistakes and welcomes the learning opportunities they provide. The fear of blame undermines the possibility of learning. Openness to learning requires personal and organisational resilience – an ability to acknowledge failure and to treat it as a learning opportunity.

**Systems and skills**

Learning from failure also depends on systems and skills to investigate, analyse and understand errors. As Syed notes, often this means looking beyond the obvious data to the underlying explanations or causes. Feedback on the reasons for mistakes and failures is essential to learning. But the identification of underpinning explanations usually requires well-honed skills of investigation.

Syed concludes that the paradox of success is that it is built on failure: ‘every error, every flaw, every failure, however small, is a marginal gain in disguise’. Evolutionary progress depends on feedback about what is working and what is not. Without such feedback, learning is severely constrained.

**How well does education learn from mistakes?**

It is interesting to reflect on our attitudes to mistakes and failures in
education.

To what extent are we prepared to call a government initiative that produced no measurable improvement a ‘mistake’? How often do we identify educational resources and programs as ‘failures’? Are we prepared to call the widespread promotion of an ineffective teaching strategy an ‘error’? Do we jump too quickly to defend and rationalise our efforts as well-intentioned and perhaps not entirely unsuccessful? And how well do we learn from such missteps?

Of course, education is not the only public policy area in which initiatives are inadequately evaluated, thereby providing little opportunity for learning. Many policy initiatives are based on personal ideologies or beliefs about what should work. Many are implemented on short timelines within the constraints of political cycles and disappear with changes in governments, ministers or senior bureaucrats. But when education policies and initiatives are denied the opportunity to fail (or to succeed), governments, education systems and the profession are denied opportunities to learn.

A willingness to learn not only means allowing time for success or failure, it also means committing to investigating why initiatives have succeeded or failed. Too often in education, policies and programs are introduced with no accompanying plan to evaluate and study their impact. Learning from mistakes requires a learning culture: a commitment to a long-term agenda to learn from evidence; a willingness to evaluate, identify and acknowledge successes and failures; and agreed systems for investigating and learning from practice.

Learning from student mistakes
There are implications, too, for our attitudes to failure in the classroom. If mistakes are essential to learning, to what extent do we design teaching to produce mistakes? Or do we instead stigmatise failure by sending students the message that we do not expect mistakes and that successful learning means not making errors?

One way to minimise mistakes is to assign tasks within students’ comfort zones. If tasks are relatively easy, failure is unlikely. But so too is learning. Successful learning is most likely when students are given challenges beyond their comfort zones – challenges that stretch and extend them to the point of making mistakes from which they can learn.

A first requirement then is a willingness to see mistakes not as something to be avoided, but as something to be embraced. This has implications for both teachers and students. Much teaching is focused on creating conditions for student success. But effective teaching often means providing opportunities for students to make mistakes and to learn. Students need to be assisted to welcome new challenges and to view mistakes not as reflections on their ability, but as vital steps in the learning process.

A second requirement is professional skill in analysing the reasons for student mistakes. Student errors are often superficial indicators of underlying misunderstandings or inadequately developed skills. There is no point creating the conditions for failure if there is no intention to investigate causes and provide feedback to guide learning. Skills of investigation and diagnosis are crucial to effective teaching and essential prerequisites for learning from mistakes.
Teaching for understanding
A meaningful education for 21st century learners

Why should we teach for understanding?
While the educational goal of helping students understand is certainly not new, there are three reasons why the need to teach for understanding has never been greater.

1. The nature of the modern world that today's students will enter.
The world is increasingly interconnected and rapidly changing, offering new potentials and problems. Search engines, computers and smartphones give most people increased and immediate access to huge amounts of information. E-mail, Twitter, Facebook and soon-to-be invented technologies enable instantaneous communication with people throughout the globe. The highly complex job market, with its array of novel and changing careers, calls for creative, innovative individuals who can apply their learning to new situations while functioning as continuous, lifelong learners.

These changes imply that we must educate in new and different ways if we are to prepare our children for a 21st century world. We can no longer focus education around the acquisition of knowledge - information is too easily accessible with the touch of a screen. Rather, education today must help students go beyond learning facts in order to develop deeper understandings of the world around them and the diverse global society in which they live. Our children need to learn how to find, sort, evaluate and apply information to new situations. They need to learn how to ask critical questions and solve difficult and messy problems. They need to develop a deeper understanding of key concepts and processes that will help them flourish in an unpredictable world.

2. The knowledge explosion.
This current reality is complicated by the fact that the knowledge base in many fields continues to expand, and most teachers today find that there is too much content to teach and not enough time to teach it all! The present curriculum simply contains too many topics and is often fragmented, without clear connections from one topic or one level to the next. The pressures of content coverage, often driven by jam-packed State standards and high stakes, standardised tests, come at the expense of engaging learners in exploring concepts in depth, addressing complex issues and problems, or investigating interesting and important questions - many of the very skills and processes needed to succeed in the modern world.

Teaching for understanding calls for a fundamental shift from a content "coverage" approach - teaching and testing a series of facts and discrete skills - to one that emphasises the "coverage" of important, transferable ideas and processes. Contemporary education must shift from an emphasis on knowledge acquisition for its own sake to preparing learners to understand ideas and processes that they can use and apply flexibly and autonomously.

Teaching for understanding is reinforced by recent insights into how people learn, and our work as educators should be guided by the most current understandings about the learning process. Over the course of the past twenty years, research in cognitive psychology and neuroscience has significantly expanded our understanding of how people learn. This research supports a set of learning principles that emphasise the importance of constructing meaning and developing understanding. Here are some of the key findings that undergird teaching for understanding and their implications:

- Views of how effective learning proceeds have shifted from the benefits of diligent drill and practice to a focus on understanding and applying knowledge.
- The knowledge of experts is not simply a list of facts and formulas that are relevant to their expertise; instead, their deeper understanding of key concepts and ideas (e.g., Newton's second law of motion) supports their ability to transfer learning to other contexts. Novices' knowledge is much less likely to be organised around big ideas; they are more likely to approach problems by searching for correct formulas and pat answers that fit their everyday intuitions.
- Knowledge learned at the level of rote memory rarely transfers. Transfer most likely occurs when the learner knows and understands underlying principles that can be applied to problems in new contexts.

Learning with understanding is more likely to promote transfer than simply memorising information from a text or a lecture.
- Skills and knowledge must be extended beyond the narrow contexts in which they are initially learned. For example, knowing how to solve a math problem in school may not transfer to solving math problems in other contexts. It is essential for learners to develop a sense of when what has been learned can be used - the conditions of application.
- Curricula that are a "mile wide and an inch deep" run the risk of developing disconnected rather than connected knowledge. Research on expertise suggests that superficial presentation of information in many topics may be a poor way to help students develop conceptual understandings and competencies or remember important information that will prepare them for future learning and work.
- It is not sufficient to provide assessments that focus primarily on testing memory of facts and formulas if the goal of learning is to enhance understanding and applicability of knowledge. Many current assessments primarily measure factual knowledge and low-level skills, and never determine whether students know when, where, why, and how to use that knowledge. Given the goal of learning with understanding, assessments and feedback must focus on understanding and not solely on remembering procedures or facts.

What is understanding?
If understanding is a worthy educational goal, then educators need clarity about its meaning. What is understanding? How would we know that a student really understands? In Understanding by Design, Wiggins and McTighe describe the nature of understanding and also propose that understanding is revealed through six facets that offer different types of evidence of understanding. Here is a brief summary of each of the six facets:

When someone truly understands, they:
- Can explain concepts, principles and processes by putting it their own words, teaching it to others, justifying their answers and showing their reasoning.
JAY MCTIGHE AND ELLIOTT SEIF

- Can interpret by making sense of data, text and experience through images, analogies, stories and models.
- Can apply by effectively using and adapting what they know in new and complex contexts.
- Can demonstrate perspective by seeing the big picture and recognizing different points of view.
- Display empathy by perceiving sensitively and walking in someone else’s shoes.
- Have self-knowledge by showing metacognitive awareness and reflecting on the meaning of the learning and experience.

The six facets offer a framework for creating rich learning activities that develop and deepen students’ understanding. They can also be used to develop assessments that determine whether students understand concepts and can apply learning to new situations. For example, we suggest that students have regular opportunities to explain a scientific principle in their own words, interpret literature and data, apply and transfer knowledge and skills to new and novel situations, form opinions based on evidence while considering the perspectives of others, self-assess their work and reflect on their learning. The use of activities and assessments based on the six facets will go a long way towards promoting understanding in schools and classrooms.

How do we teach for understanding?
Teaching for understanding involves two interrelated approaches:

1. Engage learners in meaning-making

Meaning-making occurs when learners are given the opportunity to construct their own understanding around big ideas and essential questions. Observing young children helps us to understand how learners make meaning. They often ask questions that begin their learning process. They learn through physical and mental activity that helps them make connections and construct their own meaning. For them, learning core concepts is not a linear process. They refine and revisit concepts over time and move from simple, sometimes erroneous constructs to more sophisticated, accurate concepts. Young children’s learning is mediated through thinking – asking questions, analysing and interpreting what they see, putting ideas together, making inferences, trying to solve problems and learning to reason and strategise. Moreover, learning for young children is often a social, collaborative process with adults and other children. Usually, the more opportunities there are for interaction, the greater the learning and understanding. Unlike factual information that can be transmitted by telling, understanding must be “earned” by the learner. In other words, coming to an understanding requires an active construction of meaning. We encourage teachers to help students construct meaning by focusing learning around big ideas and essential questions. Big ideas and essential questions are chosen because they are fundamental to a discipline, thought provoking and support transfer of learning to new situations. A history teacher may focus learning around the question, “How do we learn to live together in a diverse society?” and use the concept of diversity as a focus for learning. A science teacher might focus on the question “How do we know what to believe about a scientific claim?” and focus on the big idea of scientific truth. An art teacher might focus on the question “What makes great art?” and concentrate on the big idea artistic excellence. A focus on a smaller number of core ideas allows for a greater emphasis on in-depth learning. In the words of Newmann, we should create a curriculum with a “sustained examination of a few topics rather than superficial coverage of many”.

A variety of instructional activities encourage meaning-making. These include Socratic questioning and related inquiry approaches: classification and categorisation of information and data, developing and testing hypotheses, conducting research, drawing conclusions, explaining results and using project and problem-based learning strategies.

With a meaning-making perspective and appropriate instructional strategies, their learning without someone telling them what to do and when to do it. In the real world, no teacher is there to direct and remind them about which lesson to plug in here or there. Transfer is about intelligently and effectively drawing from their repertoire, independently, to handle new contexts on their own”.

The use of performance tasks help to reveal whether students understand core ideas and are able to transfer and adapt their learning to new situations. Here are several examples of such tasks:

Reading, Writing and Literature: After reading several fables and studying their characteristics, write a modern-day fable in order to teach a lesson about the characteristics of fables to younger children.

Geography: Develop a proposed route for a continental highway across central Africa, considering human and physical geography along with economic and political factors.

Mathematics: Based on a building blueprint, determine the amount of paint and cost estimates for painting the interior of a building.

Science: Design an experiment to test the capacity of different types of fabrics (e.g., cotton, wool, silk) for absorbing liquids.

Technology: Create a YouTube tutorial to teach your grandparents how to use Twitter or Facebook.

Visual Arts: Create an original mural or
3-D sculpture for your school/community to symbolise its history and values.

We suggest that teachers establish realistic, authentic contexts for the performance tasks they offer to students. Authentic performance tasks usually reflect the way in which real people in the world outside of school use knowledge and skill to address various situations. Authentic tasks typically include a goal (e.g., solve a problem, analyse an issue, conduct an investigation, communicate for a purpose) and a target audience. These tasks yield tangible products (e.g., a position paper, a poster, a 3-D model) and performances (e.g., an oral presentation, a skit, a demonstration) that are valued in the wider world. Such tasks often include realistic constraints such as time, schedule, and budget.

In order to develop transfer abilities, students need multiple opportunities to apply knowledge and skills in novel and realistic contexts, and that is precisely the opportunity that authentic performance tasks offer. Moreover, the regular use of such tasks signals to learners that a major goal of education is to enable them to use their learning in ways valued in the wider world beyond the classroom.

In transfer activities, learners have many opportunities to apply their learning and practice transferable skills in new and varied situations. When transfer is the goal, we propose that a teacher function more like a coach in athletics or the arts. A coach observes and assesses students’ efforts, and provides timely and ongoing feedback to help them improve their performance. Of course, there is a role for direct instruction and modeling of skills and strategies associated with transfer performance. But the ultimate goal is to render the teacher/coach unnecessary, since we want learners to be able to independently transfer their learning. Thus, over time, teacher support and scaffolding is gradually reduced so that students become increasingly capable of transferring their learning on their own.

Yes, but...

What about factual learning and “basic” skills? Aren’t there things that students just need to remember and specific skills they need to know how to do? How can they think and apply without knowledge and specific skills? Our emphasis on teaching for understanding through meaning making and transfer is not meant to suggest that learning core knowledge and developing proficiency in fundamental skills is not important. However, what we are proposing is that teachers help students learn facts and skills in the larger context of understanding and transfer. The learning principles cited above support the idea that learners are more likely to appreciate and remember specific facts and skills when these are connected to conceptually larger ideas used in authentic and meaningful ways.

What would we see in classrooms where teaching and assessing for understanding is emphasised?

How do these ideas play out in classrooms? A thorough examination of all the relevant instructional methods and teaching techniques for creating understanding based classrooms is beyond the scope of this article. However, in figure one (below), we describe a set of ten observable indicators that highlight the practices we would expect to see in classrooms where understanding based teaching and learning are occurring. This list can be used by teachers for self-assessment and by administrators, coaches and mentors for classroom observations.

Some Final Thoughts

A rapidly changing world with easy access to information, new social media, shifting employment needs and an explosion of knowledge requires a different way of thinking about educational practice. We can no longer dedicate a major portion of our teaching and testing to the “coverage” of facts and discrete skills.

Instead of emphasising rote learning of superficial content, teaching for understanding focuses on engaging learners in “meaning making” by exploring essential questions and engaging in meaningful applications of learning. Understanding core ideas and the ability to transfer them to new situations should be the twin goals of education today.

The principles and practices of the Understanding by Design® framework by McTighe and Wiggins offer educators a practical and proven approach for restructuring curriculum, assessment and instruction to achieve these goals. Rethinking teaching and learning using an understanding-based framework will lead to a more meaningful, authentic approach to learning in a 21st century world.

---

### Teaching and Assessing for Understanding: Observable Classroom Indicators

To what extent are...

1. Instruction and assessment focused around “big ideas” and essential questions? 4 3 2 1
2. Essential questions posted and examined throughout a unit? 4 3 2 1
3. Pre-assessments used to check students’ prior knowledge and potential misconceptions regarding new topics of study? 4 3 2 1
4. Opening “hooks” used to engage students in exploring the big ideas and essential questions? 4 3 2 1
5. Students’ understanding of the “big ideas” and core processes assessed through authentic performance tasks involving one or more of the six facets? 4 3 2 1
6. Evaluations of student products/performances based upon known criteria/rubrics, performance standards, and models (exemplars)? 4 3 2 1
7. Appropriate instructional strategies are used to help learners make meaning of the big Ideas, transfer their learning, and acquire requisite knowledge and skills? 4 3 2 1
8. Students are given multiple opportunities to demonstrate understanding using the six facets – explanation, interpretation, application, perspective, empathy and self-reflection. 4 3 2 1
9. Students given regular opportunities to rethink, revise and reflect on their work based on feedback from on-going formative assessments? 4 3 2 1
10. Students expected to self-assess, reflect on their work/learning, and set goals for improvement? 4 3 2 1

**Key:** 4 = extensively 3 = generally 2 = sometimes 1 = rarely
Improving Students’ Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology

John Dunlosky¹, Katherine A. Rawson¹, Elizabeth J. Marsh², Mitchell J. Nathan³, and Daniel T. Willingham⁴

¹Department of Psychology, Kent State University; ²Department of Psychology and Neuroscience, Duke University; ³Department of Educational Psychology, Department of Curriculum & Instruction, and Department of Psychology, University of Wisconsin–Madison; and ⁴Department of Psychology, University of Virginia

Summary
Many students are being left behind by an educational system that some people believe is in crisis. Improving educational outcomes will require efforts on many fronts, but a central premise of this monograph is that one part of a solution involves helping students to better regulate their learning through the use of effective learning techniques. Fortunately, cognitive and educational psychologists have been developing and evaluating easy-to-use learning techniques that could help students achieve their learning goals. In this monograph, we discuss 10 learning techniques in detail and offer recommendations about their relative utility. We selected techniques that were expected to be relatively easy to use and hence could be adopted by many students. Also, some techniques (e.g., highlighting and rereading) were selected because students report relying heavily on them, which makes it especially important to examine how well they work. The techniques include elaborative interrogation, self-explanation, summarization, highlighting (or underlining), the keyword mnemonic, imagery use for text learning, rereading, practice testing, distributed practice, and interleaved practice.

To offer recommendations about the relative utility of these techniques, we evaluated whether their benefits generalize across four categories of variables: learning conditions, student characteristics, materials, and criterion tasks. Learning conditions include aspects of the learning environment in which the technique is implemented, such as whether a student studies alone or with a group. Student characteristics include variables such as age, ability, and level of prior knowledge. Materials vary from simple concepts to mathematical problems to complicated science texts. Criterion tasks include different outcome measures that are relevant to student achievement, such as those tapping memory, problem solving, and comprehension.

We attempted to provide thorough reviews for each technique, so this monograph is rather lengthy. However, we also wrote the monograph in a modular fashion, so it is easy to use. In particular, each review is divided into the following sections:

1. General description of the technique and why it should work
2. How general are the effects of this technique?
   2a. Learning conditions
   2b. Student characteristics
   2c. Materials
   2d. Criterion tasks
3. Effects in representative educational contexts
4. Issues for implementation
5. Overall assessment

Corresponding Author:
John Dunlosky, Psychology, Kent State University, Kent, OH 44242
E-mail: j dunlosk@kent.edu
The review for each technique can be read independently of the others, and particular variables of interest can be easily compared across techniques.

To foreshadow our final recommendations, the techniques vary widely with respect to their generalizability and promise for improving student learning. Practice testing and distributed practice received high utility assessments because they benefit learners of different ages and abilities and have been shown to boost students’ performance across many criterion tasks and even in educational contexts. Elaborative interrogation, self-explanation, and interleaved practice received moderate utility assessments. The benefits of these techniques do generalize across some variables, yet despite their promise, they fell short of a high utility assessment because the evidence for their efficacy is limited. For instance, elaborative interrogation and self-explanation have not been adequately evaluated in educational contexts, and the benefits of interleaving have just begun to be systematically explored, so the ultimate effectiveness of these techniques is currently unknown. Nevertheless, the techniques that received moderate-utility ratings show enough promise for us to recommend their use in appropriate situations, which we describe in detail within the review of each technique.

Five techniques received a low utility assessment: summarization, highlighting, the keyword mnemonic, imagery use for text learning, and rereading. These techniques were rated as low utility for numerous reasons. Summarization and imagery use for text learning have been shown to help some students on some criterion tasks, yet the conditions under which these techniques produce benefits are limited, and much research is still needed to fully explore their overall effectiveness. The keyword mnemonic is difficult to implement in some contexts, and it appears to benefit students for a limited number of materials and for short retention intervals. Most students report rereading and highlighting, yet these techniques do not consistently boost students’ performance, so other techniques should be used in their place (e.g., practice testing instead of rereading).

Our hope is that this monograph will foster improvements in student learning, not only by showcasing which learning techniques are likely to have the most generalizable effects but also by encouraging researchers to continue investigating the most promising techniques. Accordingly, in our closing remarks, we discuss some issues for how these techniques could be implemented by teachers and students, and we highlight directions for future research.

**Introduction**

If simple techniques were available that teachers and students could use to improve student learning and achievement, would you be surprised if teachers were not being told about these techniques and if many students were not using them? What if students were instead adopting ineffective learning techniques that undermined their achievement, or at least did not improve it? Shouldn’t they stop using these techniques and begin using ones that are effective? Psychologists have been developing and evaluating the efficacy of techniques for study and instruction for more than 100 years. Nevertheless, some effective techniques are underutilized—many teachers do not learn about them, and hence many students do not use them, despite evidence suggesting that the techniques could benefit student achievement with little added effort. Also, some learning techniques that are popular and often used by students are relatively ineffective. One potential reason for the disconnect between research on the efficacy of learning techniques and their use in educational practice is that because so many techniques are available, it would be challenging for educators to sift through the relevant research to decide which ones show promise of efficacy and could feasibly be implemented by students (Pressley, Goodchild, Fleet, Zajchowski, & Evans, 1989).

Toward meeting this challenge, we explored the efficacy of 10 learning techniques (listed in Table 1) that students could use to improve their success across a wide variety of content domains. The learning techniques we consider here were chosen on the basis of the following criteria. We chose some techniques (e.g., self-testing, distributed practice) because an initial survey of the literature indicated that they could improve student success across a wide range of conditions. Other techniques (e.g., rereading and highlighting) were included because students report using them frequently. Moreover, students are responsible for regulating an increasing amount of their learning as they progress from elementary grades through middle school and high school to college. Lifelong learners also need to continue regulating their own learning, whether it takes place in the context of postgraduate education, the workplace, the development of new hobbies, or recreational activities.

Thus, we limited our choices to techniques that could be implemented by students without assistance (e.g., without requiring advanced technologies or extensive materials that would have to be prepared by a teacher). Some training may be required for students to learn how to use a technique with fidelity, but in principle, students should be able to use the techniques without supervision. We also chose techniques for which a sufficient amount of empirical evidence was available to support at least a preliminary assessment of potential efficacy. Of course, we could not review all the techniques that meet these criteria, given the in-depth nature of our reviews, and these criteria excluded some techniques that show much promise, such as techniques that are driven by advanced technologies.

Because teachers are most likely to learn about these techniques in educational psychology classes, we examined how some educational-psychology textbooks covered them (Ormrod, 2008; Santrock, 2008; Slavin, 2009; Snowman,
McCown, & Biehler, 2009; Sternberg & Williams, 2010; Woolfolk, 2007). Despite the promise of some of the techniques, many of these textbooks did not provide sufficient coverage, which would include up-to-date reviews of their efficacy and analyses of their generalizability and potential limitations. Accordingly, for all of the learning techniques listed in Table 1, we reviewed the literature to identify the generalizability of their benefits across four categories of variables—materials, learning conditions, student characteristics, and criterion tasks. The choice of these categories was inspired by Jenkins’ (1979) model (for an example of its use in educational contexts, see Marsh & Butler, in press), and examples of each category are presented in Table 2. Materials pertain to the specific content that students are expected to learn, remember, or comprehend. Learning conditions pertain to aspects of the context in which students are interacting with the to-be-learned materials. These conditions include aspects of the learning environment itself (e.g., noisiness vs. quietness in a classroom), but they largely pertain to the way in which a learning technique is implemented. For instance, a technique could be used only once or many times (a variable referred to as dosage) when students are studying, or a technique could be used when students are either reading or listening to the to-be-learned materials.

Any number of student characteristics could also influence the effectiveness of a given learning technique. For example, in comparison to more advanced students, younger students in early grades may not benefit from a technique. Students’ basic cognitive abilities, such as working memory capacity or general fluid intelligence, may also influence the efficacy of a given technique. In an educational context, domain knowledge refers to the valid, relevant knowledge a student brings to a lesson. Domain knowledge may be required for students to use some of the learning techniques listed in Table 1. For instance,
the use of imagery while reading texts requires that students know the objects and ideas that the words refer to so that they can produce internal images of them. Students with some domain knowledge about a topic may also find it easier to use self-explanation and elaborative interrogation, which are two techniques that involve answering “why” questions about a particular concept (e.g., “Why would particles of ice rise up within a cloud?”). Domain knowledge may enhance the benefits of summarization and highlighting as well. Nevertheless, although some domain knowledge will benefit students as they begin learning new content within a given domain, it is not a prerequisite for using most of the learning techniques.

The degree to which the efficacy of each learning technique obtains across long retention intervals and generalizes across different criterion tasks is of critical importance. Our reviews and recommendations are based on evidence, which typically pertains to students’ objective performance on any number of criterion tasks. Criterion tasks (Table 2, rightmost column) vary with respect to the specific kinds of knowledge that they tap. Some tasks are meant to tap students’ memory for information (e.g., “What is operant conditioning?”), others are largely meant to tap students’ comprehension (e.g., “Explain the difference between classical conditioning and operant conditioning”), and still others are meant to tap students’ application of knowledge (e.g., “How would you apply operant conditioning to train a dog to sit down?”). Indeed, Bloom and colleagues divided learning objectives into six categories, from memory (or knowledge) and comprehension of facts to their application, analysis, synthesis, and evaluation (B. S. Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956; for an updated taxonomy, see L. W. Anderson & Krathwohl, 2001).

In discussing how the techniques influence criterion performance, we emphasize investigations that have gone beyond demonstrating improved memory for target material by measuring students’ comprehension, application, and transfer of knowledge. Note, however, that although gaining factual knowledge is not considered the only or ultimate objective of schooling, we unabashedly consider efforts to improve student retention of knowledge as essential for reaching other instructional objectives; if one does not remember core ideas, facts, or concepts, applying them may prove difficult, if not impossible. Students who have forgotten principles of algebra will be unable to apply them to solve problems or use them as a foundation for learning calculus (or physics, economics, or other related domains), and students who do not remember what operant conditioning is will likely have difficulties applying it to solve behavioral problems. We are not advocating that students spend their time robotically memorizing facts; instead, we are acknowledging the important interplay between memory for a concept on one hand and the ability to comprehend and apply it on the other.

An aim of this monograph is to encourage students to use the appropriate learning technique (or techniques) to accomplish a given instructional objective. Some learning techniques are largely focused on bolstering students’ memory for facts (e.g., the keyword mnemonic), others are focused more on improving comprehension (e.g., self-explanation), and yet others may enhance both memory and comprehension (e.g., practice testing). Thus, our review of each learning technique describes how it can be used, its effectiveness for producing long-term retention and comprehension, and its breadth of efficacy across the categories of variables listed in Table 2.

**Reviewing the Learning Techniques**

In the following series of reviews, we consider the available evidence for the efficacy of each of the learning techniques. Each review begins with a brief description of the technique and a discussion about why it is expected to improve student learning. We then consider generalizability (with respect to learning conditions, materials, student characteristics, and criterion tasks), highlight any research on the technique that has been conducted in representative educational contexts, and address any identified issues for implementing the technique. Accordingly, the reviews are largely modular: Each of the 10 reviews is organized around these themes (with corresponding headers) so readers can easily identify the most relevant information without necessarily having to read the monograph in its entirety.

At the end of each review, we provide an overall assessment for each technique in terms of its relatively utility—low, moderate, or high. Students and teachers who are not already doing so should consider using techniques designated as *high utility*, because the effects of these techniques are robust and generalize widely. Techniques could have been designated as *low utility* or *moderate utility* for any number of reasons. For instance, a technique could have been designated as low utility because its effects are limited to a small subset of materials that students need to learn; the technique may be useful in some cases and adopted in appropriate contexts, but, relative to the other techniques, it would be considered low in utility because of its limited generalizability. A technique could also receive a low- or moderate-utility rating if it showed promise, yet insufficient evidence was available to support confidence in assigning a higher utility assessment. In such cases, we encourage researchers to further explore these techniques within educational settings, but students and teachers may want to use caution before adopting them widely. Most important, given that each utility assessment could have been assigned for a variety of reasons, we discuss the rationale for a given assessment at the end of each review.

Finally, our intent was to conduct exhaustive reviews of the literature on each learning technique. For techniques that have been reviewed extensively (e.g., distributed practice), however, we relied on previous reviews and supplemented them with any research that appeared after they had been published. For many of the learning techniques, too many articles have been published to cite them all; therefore, in our discussion of most of the techniques, we cite a subset of relevant articles.
I Elaborative interrogation

Anyone who has spent time around young children knows that one of their most frequent utterances is "Why?" (perhaps coming in a close second behind "No!"). Humans are inquisitive creatures by nature, attuned to seeking explanations for states, actions, and events in the world around us. Fortunately, a sizable body of evidence suggests that the power of explanatory questioning can be harnessed to promote learning. Specifically, research on both elaborative interrogation and self-explanation has shown that prompting students to answer "Why?" questions can facilitate learning. These two literatures are highly related but have mostly developed independently of one another. Additionally, they have overlapping but nonidentical strengths and weaknesses. For these reasons, we consider the two literatures separately.

1.1 General description of elaborative interrogation and why it should work. In one of the earliest systematic studies of elaborative interrogation, Pressley, McDaniel, Turnure, Wood, and Ahmad (1987) presented undergraduate students with a list of sentences, each describing the action of a particular man (e.g., "The hungry man got into the car"). In the elaborative-interrogation group, for each sentence, participants were prompted to explain "Why did that particular man do that?" Another group of participants was instead provided with an explanation for each sentence (e.g., "The hungry man got into the car to go to the restaurant"). The third group simply read each sentence. On a final test in which participants were cued to recall which man performed each action (e.g., "Who got in the car?"), the elaborative-interrogation group substantially outperformed the other two groups (collapsing across experiments, accuracy in this group was approximately 72%, compared with approximately 37% in each of the other two groups). From this and similar studies, Seifert (1993) reported average effect sizes ranging from 0.85 to 2.57.

As illustrated above, the key to elaborative interrogation involves prompting learners to generate an explanation for an explicitly stated fact. The particular form of the explanatory prompt has differed somewhat across studies—examples include "Why does it make sense that...?", "Why is this true?”, and simply "Why?" However, the majority of studies have used prompts following the general format, "Why would this fact be true of this [X] and not some other [X]?

The prevailing theoretical account of elaborative-interrogation effects is that elaborative interrogation enhances learning by supporting the integration of new information with existing prior knowledge. During elaborative interrogation, learners presumably "activate schemata... these schemata, in turn, help to organize new information which facilitates retrieval" (Willoughby & Wood, 1994, p. 140). Although the integration of new facts with prior knowledge may facilitate the organization (Hunt, 2006) of that information, organization alone is not sufficient—students must also be able to discriminate among related facts to be accurate when identifying or using the learned information (Hunt, 2006). Consistent with this account, note that most elaborative-interrogation prompts explicitly or implicitly invite processing of both similarities and differences between related entities (e.g., why a fact would be true of one province versus other provinces). As we highlight below, processing of similarities and differences among to-be-learned facts also accounts for findings that elaborative-interrogation effects are often larger when elaborations are precise rather than imprecise, when prior knowledge is higher rather than lower (consistent with research showing that preexisting knowledge enhances memory by facilitating distinctive processing; e.g., Rawson & Van Overschelde, 2008), and when elaborations are self-generated rather than provided (a finding consistent with research showing that distinctiveness effects depend on self-generating item-specific cues; Hunt & Smith, 1996).

1.2 How general are the effects of elaborative interrogation?

1.2a Learning conditions. The seminal work by Pressley et al. (1987; see also B. S. Stein & Bransford, 1979) spawned a flurry of research in the following decade that was primarily directed at assessing the generalizability of elaborative-interrogation effects. Some of this work focused on investigating elaborative-interrogation effects under various learning conditions. Elaborative-interrogation effects have been consistently shown using either incidental or intentional learning instructions (although two studies have suggested stronger effects for incidental learning: Pressley et al., 1987; Woloshyn, Wiloughby, Wood, & Pressley, 1990). Although most studies have involved individual learning, elaborative-interrogation effects have also been shown among students working in dyads or small groups (Kahl & Woloshyn, 1994; Woloshyn & Stockley, 1995).

1.2b Student characteristics. Elaborative-interrogation effects also appear to be relatively robust across different kinds of learners. Although a considerable amount of work has involved undergraduate students, an impressive number of studies have shown elaborative-interrogation effects with younger learners as well. Elaborative interrogation has been shown to improve learning for high school students, middle school students, and upper elementary school students (fourth through sixth graders). The extent to which elaborative interrogation benefits younger learners is less clear. Miller and Pressley (1989) did not find effects for kindergartners or first graders, and Wood, Miller, Symons, Canough, and Yedlicka (1993) reported mixed results for preschoolers. Nonetheless, elaborative interrogation does appear to benefit learners across a relatively wide age range. Furthermore, several of the studies involving younger students have also established elaborative-interrogation effects for learners of varying ability levels, including fourth through twelfth graders with learning disabilities (C. Greene, Symons, & Richards, 1996; Scruggs, Mastropieri, & Sullivan, 1994) and sixth through eighth graders with mild

Another key dimension along which learners differ is level of prior knowledge, a factor that has been extensively investigated within the literature on elaborative interrogation. Both correlational and experimental evidence suggest that prior knowledge is an important moderator of elaborative-interrogation effects, such that effects generally increase as prior knowledge increases. For example, Woloshyn, Pressley, and Schneider (1992) presented Canadian and German students with facts about Canadian provinces and German states. Thus, both groups of students had more domain knowledge for one set of facts and less domain knowledge for the other set. As shown in Figure 1, students showed larger effects of elaborative interrogation in their high-knowledge domain (a 24% increase) than in their low-knowledge domain (a 12% increase). Other studies manipulating the familiarity of to-be-learned materials have reported similar patterns, with significant effects for new facts about familiar items but weaker or nonexistent effects for facts about unfamiliar items. Despite some exceptions (e.g., Ozgungur & Guthrie, 2004), the overall conclusion that emerges from the literature is that high-knowledge learners will generally be best equipped to profit from the elaborative-interrogation technique. The benefit for low-knowledge learners is less certain.

One intuitive explanation for why prior knowledge moderates the effects of elaborative interrogation is that higher knowledge permits the generation of more appropriate explanations for why a fact is true. If so, one might expect final-test performance to vary as a function of the quality of the explanations generated during study. However, the evidence is mixed. Whereas some studies have found that test performance is better following adequate elaborative-interrogation responses (i.e., those that include a precise, plausible, or accurate explanation for a fact) than for inadequate responses, the differences have often been small, and other studies have failed to find differences (although the numerical trends are usually in the anticipated direction). A somewhat more consistent finding is that performance is better following an adequate response than no response, although in this case, too, the results are somewhat mixed. More generally, the available evidence should be interpreted with caution, given that outcomes are based on conditional post hoc analyses that likely reflect item-selection effects. Thus, the extent to which elaborative-interrogation effects depend on the quality of the elaborations generated is still an open question.

1.2c Materials. Although several studies have replicated elaborative-interrogation effects using the relatively artificial “man-sentences” used by Pressley et al. (1987), the majority of subsequent research has extended these effects using materials that better represent what students are actually expected to learn. The most commonly used materials involved sets of facts about various familiar and unfamiliar animals (e.g., “The Western Spotted Skunk’s hole is usually found on a sandy piece of farmland near crops”), usually with an elaborative-interrogation prompt following the presentation of each fact. Other studies have extended elaborative-interrogation effects to fact lists from other content domains, including facts about U.S. states, German states, Canadian provinces, and universities; possible reasons for dinosaur extinction; and gender-specific facts about men and women. Other studies have shown elaborative-interrogation effects for factual statements about various topics (e.g., the solar system) that are normatively consistent or inconsistent with learners’ prior beliefs (e.g., Woloshyn, Paivio, & Pressley, 1994). Effects have also been shown for facts contained in longer connected discourse, including expository texts on animals (e.g., Seifert, 1994); human digestion (B. L. Smith, Holliday, & Austin, 2010); the neuropsychology of phantom pain (Ozgungur & Guthrie, 2004); retail, merchandising, and accounting (Dornisch & Sperling, 2006); and various science concepts (McDaniel & Donnelly, 1996). Thus, elaborative-interrogation effects are relatively robust across factual material of different kinds and with different contents. However, it is important to note that elaborative interrogation has been applied (and may be applicable) only to discrete units of factual information.

1.2d Criterion tasks. Whereas elaborative-interrogation effects appear to be relatively robust across materials and learners, the extensions of elaborative-interrogation effects across measures that tap different kinds or levels of learning is somewhat more limited. With only a few exceptions, the majority of elaborative-interrogation studies have relied on the
following associative-memory measures: cued recall (generally involving the presentation of a fact to prompt recall of the entity for which the fact is true; e.g., "Which animal . . .?"
and matching (in which learners are presented with lists of facts and entities and must match each fact with the correct entity). Effects have also been shown on measures of fact recognition (B. L. Smith et al., 2010; Woloshyn et al., 1994; Woloshyn & Stockley, 1995). Concerning more generative measures, a few studies have also found elaborative-interrogation effects on free-recall tests (e.g., Woloshyn & Stockley, 1995; Woloshyn et al., 1994), but other studies have not (Dornisch & Sperling, 2006; McDaniel & Donnelly, 1996).

All of the aforementioned measures primarily reflect memory for explicitly stated information. Only three studies have used measures tapping comprehension or application of the factual information. All three studies reported elaborative-interrogation effects on either multiple-choice or verification tests that required inferences or higher-level integration (Dornisch & Sperling, 2006; McDaniel & Donnelly, 1996; Ozgungor & Guthrie, 2004). Ozgungor and Guthrie (2004) also found that elaborative interrogation improved performance on a concept-relatedness rating task (in brief, students rated the pairwise relatedness of the key concepts from a passage, and rating coherence was assessed via Pathfinder analyses; however, Dornisch and Sperling (2006) did not find significant elaborative-interrogation effects on a problem-solving test. In sum, whereas elaborative-interrogation effects on associative memory have been firmly established, the extent to which elaborative interrogation facilitates recall or comprehension is less certain.

Of even greater concern than the limited array of measures that have been used is the fact that few studies have examined performance after meaningful delays. Almost all prior studies have administered outcome measures either immediately or within a few minutes of the learning phase. Results from the few studies that have used longer retention intervals are promising. Elaborative-interrogation effects have been shown after delays of 1–2 weeks (Scruggs et al., 1994; Woloshyn et al., 1994), 1–2 months (Kahl & Woloshyn, 1994; Willoughby, Waller, Wood, & MacKinnon, 1993; Woloshyn & Stockley, 1995), and even 75 and 180 days (Woloshyn et al., 1994). In almost all of these studies, however, the delayed test was preceded by one or more criterion tests at shorter intervals, introducing the possibility that performance on the delayed test was contaminated by the practice provided by the preceding tests. Thus, further work is needed before any definitive conclusions can be drawn about the extent to which elaborative interrogation produces durable gains in learning.

1.3 Effects in representative educational contexts. Concerning the evidence that elaborative interrogation will enhance learning in representative educational contexts, few studies have been conducted outside the laboratory. However, outcomes from a recent study are suggestive (B. L. Smith et al., 2010). Participants were undergraduates enrolled in an introductory biology course, and the experiment was conducted during class meetings in the accompanying lab section. During one class meeting, students completed a measure of verbal ability and a prior-knowledge test over material that was related, but not identical, to the target material. In the following week, students were presented with a lengthy text on human digestion that was taken from a chapter in the course textbook. For half of the students, 21 elaborative interrogation prompts were interspersed throughout the text (roughly one prompt per 150 words), each consisting of a paraphrased statement from the text followed by "Why is this true?" The remaining students were simply instructed to study the text at their own pace, without any prompts. All students then completed 105 true/false questions about the material (none of which were the same as the elaborative-interrogation prompts). Performance was better for the elaborative-interrogation group than for the control group (76% versus 69%), even after controlling for prior knowledge and verbal ability.

1.4 Issues for implementation. One possible merit of elaborative interrogation is that it apparently requires minimal training. In the majority of studies reporting elaborative-interrogation effects, learners were given brief instructions and then practiced generating elaborations for 3 or 4 practice facts (sometimes, but not always, with feedback about the quality of the elaborations) before beginning the main task. In some studies, learners were not provided with any practice or illustrative examples prior to the main task. Additionally, elaborative interrogation appears to be relatively reasonable with respect to time demands. Almost all studies set reasonable limits on the amount of time allotted for reading a fact and for generating an elaboration (e.g., 15 seconds allotted for each fact). In one of the few studies permitting self-paced learning, the time-on-task difference between the elaborative-interrogation and reading-only groups was relatively minimal (32 minutes vs. 28 minutes; B. L. Smith et al., 2010). Finally, the consistency of the prompts used across studies allows for relatively straightforward recommendations to students about the nature of the questions they should use to elaborate on facts during study.

With that said, one limitation noted above concerns the potentially narrow applicability of elaborative interrogation to discrete factual statements. As Hamilton (1997) noted, "elaborative interrogation is fairly prescribed when focusing on a list of factual sentences. However, when focusing on more complex outcomes, it is not as clear to what one should direct the 'why' questions" (p. 308). For example, when learning about a complex causal process or system (e.g., the digestive system), the appropriate grain size for elaborative interrogation is an open question (e.g., should a prompt focus on an entire system or just a smaller part of it?). Furthermore, whereas the facts to be elaborated are clear when dealing with fact lists, elaborating on facts embedded in lengthier texts will require students to identify their own target facts. Thus, students may need some instruction about the kinds of content to which
elaborative interrogation may be fruitfully applied. Dosage is
also of concern with lengthier text, with some evidence sug-
gesting that elaborative-interrogation effects are substantially
diluted (Calderon & McDaniel, 2007) or even reversed (Ram-
say, Sperling, & Dornisch, 2010) when elaborative-interro-
gation prompts are administered infrequently (e.g., one prompt
every 1 or 2 pages).

1.5 Elaborative interrogation: Overall assessment. We rate
elaborative interrogation as having moderate utility. Elabo-
rate-interrogation effects have been shown across a relatively
broad range of factual topics, although some concerns remain
about the applicability of elaborative interrogation to material
that is lengthier or more complex than fact lists. Concerning
learner characteristics, effects of elaborative interrogation
have been consistently documented for learners at least as
young as upper elementary age, but some evidence suggests
that the benefits of elaborative interrogation may be limited
for learners with low levels of domain knowledge. Concerning
criterion tasks, elaborative-interrogation effects have been
firmly established on measures of associative memory admin-
istered after short delays, but firm conclusions about the extent
to which elaborative interrogation benefits comprehension
and the extent to which elaborative-interrogation effects persist
across longer delays await further research. Further research
demonstrating the efficacy of elaborative interrogation in re-
presentative educational contexts would also be useful. In sum,
the need for further research to establish the generalizability of
elaborative-interrogation effects is primarily why this tech-
nique did not receive a high-utility rating.

2 Self-explanation

2.1 General description of self-explanation and why it
should work. In the seminal study on self-explanation, Berry
(1983) explored its effects on logical reasoning using the
Wason card-selection task. In this task, a student might see
four cards labeled “A,” “4,” “D,” and “3” and be asked to indi-
cate which cards must be turned over to test the rule “if a card
has A on one side, it has 3 on the other side” (an instantiation
of the more general “if P, then Q” rule). Students were first
asked to solve a concrete instantiation of the rule (e.g., flavor
of jam on one side of a jar and the sale price on the other);
accuracy was near zero. They then were provided with a mini-
mal explanation about how to solve the “if P, then Q” rule and
were given a set of concrete problems involving the use of this
and other logical rules (e.g., “if P, then not Q”). For this set of
concrete practice problems, one group of students was
prompted to self-explain while solving each problem by stat-
ing the reasons for choosing or not choosing each card.
Another group of students solved all problems in the set and
only then were asked to explain how they had gone about solv-
ing the problems. Students in a control group were not
prompted to self-explain at any point. Accuracy on the prac-
tice problems was 90% or better in all three groups. However,
when the logical rules were instantiated in a set of abstract
problems presented during a subsequent transfer test, the two
self-explanation groups substantially outperformed the control
group (see Fig. 2). In a second experiment, another control
group was explicitly told about the logical connection between
the concrete practice problems they had just solved and the
forthcoming abstract problems, but they fared no better (28%).

As illustrated above, the core component of self-explana-
tion involves having students explain some aspect of their pro-
cessing during learning. Consistent with basic theoretical
assumptions about the related technique of elaborative inter-
rogation, self-explanation may enhance learning by support-
ing the integration of new information with existing prior
knowledge. However, compared with the consistent prompts
used in the elaborative-interrogation literature, the prompts
used to elicit self-explanations have been much more variable
across studies. Depending on the variation of the prompt used,
the particular mechanisms underlying self-explanation effects
differ somewhat. The key continuum along which self-
explanation prompts differ concerns the degree to which they
are content-free versus content-specific. For example, many
studies have used prompts that include no explicit mention of
particular content from the to-be-learned materials (e.g.,
“Explain what the sentence means to you. That is, what new
information does the sentence provide for you? And how does it
relate to what you already know?”). On the other end of the
continuum, many studies have used prompts that are much
more content-specific, such that different prompts are used for

Fig. 2. Mean percentage of logical-reasoning problems answered cor-
rectly for concrete practice problems and subsequently administered ab-
stract transfer problems in Berry (1983). During a practice phase, learners
self-explained while solving each problem, self-explained after solving all
problems, or were not prompted to engage in self-explanation. Standard
errors are not available.
different items (e.g., "Why do you calculate the total acceptable outcomes by multiplying?" "Why is the numerator 14 and the denominator 7 in this step?"). For present purposes, we limit our review to studies that have used prompts that are relatively content-free. Although many of the content-specific prompts do elicit explanations, the relatively structured nature of these prompts would require teachers to construct sets of specific prompts to put into practice, rather than capturing a more general technique that students could be taught to use on their own. Furthermore, in some studies that have been situated in the self-explanation literature, the nature of the prompts is functionally more closely aligned with that of practice testing.

Even within the set of studies selected for review here, considerable variability remains in the self-explanation prompts that have been used. Furthermore, the range of tasks and measures that have been used to explore self-explanation is quite large. Although we view this range as a strength of the literature, the variability in self-explanation prompts, tasks, and measures does not easily support a general summative statement about the mechanisms that underlie self-explanation effects.

### 2.2 How general are the effects of self-explanation?

2.2a **Learning conditions.** Several studies have manipulated other aspects of learning conditions in addition to self-explanation. For example, Rittle-Johnson (2006) found that self-explanation was effective when accompanied by either direct instruction or discovery learning. Concerning potential moderating factors, Berry (1983) included a group who self-explained after the completion of each problem rather than during problem solving. Retrospective self-explanation did enhance performance relative to no self-explanation, but the effects were not as pronounced as with concurrent self-explanation. Another moderating factor may concern the extent to which provided explanations are made available to learners. Schworm and Renkl (2006) found that self-explanation effects were significantly diminished when learners could access explanations, presumably because learners made minimal attempts to answer the explanatory prompts before consulting the provided information (see also Aleven & Koedinger, 2002).

2.2b **Student characteristics.** Self-explanation effects have been shown with both younger and older learners. Indeed, self-explanation research has relied much less heavily on samples of college students than most other literatures have, with at least as many studies involving younger learners as involving undergraduates. Several studies have reported self-explanation effects with kindergartners, and other studies have shown effects for elementary school students, middle school students, and high school students.

In contrast to the breadth of age groups examined, the extent to which the effects of self-explanation generalize across different levels of prior knowledge or ability has not been sufficiently explored. Concerning knowledge level, several studies have used pretests to select participants with relatively low levels of knowledge or task experience, but no research has systematically examined self-explanation effects as a function of knowledge level. Concerning ability level, Chi, de Leeuw, Chiu, and LaVancher (1994) examined the effects of self-explanation on learning from an expository text about the circulatory system among participants in their sample who had received the highest and lowest scores on a measure of general aptitude and found gains of similar magnitude in each group. In contrast, Didierjean and Cauzinille-Maréméché (1997) examined algebra-problem solving in a sample of ninth graders with either low or intermediate algebra skills, and they found self-explanation effects only for lower-skill students. Further work is needed to establish the generality of self-explanation effects across these important idiographic dimensions.

2.2c **Materials.** One of the strengths of the self-explanation literature is that effects have been shown not only across different materials within a task domain but also across several different task domains. In addition to the logical-reasoning problems used by Berry (1983), self-explanation has been shown to support the solving of other kinds of logic puzzles. Self-explanation has also been shown to facilitate the solving of various kinds of math problems, including simple addition problems for kindergartners, mathematical-equivalence problems for elementary-age students, and algebraic formulas and geometric theorems for older learners. In addition to improving problem solving, self-explanation improved student teachers’ evaluation of the goodness of practice problems for use in classroom instruction. Self-explanation has also helped younger learners overcome various kinds of misconceptions, improving children’s understanding of false belief (i.e., that individuals can have a belief that is different from reality), number conservation (i.e., that the number of objects in an array does not change when the positions of those objects in the array change), and principles of balance (e.g., that not all objects balance on a fulcrum at their center point). Self-explanation has improved children’s pattern learning and adults’ learning of endgame strategies in chess. Although most of the research on self-explanation has involved procedural or problem-solving tasks, several studies have also shown self-explanation effects for learning from text, including both short narratives and lengthier expository texts. Thus, self-explanation appears to be broadly applicable.

2.2d **Criterion tasks.** Given the range of tasks and domains in which self-explanation has been investigated, it is perhaps not surprising that self-explanation effects have been shown on a wide range of criterion measures. Some studies have shown self-explanation effects on standard measures of memory, including free recall, cued recall, fill-in-the-blank tests, associative matching, and multiple-choice tests tapping explicitly stated information. Studies involving text learning have also shown effects on measures of comprehension, including diagram-drawing tasks, application-based questions, and tasks in which learners must make inferences on the basis of
information implied but not explicitly stated in a text. Across those studies involving some form of problem-solving task, virtually every study has shown self-explanation effects on near-transfer tests in which students are asked to solve problems that have the same structure as, but are nonidentical to, the practice problems. Additionally, self-explanation effects on far-transfer tests (in which students are asked to solve problems that differ from practice problems not only in their surface features but also in one or more structural aspects) have been shown for the solving of math problems and pattern learning. Thus, self-explanation facilitates an impressive range of learning outcomes.

In contrast, the durability of self-explanation effects is woefully underexplored. Almost every study to date has administered criterion tests within minutes of completion of the learning phase. Only five studies have used longer retention intervals. Self-explanation effects persisted across 1-2 day delays for playing chess endgames (de Bruin, Rikkers, & Schmidt, 2007) and for retention of short narratives (Magliano, Trabasso, & Graesser, 1999). Self-explanation effects persisted across a 1-week delay for the learning of geometric theorems (although an additional study session intervened between initial learning and the final test; R. M. F. Wong, Lawson, & Keeves, 2002) and for learning from a text on the circulatory system (although the final test was an open-book test; Chi et al., 1994). Finally, Rittle-Johnson (2006) reported significant effects on performance in solving math problems after a 2-week delay; however, the participants in this study also completed an immediate test, thus introducing the possibility that testing effects influenced performance on the delayed test. Taken together, the outcomes of these few studies are promising, but considerably more research is needed before confident conclusions can be made about the longevity of self-explanation effects.

2.3 Effects in representative educational contexts. Concerning the strength of the evidence that self-explanation will enhance learning in educational contexts, outcomes from two studies in which participants were asked to learn course-relevant content are at least suggestive. In a study by Schworm and Renkl (2006), students in a teacher-education program learned how to develop example problems to use in their classrooms by studying samples of well-designed and poorly designed example problems in a computer program. On each trial, students in a self-explanation group were prompted to explain why one of two examples was more effective than the other, whereas students in a control group were not prompted to self-explain. Half of the participants in each group were also given the option to examine experimenter-provided explanations on each trial. On an immediate test in which participants selected and developed example problems, the self-explanation group outperformed the control group. However, this effect was limited to students who had not been able to view provided explanations, presumably because students made minimal attempts to self-explain before consulting the provided information.

R. M. F. Wong et al. (2002) presented ninth-grade students in a geometry class with a theorem from the course textbook that had not yet been studied in class. During the initial learning session, students were asked to think aloud while studying the relevant material (including the theorem, an illustration of its proof, and an example of an application of the theorem to a problem). Half of the students were specifically prompted to self-explain after every 1 or 2 lines of new information (e.g., “What parts of this page are new to me? What does the statement mean? Is there anything I still don’t understand?”), whereas students in a control group received nonspecific instructions that simply prompted them to think aloud during study. The following week, all students received a basic review of the theorem and completed the final test the next day. Self-explanation did not improve performance on near-transfer questions but did improve performance on far-transfer questions.

2.4 Issues for implementation. As noted above, a particular strength of the self-explanation strategy is its broad applicability across a range of tasks and content domains. Furthermore, in almost all of the studies reporting significant effects of self-explanation, participants were provided with minimal instructions and little to no practice with self-explanation prior to completing the experimental task. Thus, most students apparently can profit from self-explanation with minimal training.

However, some students may require more instruction to successfully implement self-explanation. In a study by Didier-jean and Cauzinille-Marmêche (1997), ninth graders with poor algebra skills received minimal training prior to engaging in self-explanation while solving algebra problems; analysis of think-aloud protocols revealed that students produced many more paraphrases than explanations. Several studies have reported positive correlations between final-test performance and both the quantity and quality of explanations generated by students during learning. Further suggesting that the benefit of self-explanation might be enhanced by teaching students how to effectively implement the self-explanation technique (for examples of training methods, see Ainsworth & Burcham, 2007; R. M. F. Wong et al., 2002). However, in at least some of these studies, students who produced more or better-quality self-explanations may have had greater domain knowledge; if so, then further training with the technique may not have benefited the more poorly performing students. Investigating the contribution of these factors (skill at self-explanation vs. domain knowledge) to the efficacy of self-explanation will have important implications for how and when to use this technique.

An outstanding issue concerns the time demands associated with self-explanation and the extent to which self-explanation effects may have been due to increased time on task. Unfortunately, few studies equated time on task when comparing self-explanation conditions to control conditions involving other strategies or activities, and most studies involving self-paced practice did not report participants’ time on task. In the few
studies reporting time on task, self-paced administration usually yielded nontrivial increases (30–100%) in the amount of time spent learning in the self-explanation condition relative to other conditions, a result that is perhaps not surprising, given the high dosage levels at which self-explanation was implemented. For example, Chi et al. (1994) prompted learners to self-explain after reading each sentence of an expository text, which doubled the amount of time the group spent studying the text relative to a rereading control group (125 vs. 66 minutes, respectively). With that said, Schworm and Renkl (2006) reported that time on task was not correlated with performance across groups, and Ainsworth and Burcham (2007) reported that controlling for study time did not eliminate effects of self-explanation.

Within the small number of studies in which time on task was equated, results were somewhat mixed. Three studies finding time on task reported significant effects of self-explanation (de Bruin et al., 2007; de Koning, Tabbers, Rikers, & Paas, 2011; O’Reilly, Symons, & MacLatchy-Gaudet, 1998). In contrast, Matthews and Rittle-Johnson (2009) had one group of third through fifth graders practice solving math problems with self-explanation and a control group solve twice as many practice problems without self-explanation; the two groups performed similarly on a final test. Clearly, further research is needed to establish the bang for the buck provided by self-explanation before strong prescriptive conclusions can be made.

2.5 Self-explanation: Overall assessment. We rate self-explanation as having moderate utility. A major strength of this technique is that its effects have been shown across different content materials within task domains as well as across several different task domains. Self-explanation effects have also been shown across an impressive age range, although further work is needed to explore the extent to which these effects depend on learners’ knowledge or ability level. Self-explanation effects have also been shown across an impressive range of learning outcomes, including various measures of memory, comprehension, and transfer. In contrast, further research is needed to establish the durability of these effects across educationally relevant delays and to establish the efficacy of self-explanation in representative educational contexts. Although most research has shown effects of self-explanation with minimal training, some results have suggested that effects may be enhanced if students are taught how to effectively implement the self-explanation strategy. One final concern has to do with the nontrivial time demands associated with self-explanation, at least at the dosages examined in most of the research that has shown effects of this strategy.

3 Summarization

Students often have to learn large amounts of information, which requires them to identify what is important and how different ideas connect to one another. One popular technique for accomplishing these goals involves having students write summaries of to-be-learned texts. Successful summaries identify the main points of a text and capture the gist of it while excluding unimportant or repetitive material (A. L. Brown, Campione, & Day, 1981). Although learning to construct accurate summaries is often an instructional goal in its own right (e.g., Wade-Stein & Kintsch, 2004), our interest here concerns whether doing so will boost students’ performance on later criterion tests that cover the target material.

3.1 General description of summarization and why it should work. As an introduction to the issues relevant to summarization, we begin with a description of a prototypical experiment. Bretzing and Kulhavy (1979) had high school juniors and seniors study a 2,000-word text about a fictitious tribe of people. Students were assigned to one of five learning conditions and given up to 30 minutes to study the text. After reading each page, students in a summarization group were instructed to write three lines of text that summarized the main points from that page. Students in a note-taking group received similar instructions, except that they were told to take up to three lines of notes on each page of text while reading. Students in a verbatim-copying group were instructed to locate and copy the three most important lines on each page. Students in a letter-search group copied all the capitalized words in the text, also filling up three lines. Finally, students in a control group simply read the text without recording anything. (A subset of students from the four conditions involving writing were allowed to review what they had written, but for present purposes we will focus on the students who did not get a chance to review before the final test.) Students were tested either shortly after learning or 1 week later, answering 25 questions that required them to connect information from across the text. On both the immediate and delayed tests, students in the summarization and note-taking groups performed best, followed by the students in the verbatim-copying and control groups, with the worst performance in the letter-search group (see Fig. 3).

Bretzing and Kulhavy’s (1979) results fit nicely with the claim that summarization boosts learning and retention because it involves attending to and extracting the higher-level meaning and gist of the material. The conditions in the experiment were specifically designed to manipulate how much students processed the texts for meaning, with the letter-search condition involving shallow processing of the text that did not require learners to extract its meaning (Craik & Lockhart, 1972). Summarization was more beneficial than that shallow task and yielded benefits similar to those of note-taking, another task known to boost learning (e.g., Bretzing & Kulhavy, 1981; Crawford, 1925a, 1925b; Di Vesta & Gray, 1972). More than just facilitating the extraction of meaning, however, summarization should also boost organizational processing, given that extracting the gist of a text requires learners to connect disparate pieces of the text, as opposed to simply evaluating its individual components (similar to the way in which note-taking affords organizational processing; Einstein,
Improving Student Achievement

Fig. 3. Mean number of correct responses on a test occurring shortly after study as a function of text type (immediate or delayed) and learning condition in Bretzing and Kulhavy (1979). Error bars represent standard errors.

Morris, & Smith, 1985). One last point should be made about the results from Bretzing and Kulhavy (1979)—namely, that summarization and note-taking were both more beneficial than was verbatim copying. Students in the verbatim-copying group still had to locate the most important information in the text, but they did not synthesize it into a summary or rephrase it in their own words. Thus, writing about the important points in one's own words produced a benefit over and above that of selecting important information; students benefited from the more active processing involved in summarization and note-taking (see Wittrock, 1990, and Chi, 2009, for reviews of active/generative learning). These explanations all suggest that summarization helps students identify and organize the main ideas within a text.

So how strong is the evidence that summarization is a beneficial learning strategy? One reason this question is difficult to answer is that the summarization strategy has been implemented in many different ways across studies, making it difficult to draw general conclusions about its efficacy. Pressley and colleagues described the situation well when they noted that “summarization is not one strategy but a family of strategies” (Pressley, Johnson, Symons, McGoldrick, & Kurita, 1989, p. 5). Depending on the particular instructions given, students’ summaries might consist of single words, sentences, or longer paragraphs; be limited in length or not; capture an entire text or only a portion of it; be written or spoken aloud; or be produced from memory or with the text present.

A lot of research has involved summarization in some form, yet whereas some evidence demonstrates that summarization works (e.g., L. W. Brooks, Dansereau, Holley, & Spurlin, 1983; Doctorow, Wittrock, & Marks, 1978), T. H. Anderson and Armbruster’s (1984) conclusion that “research in support of summarizing as a studying activity is sparse indeed” (p. 670) is not outmoded. Instead of focusing on discovering when (and how) summarization works, by itself and without training, researchers have tended to explore how to train students to write better summaries (e.g., Friend, 2001; Hare & Borchardt, 1984) or to examine other benefits of training the skill of summarization. Still others have simply assumed that summarization works, including it as a component in larger interventions (e.g., Carr, Bigler, & Morningstar, 1991; Lee, Lim, & Grabowski, 2010; Palincsar & Brown, 1984; Spörer, Brunstein, & Kieschke, 2009). When collapsing across findings pertaining to all forms of summarization, summarization appears to benefit students, but the evidence for any one instantiation of the strategy is less compelling.

The focus on training students to summarize reflects the belief that the quality of summaries matters. If a summary does not emphasize the main points of a text, or if it includes incorrect information, why would it be expected to benefit learning and retention? Consider a study by Bednall and Kehoe (2011, Experiment 2), in which undergraduates studied six Web units that explained different logical fallacies and provided examples of each. Of interest for present purposes are two groups: a control group who simply read the units and a group in which students were asked to summarize the material as if they were explaining it to a friend. Both groups received the following tests: a multiple-choice quiz that tested information directly stated in the Web unit; a short-answer test in which, for each of a list of presented statements, students were required to name the specific fallacy that had been committed or write “not a fallacy” if one had not occurred; and, finally, an application test that required students to write explanations of logical fallacies in examples that had been studied (near transfer) as well as explanations of fallacies in novel examples (far transfer). Summarization did not benefit overall performance, but the researchers noticed that the summaries varied a lot in content; for one studied fallacy, only 64% of the summaries included the correct definition. Table 3 shows the relationships between summary content and later performance. Higher-quality summaries that contained more information and that were linked to prior knowledge were associated with better performance.

Several other studies have supported the claim that the quality of summaries has consequences for later performance. Most similar to the Bednall and Kehoe (2011) result is Ross and Di Vesta’s (1976) finding that the length (in words) of an oral summary (a very rough indicator of quality) correlated with later performance on multiple-choice and short-answer questions. Similarly, Dyer, Riley, and Yekovich (1979) found that final-test questions were more likely to be answered correctly if the information needed to answer them had been included in an earlier summary. Garner (1982) used a different
method to show that the quality of summaries matters: Undergraduates read a passage on Dutch elm disease and then wrote a summary at the bottom of the page. Five days later, the students took an old/new recognition test; critical items were new statements that captured the gist of the passage (as in Bransford & Franks, 1971). Students who wrote better summaries (i.e., summaries that captured more important information) were more likely to falsely recognize these gist statements, a pattern suggesting that the students had extracted a higher-level understanding of the main ideas of the text.

3.2 How general are the effects of summarization?

3.2a Learning conditions. As noted already, many different types of summaries can influence learning and retention; summarization can be simple, requiring the generation of only a heading (e.g., L. W. Brooks et al., 1983) or a single sentence per paragraph of a text (e.g., Doctorow et al., 1978), or it can be as complicated as an oral presentation on an entire set of studied material (e.g., Ross & Di Vesta, 1976). Whether it is better to summarize smaller pieces of a text (more frequent summarization) or to capture more of the text in a larger summary (less frequent summarization) has been debated (Foos, 1995; Spurlin, Dansereau, O’Donnell, & Brooks, 1988). The debate remains unresolved, perhaps because what constitutes the most effective summary for a text likely depends on many factors (including students’ ability and the nature of the material).

One other open question involves whether studied material should be present during summarization. Hidi and Anderson (1986) pointed out that having the text present might help the reader to succeed at identifying its most important points as well as relating parts of the text to one another. However, summarizing a text without having it present involves retrieval, which is known to benefit memory (see the Practice Testing section of this monograph), and also prevents the learner from engaging in verbatim copying. The Dyer et al. (1979) study described earlier involved summarizing without the text present; in this study, no overall benefit from summarizing occurred, even though information that had been included in summaries was benefited (overall, this benefit was overshadowed by costs to the greater amount of information that had not been included in summaries). More generally, some studies have shown benefits from summarizing an absent text (e.g., Ross & Di Vesta, 1976), but some have not (e.g., M. C. M. Anderson & Thiede, 2008, and Thiede & Anderson, 2003, found no benefits of summarization on test performance). The answer to whether studied text should be present during summarization is most likely a complicated one, and it may depend on people’s ability to summarize when the text is absent.

3.2b Student characteristics. Benefits of summarization have primarily been observed with undergraduates. Most of the research on individual differences has focused on the age of students, because the ability to summarize develops with age. Younger students struggle to identify main ideas and tend to write lower-quality summaries that retain more of the original wording and structure of a text (e.g., A. L. Brown & Day, 1983; A. L. Brown, Day, & Jones, 1983). However, younger students (e.g., middle school students) can benefit from summarization following extensive training (e.g., Arnbuster, Anderson, & Ostertag, 1987; Bean & Steenwyk, 1984). For example, consider a successful program for sixth-grade students (Rinehart, Stahl, & Erickson, 1986). Teachers received 90 minutes of training so that they could implement summarization training in their classrooms; students then completed five 45- to 50-minute sessions of training. The training reflected principles of direct instruction, meaning that students were explicitly taught about the strategy, saw it modeled, practiced it and received feedback, and eventually learned to monitor and check their work. Students who had received the training recalled more major information from a textbook chapter (i.e., information identified by teachers as the most important for students to know) than did students who had not, and this benefit was linked to improvements in note-taking. Similar training programs have succeeded with middle school students who are learning disabled (e.g., Gajria & Salvia, 1992; Malone & Mastropieri, 1991), minority high school students (Hare & Borchardt, 1984), and underprepared college students (A. King, 1992).

Outcomes of two other studies have implications for the generality of the summarization strategy, as they involve individual differences in summarization skill (a prerequisite for
using the strategy). First, both general writing skill and interest in a topic have been linked to summarization ability in seventh graders (Head, Readence, & Buss, 1989). Writing skill was measured via performance on an unrelated essay, and interest in the topic (American history) was measured via a survey that asked students how much they would like to learn about each of 25 topics. Of course, interest may be confounded with knowledge about a topic, and knowledge may also contribute to summarization skill. Recht and Leslie (1988) showed that seventh- and eighth-grade students who knew a lot about baseball (as measured by a pretest) were better at summarizing a 625-word passage about a baseball game than were students who knew less about baseball. This finding needs to be replicated with different materials, but it seems plausible that students with more domain-relevant knowledge would be better able to identify the main points of a text and extract its gist. The question is whether domain experts would benefit from the summarization strategy or whether it would be redundant with the processing in which these students would spontaneously engage.

3.2c Materials. The majority of studies have used prose passages on such diverse topics as a fictitious primitive tribe, desert life, geology, the blue shark, an earthquake in Lisbon, the history of Switzerland, and fictional stories. These passages have ranged in length from a few hundred words to a few thousand words. Other materials have included Web modules and lectures. For the most part, characteristics of materials have not been systematically manipulated, which makes it difficult to draw strong conclusions about this factor, even though 15 years have passed since Hidi and Anderson (1986) made an argument for its probable importance. As discussed in Yu (2009), it makes sense that the length, readability, and organization of a text might all influence a reader’s ability to summarize it, but these factors need to be investigated in studies that manipulate them while holding all other factors constant (as opposed to comparing texts that vary along multiple dimensions).

3.2d Criterion tasks. The majority of summarization studies have examined the effects of summarization on either retention of factual details or comprehension of a text (often requiring inferences) through performance on multiple-choice questions, cued recall questions, or free recall. Other benefits of summarization include enhanced metacognition (with text-based summarization improving the extent to which readers can accurately evaluate what they do or do not know; M. C. M. Anderson & Thiede, 2008; Thiede & Anderson, 2003) and improved note-taking following training (A. King, 1992; Rinehart et al., 1986).

Whereas several studies have shown benefits of summarization (sometimes following training) on measures of application (e.g., B. Y. L. Wong, Wong, Perry, & Sawatsky, 1986), others have failed to find such benefits. For example, consider a study in which L. F. Annis (1985) had undergraduates read a passage on an earthquake and then examined the consequences of summarization for performance on questions designed to tap different categories of learning within Bloom et al.’s (1956) taxonomy. One week after learning, students who had summarized performed no differently than students in a control group who had only read the passages in answering questions that tapped a basic level of knowledge (fact and comprehension questions). Students benefited from summarization when the questions required the application or analysis of knowledge, but summarization led to worse performance on evaluation and synthesis questions. These results need to be replicated, but they highlight the need to assess the consequences of summarization on the performance of tasks that measure various levels of Bloom’s taxonomy.

Across studies, results have also indicated that summarization helps later performance on generative measures (e.g., free recall, essays) more than it affects performance on multiple-choice or other measures that do not require the student to produce information (e.g., Bednall & Kehoe, 2011; L. W. Brooks et al., 1983; J. R. King, Biggs, & Lipsky, 1984). Because summarizing requires production, the processing involved is likely a better match to generative tests than to tests that depend on recognition.

Unfortunately, the one study we found that used a high-stakes test did not show a benefit from summarization training (Brozo, Stahl, & Gordon, 1985). Of interest for present purposes were two groups in the study, which was conducted with college students in a remedial reading course who received training either in summarization or in self-questioning (in the self-questioning condition, students learned to write multiple-choice comprehension questions). Training lasted for 4 weeks; each week, students received approximately 4 to 5 hours of instruction and practice that involved applying the techniques to 1-page news articles. Of interest was the students’ performance on the Georgia State Regents’ examination, which involves answering multiple-choice reading-comprehension questions about passages; passing this exam is a graduation requirement for many college students in the University System of Georgia (see http://www2.gsu.edu/~wwrrtp/). Students also took a practice test before taking the actual Regents’ exam. Unfortunately, the mean scores for both groups were at or below passing, for both the practice and actual exams. However, the self-questioning group performed better than the summarization group on both the practice test and the actual Regents’ examination. This study did not report pretraining scores and did not include a no-training control group, so some caution is warranted in interpreting the results. However, it emphasizes the need to establish that outcomes from basic laboratory work generalize to actual educational contexts and suggests that summarization may not have the same influence in both contexts.

Finally, concerning test delays, several studies have indicated that when summarization does boost performance, its effects are relatively robust over delays of days or weeks (e.g., Bretzing & Kulhavy, 1979; B. I. Stein & Kirby, 1992). Similarly, benefits of training programs have persisted several weeks after the end of training (e.g., Hare & Borchardt, 1984).
3.3 Effects in representative educational contexts. Several of the large summarization-training studies have been conducted in regular classrooms, indicating the feasibility of doing so. For example, the study by A. King (1992) took place in the context of a remedial study-skills course for undergraduates, and the study by Rinehart et al. (1986) took place in sixth-grade classrooms, with the instruction led by students' regular teachers. In these and other cases, students benefited from the classroom training. We suspect it may actually be more feasible to conduct these kinds of training studies in classrooms than in the laboratory, given the nature of the time commitment for students. Even some of the studies that did not involve training were conducted outside the laboratory; for example, in the Bednall and Kehoe (2011) study on learning about logical fallacies from Web modules (see data in Table 3), the modules were actually completed as a homework assignment. Overall, benefits can be observed in classroom settings; the real constraint is whether students have the skill to successfully summarize, not whether summarization occurs in the lab or the classroom.

3.4 Issues for implementation. Summarization would be feasible for undergraduates or other learners who already know how to summarize. For these students, summarization would constitute an easy-to-implement technique that would not take a lot of time to complete or understand. The only concern would be whether these students might be better served by some other strategy, but certainly summarization would be better than the study strategies students typically favor, such as highlighting and rereading (as we discuss in the sections on those strategies below). A trickier issue would concern implementing the strategy with students who are not skilled summarizers. Relatively intensive training programs are required for middle school students or learners with learning disabilities to benefit from summarization. Such efforts are not misplaced; training has been shown to benefit performance on a range of measures, although the training procedures do raise practical issues (e.g., Gajria & Salvia, 1992: 6.5–11 hours of training used for sixth through ninth graders with learning disabilities; Malone & Mastroppieri, 1991: 2 days of training used for middle school students with learning disabilities; Rinehart et al., 1986: 45–50 minutes of instruction per day for 5 days used for sixth graders). Of course, instructors may want students to summarize material because summarization itself is a goal, not because they plan to use summarization as a study technique, and that goal may merit the efforts of training.

However, if the goal is to use summarization as a study technique, our question is whether training students would be worth the amount of time it would take, both in terms of the time required on the part of the instructor and in terms of the time taken away from students’ other activities. For instance, in terms of efficacy, summarization tends to fall in the middle of the pack when compared to other techniques. In direct comparisons, it was sometimes more useful than rereading (Rewey, Dansereau, & Pecl, 1991) and was as useful as note-taking (e.g., Bretzing & Kulhavy, 1979) but was less powerful than generating explanations (e.g., Bednall & Kehoe, 2011) or self-questioning (A. King, 1992).

3.5 Summarization: Overall assessment. On the basis of the available evidence, we rate summarization as low utility. It can be an effective learning strategy for learners who are already skilled at summarizing; however, many learners (including children, high school students, and even some undergraduates) will require extensive training, which makes this strategy less feasible. Our enthusiasm is further dampened by mixed findings regarding which tasks summarization actually helps. Although summarization has been examined with a wide range of text materials, many researchers have pointed to factors of these texts that seem likely to moderate the effects of summarization (e.g., length), and future research should be aimed at investigating such factors. Finally, although many studies have examined summarization training in the classroom, what are lacking are classroom studies examining the effectiveness of summarization as a technique that boosts students’ learning, comprehension, and retention of course content.

4 Highlighting and underlining

Any educator who has examined students’ course materials is familiar with the sight of a marked-up, multicolored textbook. More systematic evaluations of actual textbooks and other student materials have supported the claim that highlighting and underlining are common behaviors (e.g., Bell & Limber, 2010; Lonka, Lindblom-Ylänne, & Maury, 1994; Nist & Kirby, 1989). When students themselves are asked about what they do when studying, they commonly report underlining, highlighting, or otherwise marking material as they try to learn it (e.g., Cioffi, 1986; Gurung, Weidert, & Jeske, 2010). We treat these techniques as equivalent, given that, conceptually, they should work the same way (and at least one study found no differences between them; Fowler & Barker, 1974, Experiment 2). The techniques typically appeal to students because they are simple to use, do not entail training, and do not require students to invest much time beyond what is already required for reading the material. The question we ask here is, will a technique that is so easy to use actually help students learn? To understand any benefits specific to highlighting and underlining (for brevity, henceforth referred to as highlighting), we do not consider studies in which active marking of text was paired with other common techniques, such as note-taking (e.g., Arnold, 1942; L. B. Brown & Smiley, 1978; Mathews, 1938). Although many students report combining multiple techniques (e.g., L. Annis & Davis, 1978; Wade, Trathen, & Schraw, 1990), each technique must be evaluated independently to discover which ones are crucial for success.
### 4.1 General description of highlighting and underlining and why they should work

As an introduction to the relevant issues, we begin with a description of a prototypical experiment. Fowler and Barker (1974, Exp. 1) had undergraduates read articles (totaling about 8,000 words) about boredom and city life from *Scientific American* and *Science*. Students were assigned to one of three groups: a control group, in which they only read the articles; an active-highlighting group, in which they were free to highlight as much of the texts as they wanted; or a passive-highlighting group, in which they read marked texts that had been highlighted by yoked participants in the active-highlighting group. Everyone received 1 hour to study the texts (time on task was equated across groups); students in the active-highlighting condition were told to mark particularly important material. All subjects returned to the lab 1 week later and were allowed to review their original materials for 10 minutes before taking a 54-item multiple-choice test. Overall, the highlighting groups did not outperform the control group on the final test, a result that has unfortunately been echoed in much of the literature (e.g., Hoon, 1974; Idstein & Jenkins, 1972; Stordahl & Christensen, 1956).

However, results from more detailed analyses of performance in the two highlighting groups are informative about what effects highlighting might have on cognitive processing. First, within the active-highlighting group, performance was better on text items for which the relevant text had been highlighted (see Blanchard & Mikkelsen, 1987; L. L. Johnson, 1988 for similar results). Second, this benefit to highlighted information was greater for the active highlighters (who selected what to highlight) than for passive highlighters (who saw the same information highlighted, but did not select it). Third, this benefit to highlighted information was accompanied by a small cost on test questions probing information that had not been highlighted.

To explain such findings, researchers often point to a basic cognitive phenomenon known as the isolation effect, whereby a semantically or phonologically unique item in a list is much better remembered than its less distinctive counterparts (see Hunt, 1995, for a description of this work). For instance, if students are studying a list of categorically related words (e.g., “desk,” “bed,” “chair,” “table”) and a word from a different category (e.g., “cow”) is presented, the students will later be more likely to recall it than they would if it had been studied in a list of categorically related words (e.g., “goat,” “pig,” “horse,” “chicken”). The analogy to highlighting is that a highlighted, underlined, or capitalized sentence will “pop out” of the text in the same way that the word “cow” would if it were isolated in a list of words for types of furniture. Consistent with this expectation, a number of studies have shown that reading marked text promotes later memory for the marked material: Students are more likely to remember things that the experimenter highlighted or underlined in the text (e.g., Cashen & Leicht, 1970; Crouse & Idstein, 1972; Hartley, Bartlett, & Branthwaite, 1980; Klare, Mabry, & Gustafson, 1955; see Lorch, 1989 for a review).

Actively selecting information should benefit memory more than simply reading marked text (given that the former would capitalize on the benefits of generation, Slamecka & Graf, 1978, and active processing more generally, Faw & Waller, 1976). Marked text draws the reader’s attention, but additional processing should be required if the reader has to decide which material is most important. Such decisions require the reader to think about the meaning of the text and how its different pieces relate to one another (i.e., organizational processing; Hunt & Worthen, 2006). In the Fowler and Barker (1974) experiment, this benefit was reflected in the greater advantage for highlighted information among active highlighters than among passive recipients of the same highlighted text. However, active highlighting is not always better than receiving material that has already been highlighted by an experimenter (e.g., Nist & Hogrebe, 1987), probably because experimenters will usually be better than students at highlighting the most important parts of a text.

More generally, the quality of the highlighting is likely crucial to whether it helps students to learn (e.g., Wollen, Cone, Britcher, & Mindeemann, 1985), but unfortunately, many studies have not contained any measure of the amount or the appropriateness of students’ highlighting. Those studies that have examined the amount of marked text have found great variability in what students actually mark, with some students marking almost nothing and others marking almost everything (e.g., Idstein & Jenkins, 1972). Some intriguing data came from the active-highlighting group in Fowler and Barker (1974). Test performance was negatively correlated ($r = -0.29$) with the amount of text that had been highlighted in the active-highlighting group, although this result was not significant given the small sample size ($n = 19$).

Marking too much text is likely to have multiple consequences. First, overmarking reduces the degree to which marked text is distinguished from other text, and people are less likely to remember marked text if it is not distinctive (Lorch, Lorch, & Klusewitz, 1995). Second, it likely takes less processing to mark a lot of text than to single out the most important details. Consistent with this latter idea, benefits of marking text may be more likely to be observed when experimenters impose explicit limits on the amount of text students are allowed to mark. For example, Rickards and August (1975) found that students limited to underlining a single sentence per paragraph later recalled more of a science text than did a non-underlining control group. Similarly, L. L. Johnson (1988) found that marking one sentence per paragraph helped college students in a reading class to remember the underlined information, although it did not translate into an overall benefit.

### 4.2 How general are the effects of highlighting and underlining?

We have outlined hypothetical mechanisms by which highlighting might aid memory, and particular features of highlighting that would be necessary for these mechanisms to be effective (e.g., highlighting only important material). However, most studies have shown no benefit of highlighting (as it
is typically used) over and above the benefit of simply reading, and thus the question concerning the generality of the benefits of highlighting is largely moot. Because the research on highlighting has not been particularly encouraging, few investigations have systematically evaluated the factors that might moderate the effectiveness of the technique—for instance, we could not include a Learning Conditions (4.2a) subsection below, given the lack of relevant evidence. To the extent the literature permits, we sketch out the conditions known to moderate the effectiveness of highlighting. We also describe how our conclusion about the relative ineffectiveness of this technique holds across a wide range of situations.

4.2b Student characteristics. Highlighting has failed to help Air Force basic trainees (Stordahl & Christensen, 1956), children (e.g., Rickards & Denner, 1979), and remedial students (i.e., students who scored an average of 390 on the SAT verbal section; Nist & Hogrobe, 1987), as well as prototypical undergraduates (e.g., Todd & Kessler, 1971). It is possible that these groups struggled to highlight only relevant text, given that other studies have suggested that most undergraduates overmark text. Results from one study with airmen suggested that prior knowledge might moderate the effectiveness of highlighting. In particular, the airmen read a passage on aircraft engines that either was unmarked (control condition) or had key information underlined (Klare et al., 1955). The experimenters had access to participants’ previously measured mechanical-aptitude scores and linked performance in the experiment to those scores. The marked text was more helpful to airmen who had received high scores. This study involved premarked texts and did not examine what participants would have underlined on their own, but it seems likely that students with little knowledge of a topic would struggle to identify which parts of a text were more or less important (and thus would benefit less from active highlighting than knowledgeable students would).

One other interesting possibility has come from a study in which experimenters extrinsically motivated participants by promising them that the top scorers on an exam would receive $5 (Fass & Schumacher, 1978). Participants read a text about enzymes; half the participants were told to underline key words and phrases. All participants then took a 15-item multiple-choice test. A benefit from underlining was observed among students who could earn the $5 bonus, but not among students in a control group. Thus, although results from this single study need to be replicated, it does appear that some students may have the ability to highlight effectively, but do not always do so.

4.2c Materials. Similar conclusions about marking text have come from studies using a variety of different text materials on topics as diverse as aerodynamics, ancient Greek schools, aggression, and Tanzania, ranging in length from a few hundred words to a few thousand. Todd and Kessler (1971) manipulated text length (all of the materials were relatively short, with lengths of 44, 140, or 256 words) and found that underlining was ineffective regardless of the text length. Fass and Schumacher (1978) manipulated whether a text about enzymes was easy or difficult to read; the easy version was at a seventh-grade reading level, whereas the difficult version was at high school level and contained longer sentences. A larger difference between the highlighting and control groups was found for performance on multiple-choice tests for the difficult text as opposed to the easy text.

4.2d Criterion tasks. A lack of benefit from highlighting has been observed on both immediate and delayed tests, with delays ranging from 1 week to 1 month. A variety of dependent measures have been examined, including free recall, factual multiple-choice questions, comprehension multiple-choice questions, and sentence-completion tests.

Perhaps most concerning are results from a study that suggested that underlining can be detrimental to later ability to make inferences. Peterson (1992) had education majors read a 10,000-word chapter from a history textbook; two groups underlined while studying for 90 minutes, whereas a third group was allowed only to read the chapter. One week later, all groups were permitted to review the material for 15 minutes prior to taking a test on it (the two underlining groups differed in whether they reviewed a clean copy of the original text or one containing their underlining). Everyone received the same test again 2 months later, without having another chance to review the text. The multiple-choice test consisted of 20 items that probed facts (and could be linked to specific references in the text) and 20 items that required inferences (which would have to be based on connections across the text and could not be linked to specific, underlined information). The three groups performed similarly on the factual questions, but students who had underlined (and reviewed their marked texts) were at a disadvantage on the inference questions. This pattern of results requires replication and extension, but one possible explanation for it is that standard underlining draws attention more to individual concepts (supporting memory for facts) than to connections across concepts (as required by the inference questions). Consistent with this idea, in another study, underliners who expected that a final test would be in a multiple-choice format scored higher on it than did underliners who expected it to be in a short-answer format (Kulhavy, Dyer, & Silver, 1975), regardless of the actual format of the final-test questions. Underlined information may naturally line up with the kinds of information students expect on multiple-choice tests (e.g., S. R. Schmidt, 1988), but students may be less sure about what to underline when studying for a short-answer test.

4.5 Effects in representative educational contexts. As alluded to at the beginning of this section, surveys of actual textbooks and other student materials have supported the frequency of highlighting and underlining in educational contexts (e.g., Bell & Limber, 2010; Lonka et al., 1994). Less clear are the consequences of such real-world behaviors. Classroom studies have examined whether instructor-provided markings affect examination performance. For example,
Cashen and Leicht (1970) had psychology students read *Scientific American* articles on animal learning, suicide, and group conflict, each of which contained five critical statements, which were underlined in red for half of the students. The articles were related to course content but were not covered in lectures. Exam scores on items related to the critical statements were higher when the statements had been underlined in red than when they had not. Interestingly, students in the underlining condition also scored better on exam questions about information that had been in sentences adjacent to the critical statements (as opposed to scoring worse on questions about nonunderlined information). The benefit to underlined items was replicated in another psychology class (Leicht & Cashen, 1972), although the effects were weaker. However, it is unclear whether the results from either of these studies would generalize to a situation in which students were in charge of their own highlighting, because they would likely mark many more than five statements in an article (and hence would show less discrimination between important and trivial information).

4.4 Issues for implementation. Students already are familiar with and spontaneously adopt the technique of highlighting; the problem is that the way the technique is typically implemented is not effective. Whereas the technique as it is typically used is not normally detrimental to learning (but see Peterson, 1992, for a possible exception), it may be problematic to the extent that it prevents students from engaging in other, more productive strategies.

One possibility that should be explored is whether students could be trained to highlight more effectively. We located three studies focused on training students to highlight. In two of these cases, training involved one or more sessions in which students practiced reading texts to look for main ideas before marking any text. Students received feedback about practice texts before marking (and being tested on) the target text, and training improved performance (e.g., Amer, 1994; Hayati & Shariatifar, 2009). In the third case, students received feedback on their ability to underline the most important content in a text; critically, students were instructed to underline as little as possible. In one condition, students even lost points for underlining extraneous material (Glover, Zimmer, Filbeck, & Plake, 1980). The training procedures in all three cases involved feedback, and they all had some safeguard against overuse of the technique. Given students’ enthusiasm for highlighting and underlining (or perhaps overenthusiasm, given that students do not always use the technique correctly), discovering fail-proof ways to ensure that this technique is used effectively might be easier than convincing students to abandon it entirely in favor of other techniques.

4.5 Highlighting and underlining: Overall assessment. On the basis of the available evidence, we rate highlighting and underlining as having low utility. In most situations that have been examined and with most participants, highlighting does little to boost performance. It may help when students have the knowledge needed to highlight more effectively, or when texts are difficult, but it may actually hurt performance on higher-level tasks that require inference making. Future research should be aimed at teaching students how to highlight effectively, given that students are likely to continue to use this popular technique despite its relative ineffectiveness.

5 The keyword mnemonic

Develop a mental image of students hunched over textbooks, struggling with a science unit on the solar system, trying to learn the planets’ names and their order in distance from the sun. Or imagine students in a class on language arts, reading a classic novel, trying to understand the motives of the main characters and how they may act later in the story. By visualizing these students in your “mind’s eye,” you are using one of the oldest strategies for enhancing learning—dating back to the ancient Greeks (Yates, 1966)—and arguably a powerful one: mental imagery. The earliest systematic research on imagery was begun in the late 1800s by Francis Galton for a historical review, see Thompson, 1990); since then, many debates have arisen about its nature (e.g., Kosslyn, 1981; Pylyshyn, 1981), such as whether its power accrues from the storage of dual codes (one imaginal and one propositional) or the storage of a distinctive propositional code (e.g., Marschark & Hunt, 1989), and whether mental imagery is subserved by the same brain mechanisms as visual imagery (e.g., Goldenberg, 1998).

Few of these debates have been entirely resolved, but fortunately, their resolution is not essential for capitalizing on the power of mental imagery. In particular, it is evident that the use of imagery can enhance learning and comprehension for a wide variety of materials and for students with various abilities. A review of this entire literature would likely go beyond a single monograph or perhaps even a book, given that mental imagery is one of the most highly investigated mental activities and has inspired enough empirical research to warrant its own publication (i.e., the *Journal of Mental Imagery*). Instead of an exhaustive review, we briefly discuss two specific uses of mental imagery for improving student learning that have been empirically scrutinized: the use of the keyword mnemonic for learning foreign-language vocabulary, and the use of mental imagery for comprehending and learning text materials.

5.1 General description of the keyword mnemonic and why it works. Imagine a student struggling to learn French vocabulary, including words such as la dent (tooth), la clef (key), revenir (to come back), and mourir (to die). To facilitate learning, the student uses the keyword mnemonic, which is a technique based on interactive imagery that was developed by Atkinson and Raugh (1975). To use this mnemonic, the student would first find an English word that sounds similar to the foreign cue word, such as dentist for “la dent” or cliff for
“la clef.” The student would then develop a mental image of the English keyword interacting with the English translation. So, for la dent—tooth, the student might imagine a dentist holding a large molar with a pair of pliers. Raugh and Atkinson (1975) had college students use the keyword mnemonic to learn Spanish-English vocabulary (e.g., gusano—worm): the students first learned to associate each experimenter-provided keyword with the appropriate Spanish cue (e.g., “gusano” is associated with the keyword “goose”), and then they developed interactive images to associate the keywords with their English translations. In a later test, the students were asked to generate the English translation when presented with the Spanish cue (e.g., “gusano”—tooth). Students who used the keyword mnemonic performed significantly better on the test than did a control group of students who studied the translation equivalents without keywords.

Beyond this first demonstration, the potential benefits of the keyword mnemonic have been extensively explored, and its power partly resides in the use of interactive images. In particular, the interactive image involves elaboration that integrates the words meaningfully, and the images themselves should help to distinguish the sought-after translation from other candidates. For instance, in the example above, the image of the “large molar” distinguishes “tooth” (the target) from other candidates relevant to dentists (e.g., gums, drills, floss). As we discuss next, the keyword mnemonic can be effectively used by students of different ages and abilities for a variety of materials. Nevertheless, our analysis of this literature also uncovered limitations of the keyword mnemonic that may constrain its utility for teachers and students. Given these limitations, we did not separate our review of the literature into separate sections that pertain to each variable category (Table 2) but instead provide a brief overview of the most relevant evidence concerning the generalizability of this technique.

5.2 a–d How general are the effects of the keyword mnemonic? The benefits of the keyword mnemonic generalize to many different kinds of material: (a) foreign-language vocabulary from a variety of languages (French, German, Italian, Latin, Russian, Spanish, and Tagalog); (b) the definitions of obscure English vocabulary words and science terms; (c) state-capital associations (e.g., Lincoln is the capital of Nebraska); (d) medical terminology; (e) people’s names and accomplishments or occupations; and (f) minerals and their attributes (e.g., the mineral wolframite is soft, dark in color, and used in the home). Equally impressive, the keyword mnemonic has also been shown to benefit learners of different ages (from second graders to college students) and students with learning disabilities (for a review, see Jitendra, Edwards, Sacks, & Jacobson, 2004). Although the bulk of research on the keyword mnemonic has focused on students’ retention of target materials, the technique has also been shown to improve students’ performance on a variety of transfer tasks: It helps them (a) to generate appropriate sentences using newly learned English vocabulary (McDaniel & Pressley, 1984) and (b) to adapt newly acquired vocabulary to semantically novel contexts (Mastropieri, Scruggs, & Mushinski Fulk, 1990).

The overwhelming evidence that the keyword mnemonic can boost memory for many kinds of material and learners has made it a relatively popular technique. Despite the impressive outcomes, however, some aspects of these demonstrations imply limits to the utility of the keyword mnemonic. First, consider the use of this technique for its originally intended domain—the learning of foreign-language vocabulary. In the example above, la dent easily supports the development of a concrete keyword (“dentist”) that can be easily imagined, whereas many vocabulary terms are much less amenable to the development and use of keywords. In the case of revenir (to come back), a student could perhaps use the keyword “revenge” (e.g., one might need “to come back” to taste its sweetness), but imaging this abstract term would be difficult and might even limit retention. Indeed, Hall (1988) found that a control group (which received task practice but no specific instructions on how to study) outperformed a keyword group in a test involving English definitions that did not easily afford keyword generation, even when the keywords were provided. Proponents of the keyword mnemonic do acknowledge that its benefits may be limited to keyword-friendly materials (e.g., concrete nouns), and in fact, the vast majority of the research on the keyword mnemonic has involved materials that afforded its use.

Second, in most studies, the keywords have been provided by the experimenters, and in some cases, the interactive images (in the form of pictures) were provided as well. Few studies have directly examined whether students can successfully generate their own keywords, and those that have have offered mixed results: Sometimes students’ self-generated keywords facilitate retention as well as experimenter-provided keywords do (Shapiro & Waters, 2005), and sometimes they do not (Shriberg, Levin, McCormick, & Pressley, 1982; Thomas & Wang, 1996). For more complex materials (e.g., targets with multiple attributes, as in the wolframite example above), the experimenter-provided “keywords” were pictures, which some students may have difficulties generating even after extensive training. Finally, young students who have difficulties generating images appear to benefit from the keyword mnemonic only if keywords and an associated interactive image (in the form of a picture) are supplied during learning (Pressley & Levin, 1978). Thus, although teachers who are willing to construct appropriate keywords may find this mnemonic useful, even these teachers (and students) would be able to use the technique only for subsets of target materials that are keyword friendly.

Third, and perhaps most disconcerting, the keyword mnemonic may not produce durable retention. Some of the studies investigating the long-term benefits of the keyword mnemonic included a test soon after practice as well as one after a longer delay of several days or even weeks (e.g., Condas, Marshall, & Miller, 1986; Raugh & Atkinson, 1975). These studies
generally demonstrated a benefit of keywords at the longer delay (for a review, see Wang, Thomas, & Ouellette, 1992). Unfortunately, these promising effects were compromised by the experimental designs. In particular, all items were tested on both the immediate and delayed tests. Given that the keyword mnemonic yielded better performance on the immediate tests, this initial increase in successful recall could have boosted performance on the delayed tests and thus inappropriately disadvantaged the control groups. Put differently, the advantage in delayed test performance could have been largely due to the effects of retrieval practice (i.e., from the immediate test) and not to the use of keyword mnemonics per se (because retrieval can slow forgetting; see the Practice Testing section below).

This possibility was supported by data from Wang et al. (1992; see also Wang & Thomas, 1995), who administered immediate and delayed tests to different groups of students. As shown in Figure 4 (top panel), for participants who received the immediate test, the keyword-mnemonic group outperformed a rote-repetition control group. By contrast, this benefit vanished for participants who received only the delayed test. Even more telling, as shown in the bottom panel of Figure 4, when the researchers equated the performance of the two groups on the immediate test (by giving the rote-repetition group more practice), performance on the delayed test was significantly better for the rote-repetition group than for the keyword-mnemonic group (Wang et al., 1992).

These data suggest that the keyword mnemonic leads to accelerated forgetting. One explanation for this surprising outcome concerns decoding at retrieval: Students must decode each image to retrieve the appropriate target, and at longer delays, such decoding may be particularly difficult. For instance, when a student retrieves “a dentist holding a large molar with a pair of pliers,” he or she may have difficulty deciding whether the target is “molar,” “tooth,” “pliers,” or “enamel.”

5.3 Effects in representative educational contexts. The keyword mnemonic has been implemented in classroom settings, and the outcomes have been mixed. On the promising side, Levin, Pressley, McCormick, Miller, and Shriberg (1979) had fifth graders use the keyword mnemonic to learn Spanish vocabulary words that were keyword friendly. Students were trained to use the mnemonic in small groups or as an entire class, and in both cases, the groups who used the keyword mnemonic performed substantially better than did control groups who were encouraged to use their own strategies while studying. Less promising are results for high school students who Levin et al. (1979) trained to use the keyword mnemonic. These students were enrolled in a 1st-year or 2nd-year language course, which is exactly the context in which one would expect the keyword mnemonic to help. However, the keyword mnemonic did not benefit recall, regardless of whether students were trained individually or in groups. Likewise, Willerman and Melvin (1979) did not find benefits of keyword-mnemonic training for college students enrolled in an elementary French course (cf. van Hell & Mahn, 1997; but see Lawson & Hogben, 1998).

5.4 Issues for implementation. The majority of research on the keyword mnemonic has involved at least some (and occasionally extensive) training, largely aimed at helping students develop interactive images and use them to subsequently retrieve targets. Beyond training, implementation also requires the development of keywords, whether by students, teachers, or textbook designers. The effort involved in generating some keywords may not be the most efficient use of time for students (or teachers), particularly given that at least one easy-to-use technique (i.e., retrieval practice, Fritz, Morris, Acton, Voelkel, & Etkind, 2007) benefits retention as much as the keyword mnemonic does.
5.5 The keyword mnemonic: Overall assessment. On the basis of the literature reviewed above, we rate the keyword mnemonic as low utility. We cannot recommend that the keyword mnemonic be widely adopted. It does show promise for keyword-friendly materials, but it is not highly efficient (in terms of time needed for training and keyword generation), and it may not produce durable learning. Moreover, it is not clear that students will consistently benefit from the keyword mnemonic when they have to generate keywords; additional research is needed to more fully explore the effectiveness of keyword generation (at all age levels) and whether doing so is an efficient use of students’ time, as compared to other strategies. In one head-to-head comparison, cued recall of foreign-language vocabulary was either no different after using the keyword mnemonic (with experimenter-provided keywords) than after practice testing, or was lower on delayed criterion tests 1 week later (Fritz, Morris, Acton, et al., 2007). Given that practice testing is easier to use and more broadly applicable (as reviewed below in the Practice Testing section), it seems superior to the keyword mnemonic.

6 Imagery use for text learning

6.1 General description of imagery use and why it should work. In one demonstration of the potential of imagery for enhancing text learning, Leutner, Leopold, and Sumfleth (2009) gave tenth graders 35 minutes to read a lengthy science text on the dipole character of water molecules. Students either were told to read the text for comprehension (control group) or were told to read the text and to mentally imagine the content of each paragraph using simple and clear mental images. Imagery instructions were also crossed with drawing: Some students were instructed to draw pictures that represented the content of each paragraph, and others did not draw. Soon after reading, the students took a multiple-choice test that included questions for which the correct answer was not directly available from the text but needed to be inferred from it. As shown in Figure 5, the instructions to mentally imagine the content of each paragraph significantly boosted the comprehension-test performance of students in the mental-imagery group, in comparison to students in the control group (Cohen’s $d = 0.72$). This effect is impressive, especially given that (a) training was not required, (b) the text involved complex science content, and (c) the criterion test required learners to make inferences about the content. Finally, drawing did not improve comprehension, and it actually negated the benefits of imagery instructions. The potential for another activity to interfere with the potency of imagery is discussed further in the subsection on learning conditions (6.2a) below.

A variety of mechanisms may contribute to the benefits of imaging text material on later test performance. Developing images can enhance one’s mental organization or integration of information in the text, and idiosyncratic images of particular referents in the text could enhance learning as well (cf. distinctive processing; Hunt, 2006). Moreover, using one’s prior knowledge to generate a coherent representation of a narrative may enhance a student’s general understanding of the text; if so, the influence of imagery use may be robust across criterion tasks that tap memory and comprehension. Despite these possibilities and the dramatic effect of imagery demonstrated by Leutner et al. (2009), our review of the literature suggests that the effects of using mental imagery to learn from text may be rather limited and not robust.

6.2 How general are the effects of imagery use for text learning? Investigations of imagery use for learning text materials have focused on single sentences and longer text materials. Evidence concerning the impact of imagery on sentence learning largely comes from investigations of other mnemonic techniques (e.g., elaborative interrogation) in which imagery instructions have been included in a comparison condition. This research has typically demonstrated that groups who receive imagery instructions have better memory for sentences than do no-instruction control groups (e.g., R. C. Anderson & Hirde, 1971; Wood, Pressley, & Winne, 1990). In the remainder of this section, we focus on the degree to which imagery instructions improve learning for longer text materials.

6.2a Learning conditions. Learning conditions play a potentially important role in moderating the benefits of imagery, so we briefly discuss two conditions here—namely, the modality of text presentation and learners’ actual use of imagery after receiving imagery instructions. Modality pertains to whether students are asked to use imagery as they read a text or as they listen to a narration of a text. L. R. Brooks (1967, 1968)
reported that participants’ visualization of a pathway through a matrix was disrupted when they had to read a description of it; by contrast, visualization was not disrupted when participants listened to the description. Thus, it is possible that the benefits of imagery are not fully actualized when students read text and would be most evident if they listened. Two observations are relevant to this possibility. First, the majority of imagery research has involved students reading texts; the fact that imagery benefits have sometimes been found indicates that reading does not entirely undermine imaginal processing. Second, in experiments in which participants either read or listened to a text, the results have been mixed. As expected, imagery has benefited performance more among students who have listened to texts than among students who have read them (De Beni & Moë, 2003; Levin & Divine-Hawkins, 1974), but in one case, imagery benefited performance similarly for both modalities in a sample of fourth graders (Maher & Sullivan, 1982).

The actual use of imagery as a learning technique should also be considered when evaluating the imagery literature. In particular, even if students are instructed to use imagery, they may not necessarily use it. For instance, R. C. Anderson and Kulhavy (1972) had high school seniors read a lengthy text passage about a fictitious primitive tribe; some students were told to generate images while reading, whereas others were told to read carefully. Imagery instructions did not influence performance, but reported use of imagery was significantly correlated with performance (see also Denis, 1982). The problem here is that some students who were instructed to use imagery did not, whereas some un instructed students spontaneously used it. Both circumstances would reduce the observed effect of imagery instructions, and students’ spontaneous use of imagery in control conditions may be partly responsible for the failure of imagery to benefit performance in some cases. Unfortunately, researchers have typically not measured imagery use, so evaluation of these possibilities must await further research.

6.2b Student characteristics. The efficacy of imagery instructions have been evaluated across a wide range of student ages and abilities. Consider data from studies involving fourth graders, given that this particular grade level has been popular in imagery research. In general, imagery instructions have tended to boost criterion performance for fourth graders, but even here the exceptions are noteworthy. For instance, imagery instructions boosted the immediate test performance of fourth graders who studied short (e.g., 12-sentence) stories that could be pictorially represented (e.g., Levin & Divine-Hawkins, 1974), but in some studies, this benefit was found only for students who were biased to use imagery or for skilled readers (Levin, Divine-Hawkins, Kerst, & Gutman, 1974). For reading longer narratives (e.g., narratives of 400 words or more), imagery instructions have significantly benefited fourth graders’ free recall of text material (Gambrell & Jatw, 1993; Rasco, Tennyson, & Boutwell, 1975; see also Lesgold, McCormick, & Golinkoff, 1975) and performance on multiple-choice questions about the text (Maher & Sullivan, 1982; this latter benefit was apparent for both high- and low-skilled readers), but even after extensive training and a reminder to use imagery, fourth graders’ performance on a standardized reading-comprehension test did not improve (Lesgold et al., 1975).

Despite the promise of imagery, this patchwork of inconsistent effects for fourth graders has also been found for studies of other ages. College students have been shown to reap the benefits of imagery, but these benefits depend on the nature of the criterion test (an issue we discuss below). In two studies, high school students who read a long passage did not benefit from imagery instructions (R. C. Anderson & Kulhavy, 1972; Rasco et al., 1975). Studies with fifth and sixth grade students have shown some benefits of imagery, but these trends have not all been significant (Kulhavy & Swenson, 1975) and did not arise on some criterion tests (e.g., standardized achievement tests; Micciniati, 1982). Third graders have been shown to benefit from using imagery (Oakhill & Patel, 1991; Pressley, 1976), but younger students do not appear to benefit from attempting to generate mental images while listening to a story (Gutman, Levin, & Pressley, 1977).

6.2c Materials. Similar to studies on the keyword mnemonic, investigations of imagery use for text learning have often used texts that are imagery friendly, such as narratives that can be visualized or short stories that include concrete terms. Across investigations, the specific texts have varied widely and include long passages (of 2,000 words or more; e.g., R. C. Anderson & Kulhavy, 1972; Giesen & Peccei, 1984), relatively short stories (e.g., L. K. S. Chan, Cole, & Morris, 1990; Maher & Sullivan, 1982), and brief 10-sentence passages (Levin & Divine-Hawkins, 1974; Levin et al., 1974). With regard to these variations in materials, the safest conclusion is that sometimes imagery instructions boost performance and sometimes they do not. The literature is filled with interactions whereby imagery helped for one kind of material but not for another kind of material. In these cases, failures to find an effect for any given kind of material may not be due to the material per se, but instead may reflect the effect of other, uncontrolled factors, making it is impossible to tell which (if any) characteristics of the materials predict whether imagery will be beneficial.

Fortunately, some investigators have manipulated the content of text materials when examining the benefits of imagery use. In De Beni and Moë (2003), one text included descriptions that were easy to imagine, another included a spatial description of a pathway that was easy to imagine and verbalize, and another was abstract and presumably not easy to imagine. As compared with instructions to just rehearse the texts, instructions to use imagery benefited free recall of the easy-to-imagine texts and the spatial texts but did not benefit recall of the abstract texts. Moreover, the benefits were evident only when students listened to the text, not when they read it (as discussed under “Learning Conditions,” 6.2a, above). Thus, the benefits of imagery may be largely constrained to texts that directly support imaginal representations.
Although the bulk of the research on imagery has used texts that were specifically chosen to support imagery, two studies have used the Metropolitan Achievement Test, which is a standardized test that taps comprehension. Both studies used extensive training in the use of imagery while reading, and both studies failed to find an effect of imagery training on test performance (Lesgold, et al., 1975; Miccinati, 1982), even when participants were explicitly instructed to use their trained skills to complete the test (Lesgold et al., 1975).

6.2d Criterion tasks. The inconsistent benefits of imagery within groups of students can in part be explained by interactions between imagery (vs. reading) instructions and the criterion task. Consider first the results from studies involving college students. When the criterion test comprises free-recall or short-answer questions tapping information explicitly stated in the text, college students tend to benefit from instructions to image (e.g., Gyeselincz, Meneghetti, De Beni, & Pazzaglia, 2009; Hodes, 1992; Rasco et al., 1975; although, as discussed earlier, these effects may be smaller when students read the passages rather than listen to them; De Beni & Moè, 2003). By contrast, despite the fact that imagery presumably helps students develop an integrated visual model of a text, imagery instructions did not significantly help college students answer questions that required them to make inferences based on information in a text (Giesen & Pecck, 1984) or comprehension questions about a passage on the human heart (Hodes, 1992).

This pattern is also apparent from studies with sixth graders, who do show significant benefits of imagery use on measures involving the recall or summarization of text information (e.g., Kulhavy & Swenson, 1975), but show reduced or nonexistent benefits on comprehension tests and on criterion tests that require application of the knowledge (Gagne & Memory, 1978; Miccinati, 1982). In general, imagery instructions tend not to enhance students' understanding or application of the content of a text. One study demonstrated that training improved 8- and 9-year-olds' performance on inference questions, but in this case, training was extensive (three sessions), which may not be practical in other settings.

When imagery instructions do improve criterion performance, a question arises as to whether these effects are long lasting. Unfortunately, the question of whether the use of imagery protects against the forgetting of text content has not been widely investigated; in the majority of studies, criterion tests have been administered immediately or shortly after the target material was studied. In one exception, Kulhavy and Swenson (1975) found that imagery instructions benefited fifth and sixth graders' accuracy in answering questions that tapped the gist of the texts, and this effect was even apparent 1 week after the texts were initially read. The degree to which these long-term benefits are robust and generalize across a variety of criterion tasks is an open question.

6.3 Effects in representative educational contexts. Many of the studies on imagery use and text learning have involved students from real classrooms who were reading texts that were written to match the students' grade level. Most studies have used fabricated materials, and few studies have used authentic texts that students would read. Exceptions have involved the use of a science text on the dipole character of water molecules (Leutner et al., 2009) and texts on cause-effect relationships that were taken from real science and social-science textbooks (Gagne & Memory, 1978); in both cases, imagery instructions improved test performance (although the benefits were limited to a free-recall test in the latter case). Whether instructions to use imagery will help students learn materials in a manner that will translate into improved course grades is unknown, and research investigating students' performance on achievement tests has shown imagery use to be a relatively inert strategy (Lesgold et al., 1975; Miccinati, 1982; but see Rose, Parks, Androe, & McMahon, 2000, who supplemented imagery by having students act out narrative stories).

6.4 Issues for implementation. The majority of studies have examined the influence of imagery by using relatively brief instructions that encouraged students to generate images of text content while studying. Given that imagery does not appear to undermine learning (and that it does boost performance in some conditions), teachers may consider instructing students (third grade and above) to attempt to use imagery when they are reading texts that easily lend themselves to imaginal representations. How much training would be required to ensure that students consistently and effectively use imagery under the appropriate conditions is unknown.

6.5 Imagery use for learning text: Overall assessment. Imagery can improve students' learning of text materials, and the promising work by Leutner et al. (2009) speaks to the potential utility of imagery use for text learning. Imagery production is also more broadly applicable than the keyword mnemonic. Nevertheless, the benefits of imagery are largely constrained to imagery-friendly materials and to texts of memory, and further demonstrations of the effectiveness of the technique (across different criterion tests and educationally relevant retention intervals) are needed. Accordingly, we rated the use of imagery for learning text as low utility.

7 Rereading

Rereading is one of the techniques that students most frequently report using during self-regulated study (Carrier, 2003; Hartwig & Dunlosky, 2012; Karpicke, Butler, & Roediger, 2009; Kornell & Bjork, 2007; Wissman, Rawson, & Pyc, 2012). For example, Carrier (2003) surveyed college students in an upper-division psychology course, and 65% reported using rereading as a technique when preparing for course exams. More recent surveys have reported similar results. Kornell and Bjork (2007) and Hartwig and Dunlosky (2012) asked students if they typically read a textbook, article, or
other source material more than once during study. Across these two studies, 18% of students reported rereading entire chapters, and another 62% reported rereading parts or sections of the material. Even high-performing students appear to use rereading regularly. Karpicke et al. (2009) asked undergraduates at an elite university (where students’ average SAT scores were above 1400) to list all of the techniques they used when studying and then to rank them in terms of frequency of use. Eighty-four percent of students included rereading textbook/notes in their list, and rereading was also the top-ranked technique (listed as the most frequently used technique by 55% of students). Students’ heavy reliance on rereading during self-regulated study raises an important question: Is rereading an effective technique?

7.1 General description of rereading and why it should work. In an early study by Rothkopf (1968), undergraduates read an expository text (either a 1,500-word passage about making leather or a 750-word passage about Australian history) zero, one, two, or four times. Reading was self-paced, and rereading was massed (i.e., each presentation of a text occurred immediately after the previous presentation). After a 10-minute delay, a cloze test was administered in which 10% of the content words were deleted from the text and students were to fill in the missing words. As shown in Figure 6, performance improved as a function of number of readings.

Why does rereading improve learning? Mayer (1983; Bormage & Mayer, 1986) outlined two basic accounts of rereading effects. According to the quantitative hypothesis, rereading simply increases the total amount of information encoded, regardless of the kind or level of information within the text. In contrast, the qualitative hypothesis assumes that rereading differentially affects the processing of higher-level and lower-level information within a text, with particular emphasis placed on the conceptual organization and processing of main ideas during rereading. To evaluate these hypotheses, several studies have examined free recall as a function of the kind or level of text information. The results have been somewhat mixed, but the evidence appears to favor the qualitative hypothesis. Although a few studies found that rereading produced similar improvements in the recall of main ideas and of details (a finding consistent with the quantitative hypothesis), several studies have reported greater improvement in the recall of main ideas than in the recall of details (e.g., Bormage & Mayer, 1986; Kiewra, Mayer, Christensen, Kim, & Risch, 1991; Rawson & Kintsch, 2005).

7.2 How general are the effects of rereading?

7.2a Learning conditions. Following the early work of Rothkopf (1968), subsequent research established that the effects of rereading are fairly robust across other variations in learning conditions. For example, rereading effects obtain regardless of whether learners are forewarned that they will be given the opportunity to study more than once, although Barnett and Scefeldt (1989) found a small but significant increase in the magnitude of the rereading effect among learners who were forewarned, relative to learners who were not forewarned. Furthermore, rereading effects obtain with both self-paced reading and experimenter-paced presentation. Although most studies have involved the silent reading of written material, effects of repeated presentations have also been shown when learners listen to an auditory presentation of text material (e.g., Bormage & Mayer, 1986; Mayer, 1983).

One aspect of the learning conditions that does significantly moderate the effects of rereading concerns the lag between initial reading and rereading. Although advantages of rereading over reading only once have been shown with massed rereading and with spaced rereading (in which some amount of time passes or intervening material is presented between initial study and restudy), spaced rereading usually outperforms massed rereading. However, the relative advantage of spaced reading over massed rereading may be moderated by the length of the retention interval, an issue that we discuss further in the subsection on criterion tasks below (7.2d). The effect of spaced rereading may also depend on the length of the lag between initial study and restudy. In a recent study by Verkoeijen, Rikers, and Özsoy (2008), learners read a lengthy expository text and then reread it immediately afterward, 4 days later, or 3.5 weeks later. Two days after rereading, all participants completed a final test. Performance was greater for the group who reread after a 4-day lag than for the massed rereaders, whereas performance for the group who reread after a 3.5-week lag was intermediate and did not significantly differ from performance in either of the other two groups. With that said, spaced rereading appears to be effective at least across

---

**Fig. 6.** Mean percentage of correct responses on a final cloze test for learners who read an expository text zero, one, two, or four times in Rothkopf (1968). Means shown are overall means for two conditions, one in which learners read a 1,500-word text and one in which learners read a 750-word text. Values are estimated from original figures in Rothkopf (1968). Standard errors are not available.
moderate lags, with studies reporting significant effects after lags of several minutes, 15–30 minutes, 2 days, and 1 week.

One other learning condition that merits mention is amount of practice, or dosage. Most of the benefits of rereading over a single reading appear to accrue from the second reading. The majority of studies that have involved two levels of rereading have shown diminishing returns from additional rereading trials. However, an important caveat is that all of these studies involved massed rereading. The extent to which additional spaced rereading trials produce meaningful gains in learning remains an open question.

Finally, although learners in most experiments have studied only one text, rereading effects have also been shown when learners are asked to study several texts, providing suggestive evidence that rereading effects can withstand interference from other learning materials.

7.2b Student characteristics. The extant literature is severely limited with respect to establishing the generality of rereading effects across different groups of learners. To our knowledge, all but two studies of rereading effects have involved undergraduate students. Concerning the two exceptions, Amlund, Kardash, and Kulhavy (1986) reported rereading effects with graduate students, and O’Shea, Sindeler, and O’Shea (1985) reported effects with third graders.

The extent to which rereading effects depend on knowledge level is also woefully unexplored. In the only study to date that has provided any evidence about the extent to which knowledge may moderate rereading effects (Arnold, 1942), both high-knowledge and low-knowledge readers showed an advantage of massed rereading over outlining or summarizing a passage for the same amount of time. Additional suggestive evidence that relevant background knowledge is not requisite for rereading effects has come from three recent studies that used the same text (Rawson, 2012; Rawson & Kintsch, 2005; Verkuilen et al., 2008) and found significant rereading effects for learners with virtually no specific prior knowledge about the main topics of the text (the charge of the Light Brigade in the Crimean War and the Hollywood film portraying the event).

Similarly, few studies have examined rereading effects as a function of ability, and the available evidence is somewhat mixed. Arnold (1942) found an advantage of massed rereading over outlining or summarizing a passage for the same amount of time among learners with both higher and lower levels of intelligence and both higher and lower levels of reading ability (but see Callender & McDaniel, 2009, who did not find an effect of massed rereading over single reading for either higher- or lower-ability readers). Raney (1993) reported a similar advantage of massed rereading over a single reading for readers with either higher or lower working-memory spans. Finally, Barnett and Seefeldt (1989) defined high- and low-ability groups by a median split of ACT scores; both groups showed an advantage of massed rereading over a single reading for short-answer factual questions, but only high-ability learners showed an effect for questions that required application of the information.

7.2c Materials. Rereading effects are robust across variations in the length and content of text material. Although most studies have used expository texts, rereading effects have also been shown for narratives. Those studies involving expository text material have used passages of considerably varying lengths, including short passages (e.g., 99–125 words), intermediate passages (e.g., 390–750 words), lengthy passages (e.g., 900–1,500 words), and textbook chapters or magazine articles with several thousand words. Additionally, a broad range of content domains and topics have been covered—an illustrative but nonexhaustive list includes physics (e.g., Ohm’s law), law (e.g., legal principles of evidence), history (e.g., the construction of the Brooklyn Bridge), technology (e.g., how a camera exposure meter works), biology (e.g., insects), geography (e.g., of Africa), and psychology (e.g., the treatment of mental disorders).

7.2d Criterion tasks. Across rereading studies, the most commonly used outcome measure has been free recall, which has consistently shown effects of both massed and spaced rereading with very few exceptions. Several studies have also shown rereading effects on cue-based recall measures, such as fill-in-the-blank tests and short-answer questions tapping factual information. In contrast, the effects of rereading on recognition are less certain, with weak or nonexistent effects on sentence-verification tasks and multiple-choice questions tapping factual information explicitly stated in the text (Callender & McDaniel, 2009; Dunlosky & Rawson, 2005; Hinze & Wiley, 2011; Kardash & Scholes, 1995). The evidence concerning the effects of rereading on comprehension is somewhat muddy. Although some studies have shown positive effects of rereading on answering problem-solving essay questions (Mayer, 1983) and short-answer application or inference questions (Karpicke & Blunt, 2011; Rawson & Kintsch, 2005), other studies using application or inference-based questions have reported effects only for higher-ability students (Barnett & Seefeldt, 1989) or no effects at all (Callender & McDaniel, 2009; Dunlosky & Rawson, 2005; Durgunoglu, Mir, & Ariño-Martí, 1993; Griffin, Wiley, & Thiede, 2008).

Concerning the durability of learning, most of the studies that have shown significant rereading effects have administered criterion tests within a few minutes after the final study trial, and most of these studies reported an advantage of massed rereading over a single reading. The effects of massed rereading after longer delays are somewhat mixed. Agarwal, Karpicke, Kang, Roediger, and McDermott (2008; see also Karpicke & Blunt, 2011) reported massed rereading effects after 1 week, but other studies have failed to find significant effects after 1–2 days (Callender & McDaniel, 2009; Cranney, Ahn, McKinnon, Morris, & Watts, 2009; Hinze & Wiley, 2011; Rawson & Kintsch, 2005).

Fewer studies have involved spaced rereading, although a relatively consistent advantage for spaced rereading over a single reading has been shown both on immediate tests and on tests administered after a 2-day delay. Regarding the comparison of massed rereading with spaced rereading, neither
schedule shows a consistent advantage on immediate tests. A similar number of studies have shown an advantage of spacing over massing, an advantage of massing over spacing, and no differences in performance. In contrast, spaced rereading consistently outperforms massed rereading on delayed tests. We explore the benefits of spacing more generally in the Distributed Practice section below.

7.3 Effects in representative educational contexts. Given that rereading is the study technique that students most commonly report using, it is perhaps ironic that no experimental research has assessed its impact on learning in educational contexts. Although many of the topics of the expository texts used in rereading research are arguably similar to those that students might encounter in a course, none of the aforementioned studies have involved materials taken from actual course content. Furthermore, none of the studies were administered in the context of a course, nor have any of the outcome measures involved course-related tests. The only available evidence involves correlational findings reported in survey studies, and it is mixed. Carrier (2003) found a nonsignificant negative association between self-reported rereading of textbook chapters and exam performance but a significantly positive association between self-reported review of lecture notes and exam performance. Hartwig and Dunlosky (2012) found a small but significant positive association between self-reported rereading of textbook chapters or notes and self-reported grade point average, even after controlling for self-reported use of other techniques.

7.4 Issues for implementation. One advantage of rereading is that students require no training to use it, other than perhaps being instructed that rereading is generally most effective when completed after a moderate delay rather than immediately after an initial reading. Additionally, relative to some other learning techniques, rereading is relatively economical with respect to time demands (e.g., in those studies permitting self-paced study, the amount of time spent rereading has typically been less than the amount of time spent during initial reading). However, in head-to-head comparisons of learning techniques, rereading has not fared well against some of the more effective techniques discussed here. For example, direct comparisons of rereading to elaborative interrogation, self-explanation, and practice testing (described in the Practice Testing section below) have consistently shown rereading to be an inferior technique for promoting learning.

7.5 Rereading: Overall assessment. Based on the available evidence, we rate rereading as having low utility. Although benefits from rereading have been shown across a relatively wide range of text materials, the generality of rereading effects across the other categories of variables in Table 2 has not been well established. Almost no research on rereading has involved learners younger than college-age students, and an insufficient amount of research has systematically examined the extent to which rereading effects depend on other student characteristics, such as knowledge or ability. Concerning criterion tasks, the effects of rereading do appear to be durable across at least modest delays when rereading is spaced. However, most effects have been shown with recall-based memory measures, whereas the benefit for comprehension is less clear. Finally, although rereading is relatively economical with respect to time demands and training requirements when compared with some other learning techniques, rereading is also typically much less effective. The relative disadvantage of rereading to other techniques is the largest strike against rereading and is the factor that weighed most heavily in our decision to assign it a rating of low utility.

8 Practice testing

Testing is likely viewed by many students as an undesirable necessity of education, and we suspect that most students would prefer to take as few tests as possible. This view of testing is understandable, given that most students’ experience with testing involves high-stakes summative assessments that are administered to evaluate learning. This view of testing is also unfortunate, because it overshadows the fact that testing also improves learning. Since the seminal study by Abbott (1909), more than 100 years of research has yielded several hundred experiments showing that practice testing enhances learning and retention (for recent reviews, see Rawson & Dunlosky, 2011; Roediger & Butler, 2011; Roediger, Putnam, & Smith, 2011). Even in 1906, Edward Thorndike recommended that “the active recall of a fact from within is, as a rule, better than its impression from without” (p. 123, Thorndike, 1906).

The century of research on practice testing since then has supported Thorndike’s recommendation by demonstrating the broad generalizability of the benefits of practice testing. Note that we use the term practice testing here (a) to distinguish testing that is completed as a low-stakes or no-stakes practice or learning activity outside of class from summative assessments that are administered by an instructor in class, and (b) to encompass any form of practice testing that students would be able to engage in on their own. For example, practice testing could involve practicing recall of target information via the use of actual or virtual flashcards, completing practice problems or questions included at the end of textbook chapters, or completing practice tests included in the electronic supplemental materials that increasingly accompany textbooks.

8.1 General description of practice testing and why it should work. As an illustrative example of the power of testing, Ranquist (1983) presented undergraduates with a list of word pairs for initial study. After a brief interval during which participants completed filler tasks, half of the pairs were tested via cued recall and half were not. Participants completed a final cued-recall test for all pairs either 10 minutes or 1 week later. Final-test performance was better for pairs that were practice tested than pairs that were not (53% versus 36% after
10 minutes, 35% versus 4% after 1 week). Whereas this study illustrates the method of comparing performance between conditions that do and do not involve a practice test, many other studies have compared a practice-testing condition with more stringent conditions involving additional presentations of the to-be-learned information. For example, Roediger and Karpicke (2006b) presented undergraduates with a short expository text for initial study followed either by a second study trial or by a practice free-recall test. One week later, free recall was considerably better among the group that had taken the practice test than among the group that had restudied (56% versus 42%). As another particularly compelling demonstration of the potency of testing as compared with restudy, Karpicke and Roediger (2008) presented undergraduates with Swahili-English translations for cycles of study and practice cued recall until items were correctly recalled once. After the first correct recall, items were presented only in subsequent study cycles with no further testing, or only in subsequent test cycles with no further study. Performance on a final test 1 week later was substantially greater after continued testing (80%) than after continued study (36%).

Why does practice testing improve learning? Whereas a wealth of studies have established the generality of testing effects, theories about why it improves learning have lagged behind. Nonetheless, theoretical accounts are increasingly emerging to explain two different kinds of testing effects, which are referred to as direct effects and mediated effects of testing (Roediger & Karpicke, 2006a). Direct effects refer to changes in learning that arise from the act of taking a test itself, whereas mediated effects refer to changes in learning that arise from an influence of testing on the amount or kind of encoding that takes place after the test (e.g., during a subsequent restudy opportunity).

Concerning direct effects of practice testing, Carpenter (2009) recently proposed that testing can enhance retention by triggering elaborative retrieval processes. Attempting to retrieve target information involves a search of long-term memory that activates related information, and this activated information may then be encoded along with the retrieved target, forming an elaborated trace that affords multiple pathways to facilitate later access to that information. In support of this account, Carpenter (2011) had learners study weakly related word pairs (e.g., "mother"—"child") followed either by additional study or a practice cued-recall test. On a later final test, recall of the target word was prompted via a previously unpresented but strongly related word (e.g., "father"). Performance was greater following a practice test than following restudy, presumably because the practice test increased the likelihood that the related information was activated and encoded along with the target during learning.

Concerning mediated effects of practice testing, Pyc and Rawson (2010, 2012b) proposed a similar account, according to which practice testing facilitates the encoding of more effective mediators (i.e., elaborative information connecting cues and targets) during subsequent restudy opportunities. Pyc and Rawson (2010) presented learners with Swahili-English translations in an initial study block, which was followed by three blocks of restudy trials; for half of the participants, each restudy trial was preceded by practice cued recall. All learners were prompted to generate and report a keyword mediator during each restudy trial. When tested 1 week later, compared with students who had only restudied, students who had engaged in practice cued recall were more likely to recall their mediators when prompted with the cue word and were more likely to recall the target when prompted with their mediator.

Recent evidence also suggests that practice testing may enhance how well students mentally organize information and how well they process idiosyncratic aspects of individual items, which together can support better retention and test performance (Hunt, 1995, 2006). Zaromb and Roediger (2010) presented learners with lists consisting of words from different taxonomic categories (e.g., vegetables, clothing) either for eight blocks of study trials or for four blocks of study trials with each trial followed by a practice free-recall test. Replicating basic testing effects, final free recall 2 days later was greater when items had received practice tests (39%) than when they had only been studied (17%). Importantly, the practice test condition also outperformed the study condition on secondary measures primarily tapping organizational processing and idiosyncratic processing.

8.2 How general are the effects of practice testing? Given the volume of research on testing effects, an exhaustive review of the literature is beyond the scope of this article. Accordingly, our synthesis below is primarily based on studies from the past 10 years (which include more than 120 articles), which we believe represent the current state of the field. Most of these studies compared conditions involving practice tests with conditions not involving practice tests or involving only restudy; however, we also considered more recent work pitting different practice-testing conditions against one another to explore when practice testing works best.

8.2a Learning conditions. The majority of research on practice testing has used test formats that involve cued recall of target information from memory, but some studies have also shown testing effects with other recall-based practice-test formats, including free recall, short-answer questions, and fill-in-the-blank questions. A growing number of studies using multiple-choice practice tests have also reported testing effects. Across these formats, most prior research has involved practice tests that tap memory for explicitly presented information. However, several studies have also shown testing effects for practice tests that tap comprehension, including short-answer application and multiple-choice inference-based questions (e.g., Agarwal & Roediger, 2011; Butler, 2010; C. I. Johnson & Mayer, 2009). Testing effects have also been shown in a study in which practice involved predicting (vs. studying) input-output values in an inductive function learning task (Kang, McDaniel, & Pashler, 2011) and a study in which participants practiced (vs. restudied) resuscitation procedures (Kromann,
Jensen, & Ringsted, 2009). Some research has demonstrated testing effects even when practice tests are open book (Agarwal et al., 2008; Weinstein, McDermott, & Roediger, 2010).

It is important to note that practice tests can benefit learning even when the format of the practice test does not match the format of the criterion test. For example, research has shown cross-format effects of multiple-choice practice tests on subsequent cued recall (Fazio, Agarwal, Marsh, & Roediger, 2010; Marsh, Agarwal, & Roediger, 2009; Roediger & Marsh, 2005), practice free recall on subsequent multiple-choice and short-answer inference tests (McDaniel, Howard, & Einstein, 2009), and practice cued recall on subsequent free recall and recognition (Carpenter, Pashler, & Vul, 2006; Vaughn & Rawson, 2011).

Although various practice-test formats work, some work better than others. Glover (1989) presented students with a short expository text for initial study and then manipulated the format of the practice test (free recall, fill in the blank, or recognition) and the format of the final test (free recall, fill in the blank, or recognition). On all three final-test formats, performance was greater following free-recall practice than following fill-in-the-blank practice, which in turn was greater than performance following recognition practice. Similarly, Carpenter and DeLosh (2006) found that free-recall practice outperformed cued-recall and recognition practice regardless of whether the final test was in a free-recall, cued-recall, or recognition format, and Hinze and Wiley (2011) found that performance on a multiple-choice final test was better following cued recall of paragraphs than following fill-in-the-blank practice. Further work is needed to support strong prescriptive conclusions, but the available evidence suggests that practice tests that require more generative responses (e.g., recall or short answer) are more effective than practice tests that require less generative responses (e.g., fill in the blank or recognition).

In addition to practice-test format, two other conditions of learning that strongly influence the benefits of practice testing are dosage and timing. Concerning dosage, the simplest conclusion is that more is better. Some studies supporting this conclusion have manipulated the number of practice tests, and final-test performance has consistently been better following multiple practice tests than the single practice test (e.g., Karpicke & Roediger, 2007a, 2010; Logan & Balota, 2008; Pavlik & Anderson, 2005). In other studies, experimenters have varied the number of practice tests to manipulate the level of success achieved during practice. For example, Vaughn and Rawson (2011) observed significantly greater final-test performance when students engaged in cued-recall practice until target items were recalled four to five times versus only once. Several other studies have shown that final-test performance improves as the number of correct responses during practice increases (e.g., Karpicke & Roediger, 2007b, 2008; Pyc & Rawson, 2009, 2012a; Rawson & Dunlosky, 2011), albeit with diminishing returns as higher criterion levels are achieved. Whereas these studies have involved manipulations of dosage within a practice session, other studies that have manipulated the number of practice sessions have also found that more is better (Bahrick, 1979; Bahrick, Bahrick, & Bahrick, 1993; Morris & Fritz, 2002; Rawson & Dunlosky, 2011).

However, the benefit of repeated practice testing in turn depends on the timing of the practice tests. Several studies have increased the number of tests presented in immediate succession within a session and have found minimal or nonexistent effects, in contrast to the sizable benefits observed when repeated tests are spaced (e.g., Carpenter & DeLosh, 2005; Cull, 2000; Glover, 1989; Karpicke & Bauernschmidt, 2011). Concerning the time intervals involved with spacing, longer is better. Repeated practice testing produces greater benefits when lags between trials within a session are longer rather than shorter (e.g., Pashler, Zarow, & Triplett, 2003; Pavlik & Anderson, 2005; Pyc & Rawson, 2009, 2012b), when trials are completed in different practice sessions rather than all in the same session (e.g., Bahrick, 1979; Bahrick & Hall, 2005; Kornell, 2009; Rohrer, 2009; Rohrer & Taylor, 2006), and when intervals between practice sessions are longer rather than shorter (Bahrick et al., 1993; Carpenter, Pashler, & Cepeda, 2009, although the optimal lag between sessions may depend on retention interval—see Cepeda et al., 2009; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008). We discuss lag effects further in the Distributed Practice section below.

**8.2b Student characteristics.** A large majority of studies have involved college students as participants, but testing effects have also been demonstrated across participants of widely varying ages. Studies involving nonundergraduate samples have differed somewhat in the kind, dosage, or timing of practice testing involved, but some form of testing effect has been demonstrated with preschoolers and kindergartners (Fritz, Morris, Nolan, & Singleton, 2007; Kratcohill, Demuth, & Conzemius, 1977), elementary school students (Atkinson & Paulson, 1972; Bouwmeester & Verkoeijen, 2011; Fishman, Keller, & Atkinson, 1968; Gates, 1917; Metcalfe & Kornell, 2007; Metcalfe, Kornell, & Finn, 2009; Myers, 1914; Rea & Modigliani, 1985; Rohrer, Taylor, & Sholar, 2010; Spitzer, 1939), middle school students (Carpenter et al., 2009; Fritz, Morris, Nolan, et al., 2007; Glover, 1989; McDaniell, Agarwal, Huehler, McDermott, & Roediger, 2011; Metcalfe, Kornell, & Son, 2007; Sones & Stroud, 1940), high school students (Duchastel, 1981; Duchastel & Nungester, 1982; Marsh et al., 2009; Nungester & Duchastel, 1982), and more advanced students, such as 3rd- and 4th-year medical-school students (Krohn et al., 2009; Rees, 1986; Schmidmaier et al., 2011). On the other end of the continuum, testing effects have also been shown with middle-aged learners and with older adults (Balota, Duchek, Sergent-Marshall, & Roediger, 2006; Bishara & Jacoby, 2008; Logan & Balota, 2008; Maddox, Balota, Coane, & Duchek, 2011; Sumowski, Chiaramavalli, & DeLuca, 2010; Tse, Balota, & Roediger, 2010).

In contrast to the relatively broad range of ages covered in the testing-effect literature, surprisingly minimal research has examined testing effects as a function of individual differences
in knowledge or ability. In the only study including groups of learners with different knowledge levels, Carroll, Campbell-Ratcliffe, Murnane, and Perfelt (2007) presented first-year undergraduates and advanced psychology majors with two passages from an abnormal-psychology textbook. Students completed a short-answer practice test on one of the passages and then took a final test over both passages either 15 minutes or 1 day later. Both groups showed similar testing effects at both time points (with 33% and 38% better accuracy, respectively, on the material that had been practice tested relative to the material that had not). Although these initial results provide encouraging evidence that testing effects may be robust across knowledge levels, further work is needed before strong conclusions can be drawn about the extent to which knowledge level moderates testing effects.

Likewise, minimal research has examined testing effects as a function of academically relevant ability levels. In a study by Spitzer (1939), 3,605 sixth graders from 91 different elementary schools read a short text and took an immediate test, to provide a baseline measure of reading comprehension ability. In the groups of interest here, all students read an experimental text, half completed a practice multiple-choice test, and then all completed a multiple-choice test either 1 or 7 days later. Spitzer reported final-test performance for the experimental text separately for the top and bottom thirds of performers on the baseline measure. As shown in Figure 7, taking the practice test benefited both groups of students. With that said, the testing effect appeared to be somewhat larger for higher-ability readers than for lower-ability readers (with approximately 20%, vs. 12%, improvements in accuracy), although Spitzer did not report the relevant inferential statistics.

Finally, evidence from studies involving patient populations is at least suggestive with respect to the generality of testing effects across different levels of learning capacity. For example, Balota et al. (2006) found that spaced practice tests improved retention over short time intervals not only for younger adults and healthy older adults but also for older adults with Alzheimer's disease. Similarly, Sumowski et al. (2010) found that a practice test produced larger testing effects for memory-impaired, versus memory-intact, subsets of middle-aged individuals with multiple sclerosis ($d = 0.95$ vs. $d = 0.54$, respectively, with grouping based on performance on a baseline measure of memory). In sum, several studies have suggested that practice testing may benefit individuals with varying levels of knowledge or ability, but the extent to which the magnitude of the benefit depends on these factors remains an open question.

8.2c Materials. Many of the studies that have demonstrated testing effects have involved relatively simple verbal materials, including word lists and paired associates. However, most of the sets of materials used have had some educational relevance. A sizable majority of studies using paired-associate materials have included foreign-language translations (including Chinese, Ilokpaq, Japanese, Lithuanian, Spanish, and Swahili) or vocabulary words paired with synonyms. Other studies have extended effects to paired book titles and author names, names and faces, objects and names, and pictures and foreign-language translations (e.g., Barcroft, 2007; Carpenter & Vul, 2011; Morris & Fritz, 2002; Rohrer, 2009).

A considerable number of studies have also shown testing effects for factual information, including trivia facts and general knowledge questions (e.g., Butler, Karpicek, & Roediger,
2008; T. A. Smith & Kimball, 2010) and facts drawn from classroom units in science, history, and psychology (e.g., Carpenter et al., 2009; McDaniel et al., 2011; McDaniel, Wildman, & Anderson, 2012). Earlier research showed that practice tests helped children learn multiplication facts and spelling lists (Atkinson & Paulson, 1972; Fishman et al., 1968; Rea & Modigliani, 1985), and recent studies have reported enhanced learning of definitions of vocabulary words (Metcalfe et al., 2007) and definitions of key term concepts from classroom material (Rawson & Dunlosky, 2011).

An increasing number of studies have shown benefits for learning from text materials of various lengths (from 160 words to 2,000 words or more), of various text genres (e.g., encyclopedia entries, scientific journal articles, textbook passages), and on a wide range of topics (e.g., Civil War economics, bat echolocation, sea otters, the big bang theory, fossils, Arctic exploration, toucans). Practice tests have improved learning from video lectures and from narrated animations on topics such as adult development, lighting, neuroanatomy, and art history (Butler & Roediger, 2007; Cranney et al., 2009; Vojdanoska, Cranney, & Newell, 2010).

Although much of the work on testing effects has used verbal materials, practice testing has also been shown to support learning of materials that include visual or spatial information, including learning of features and locations on maps (Carpenter & Pashler, 2007; Rohrer et al., 2010), identifying birds (Jacob, Wahlheim, & Coane, 2010), naming objects (Cepeda et al., 2009; Fritz et al., 2007), associating names with faces (Heldr & Shaughnessy, 2008; Morris & Fritz, 2002), learning spatial locations of objects (Sommer, Schoell, & Büchel, 2008), learning symbols (Coppens, Verkoeijen, & Rikers, 2011), and identifying depicted parts of a flower (Glover, 1989). Finally, recent work has extended testing effects to nondeclarative learning, including the learning of resuscitation skills (Kromann et al., 2009) and inductive learning of input-output functions (Kang, McDaniel, et al., 2011).

8.2d Criterion tasks. Although cued recall is the most commonly used criterion measure, testing effects have also been shown with other forms of memory tests, including free-recall, recognition, and fill-in-the-blank tests, as well as short-answer and multiple-choice questions that tap memory for information explicitly stated in test material.

Regarding transfer, the modal method in testing-effect research has involved using the same questions tapping the same target information (e.g., the same cued-recall prompts or multiple-choice questions) on practice tests and criterion tests. However, as described in the subsection on learning conditions (8.2a) above, many studies have also shown testing effects when learning of the same target information is evaluated using different test formats for practice and criterion tests. Furthermore, an increasing number of studies have shown that practice testing a subset of information influences memory for related but untested information (J. C. K. Chan, 2009, 2010; J. C. K. Chan, McDermott, & Roediger, 2006; Cranney et al., 2009), although benefits have not always accrued to related information (see Carroll et al., 2007; Duchastel, 1981).

Although most research has involved memory-based practice tests and criterion measures, several recent studies have also reported encouraging results concerning the extent to which practice testing can benefit comprehension. Positive effects have been shown on criterion tests that require inferences or the application of previously learned information (Agarwal & Roediger, 2011; Butler, 2010; Foos & Fisher, 1988; C. I. Johnson & Mayer, 2009; Karpicke & Blunt, 2011; McDaniel et al., 2009), including criterion tests that used different questions or different test formats than those used during practice. For example, Karpicke and Blunt (2011) found that practicing free recall of text material facilitated performance on a subsequent criterion test involving inference-based short-answer questions, as well as on a concept-mapping test. In fact, concept-mapping performance was better following free-recall practice during study than following concept mapping during study. Similarly, Butler (2010) presented students with expository texts for initial study, which was followed either by repeated restudy or by repeated practice short-answer tests (with feedback) tapping key facts and concepts from the texts. One week later, performance on new inference-based short-answer questions tapping the key facts and concepts was better following practice testing than following restudy (see Fig. 8). The outcomes of a follow-up experiment are particularly striking, given that the criterion test involved far transfer, in that questions required the concepts from one domain to be applied in a novel domain (e.g., students had to apply information learned about bat wings to make inferences about the development of new kinds of aircraft).

Finally, recent studies have also shown testing effects involving other forms of transfer. Jacoby et al. (2010)

![Fig. 8. Accuracy on final tests that consisted of inference-based transfer questions tapping key facts or concepts, administered 1 week after a learning session that involved either practice tests or restudy, in Butler (2010). Error bars represent standard errors.](image_url)
presented learners with pictures of birds and their family names for initial study, which was followed either by additional study of the picture-name pairs or by practice tests in which learners were shown each picture and attempted to retrieve the appropriate family name prior to being shown the correct answer. The subsequent criterion test involved the same families of birds but included new pictures of birds from those families. Learners were more accurate in classifying new birds following practice testing than following restudy only. Similarly, Kang, McDaniel, & Pashler (2011) examined inductive function learning under conditions in which learners either studied pairs of input-output values or predicted output for a given input value prior to being shown the correct output. The prediction group outperformed the study-only group on a criterion test for both trained pairs and untrained extrapolation pairs.

In addition to establishing testing effects across an array of outcome measures, studies have also demonstrated testing effects across many retention intervals. Indeed, in contrast to literatures on other learning techniques, contemporary research on testing effects has actually used short retention intervals less often than longer retention intervals. Although a fair number of studies have shown testing effects after short delays (0–20 minutes), the sizable majority of recent research has involved delays of at least 1 day, and the modal retention interval used is 1 week. The preference for using longer retention intervals may be due in part to outcomes from several studies reporting that testing effects are larger when final tests are administered after longer delays (J. C. K. Chan, 2009; Coppens et al., 2011; C. I. Johnson & Mayer, 2009; Kornell, Bjork, & Garcia, 2011; Roediger & Karpicke, 2006b; Runquist, 1983; Schmidmaier et al., 2011; Toppino & Cohen, 2009; Wenger, Thompson, & Bartling, 1980; Wheeler, Ewers, & Buonomo, 2003). It is impressive that testing effects have been observed after even longer intervals, including intervals of 2 to 4 weeks (e.g., Bahrick & Hall, 2005; Butler & Roediger, 2007; Carpenter, Pashler, Wixted, & Vul, 2008; Kromann et al., 2009; Rohrer, 2009), 2 to 4 months (e.g., McDaniel, Anderson, Derbish, & Morrisette, 2007; Morris & Fritz, 2002; Rawson & Dunlosky, 2011), 5 to 8 months (McDaniel et al., 2011; Rees, 1986), 9–11 months (Carpenter et al., 2009), and even 1 to 5 years (Bahrick et al., 1993). These findings are great news for students and educators, given that a key educational goal is durable knowledge and not just temporary improvements in learning.

8.3 Effects in representative educational contexts. As described above, much of the research on testing effects has involved educationally relevant materials, tasks, and retention intervals. Additionally, several studies have reported testing effects using authentic classroom materials (i.e., material taken from classes in which student participants were enrolled; Carpenter et al., 2009; Cranney et al., 2009; Kromann et al., 2009; McDaniel et al., 2007; Rawson & Dunlosky, 2011; Rees, 1986; Vojdanoska et al., 2010). Whereas the criterion measures in these studies involved experimenter-devised tests or no-stakes pop quizzes, research has also shown effects of practice testing on actual summative course assessments (Balch, 1998; Daniel & Brodria, 2004; Lyle & Crawford, 2011; McDaniel et al., 2011; McDaniel et al., 2012).

For example, a study by McDaniel et al. (2012) involved undergraduates enrolled in an online psychology course on the brain and behavior. Each week, students could earn course points by completing an online practice activity up to four times. In the online activity, some information was presented for practice testing with feedback, some information was presented for restudy, and some information was not presented. Subsequent unit exams included questions that had been presented during the practice tests and also new, related questions focusing on different aspects of the practiced concepts. As shown in Figure 9, grades on unit exams were higher for information that had been practice tested than for restudied information or unpracticed information, for both repeated questions and for new related questions.

8.4 Issues for implementation. Practice testing appears to be relatively reasonable with respect to time demands. Most research has shown effects of practice testing when the amount of time allotted for practice testing is modest and is equated with the time allotted for restudying. Another merit of practice testing is that it can be implemented with minimal training. Students can engage in recall-based self-testing in a relatively straightforward fashion. For example, students can self-test via cued recall by creating flashcards (free and low-cost flashcard software is also readily available) or by using the Cornell

![Fig. 9. Grades on course exams covering items that were presented for practice testing, presented for restudy, or not presented during online learning activities that students completed for course points. The course exam included some questions that had been presented during practice tests as well as new questions tapping the same information. For simplicity, outcomes reported here are collapsed across two experiments reported by McDaniel, Wildman, and Anderson (2012).](image-url)
note-taking system (which involves leaving a blank column when taking notes in class and entering key terms or questions in it shortly after taking notes to use for self-testing when reviewing notes at a later time; for more details, see Pauk & Ross, 2010). More structured forms of practice testing (e.g., multiple-choice, short-answer, and fill-in-the-blank tests) are often readily available to students via practice problems or questions included at the end of textbook chapters or in the electronic supplemental materials that accompany many textbooks. With that said, students would likely benefit from some basic instruction on how to most effectively use practice tests, given that the benefits of testing depend on the kind of test, dosage, and timing. As described above, practice testing is particularly advantageous when it involves retrieval and is continued until items are answered correctly more than once within and across practice sessions, and with longer as opposed to shorter intervals between trials or sessions.

Concerning the effectiveness of practice testing relative to other learning techniques, a few studies have shown benefits of practice testing over concept mapping, note-taking, and imagery use (Fritz et al., 2007; Karpicke & Blunt, 2011; McDaniel et al., 2009; Neuschatz, Preston, Toglia, & Neuschatz, 2005), but the most frequent comparisons have involved pitting practice testing against unguided restudy. The modal outcome is that practice testing outperforms restudying, although this effect depends somewhat on the extent to which practice tests are accompanied by feedback involving presentation of the correct answer. Although many studies have shown that testing alone outperforms restudy, some studies have failed to find this advantage (in most of these cases, accuracy on the practice test has been relatively low). In contrast, the advantage of practice testing with feedback over restudy is extremely robust. Practice testing with feedback also consistently outperforms practice testing alone.

Another reason to recommend the implementation of feedback with practice testing is that it protects against perseveration errors when students respond incorrectly on a practice test. For example, Butler and Roediger (2008) found that a multiple-choice practice test increased intrusions of false alternatives on a final cued-recall test when no feedback was provided, whereas no such increase was observed when feedback was given. Fortunately, the corrective effect of feedback does not require that it be presented immediately after the practice test. Metcalfe et al. (2009) found that final-test performance for initially incorrect responses was actually better when feedback had been delayed than when it had been immediate. Also encouraging is evidence suggesting that feedback is particularly effective for correcting high-confidence errors (e.g., Butterfield & Metcalfe, 2001). Finally, we note that the effects of practice-test errors on subsequent performance tend to be relatively small, often do not obtain, and are heavily outweighed by the positive benefits of testing (e.g., Fazio et al., 2010; Kang, Pashler, et al., 2011; Roediger & Marsh, 2005). Thus, potential concerns about errors do not constitute a serious issue for implementation, particularly when feedback is provided.

Finally, although we have focused on students’ use of practice testing, in keeping with the purpose of this monograph, we briefly note that instructors can also support student learning by increasing the use of low-stakes or no-stakes practice testing in the classroom. Several studies have also reported positive outcomes from administering summative assessments that are shorter and more frequent rather than longer and less frequent (e.g., one exam per week rather than only two or three exams per semester), not only for learning outcomes but also on students’ ratings of factors such as course satisfaction and preference for more frequent testing (e.g., Keys, 1934; Kika, McLaughlin, & Dixon, 1992; Leeming, 2002; for a review, see Bangert-Drowns, Kulik, & Kulik, 1991).

8.5 Practice testing: Overall assessment. On the basis of the evidence described above, we rate practice testing as having high utility. Testing effects have been demonstrated across an impressive range of practice-test formats, kinds of material, learner ages, outcome measures, and retention intervals. Thus, practice testing has broad applicability. Practice testing is not particularly time intensive relative to other techniques, and it can be implemented with minimal training. Finally, several studies have provided evidence for the efficacy of practice testing in representative educational contexts. Regarding recommendations for future research, one gap identified in the literature concerns the extent to which the benefits of practice testing depend on learners’ characteristics, such as prior knowledge or ability. Exploring individual differences in testing effects would align well with the aim to identify the broader generalizability of the benefits of practice testing. Moreover, research aimed at more thoroughly identify the causes of practice-test effects may provide further insights into maximizing these effects.

9 Distributed practice
To-be-learned material is often encountered on more than one occasion, such as when students review their notes and then later use flashcards to restudy the materials, or when a topic is covered in class and then later studied in a textbook. Even so, students mass much of their study prior to tests and believe that this popular cramming strategy is effective. Although cramming is better than not studying at all in the short term, given the same amount of time for study, would students be better off spreading out their study of content? The answer to this question is a resounding “yes.” The term distributed-practice effect refers to the finding that distributing learning over time (either within a single study session or across sessions) typically benefits long-term retention more than does massing learning opportunities back-to-back or in relatively close succession.
Given the volume of research on distributed practice, an exhaustive review of the literature is beyond the scope of this article. Fortunately, this area of research benefits from extensive review articles (e.g., Benjamin & Tullis, 2010; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Delaney, Vercelje, & Spiegel, 2010; Dempster & Farris, 1990; Donovan & Radesvich, 1999; Janiszewski, Noel, & Sawyer, 2003), which provided foundations for the current review. In keeping with recent reviews (Cepeda et al., 2006; Delaney et al., 2010), we use the term distributed practice to encompass both spacing effects (i.e., the advantage of spaced over massed practice) and lag effects (i.e., the advantage of spacing with longer lags over spacing with shorter lags), and we draw on both literatures for our summary.

9.1 General description of distributed practice and why it should work. To illustrate the issues involved, we begin with a description of a classic experiment on distributed practice, in which students learned translations of Spanish words to criterion in an original session (Bahrick, 1979). Students then participated in six additional sessions in which they had the chance to retrieve and relearn the translations (feedback was provided). Figure 10 presents results from this study. In the zero-spacing condition (represented by the circles in Fig. 10), the learning sessions were back-to-back, and learning was rapid across the six massed sessions. In the 1-day condition (represented by the squares in Fig. 10), learning sessions were spaced 1 day apart, resulting in slightly more forgetting across sessions (i.e., lower performance on the initial test in each session) than in the zero-spacing condition, but students in the 1-day condition still obtained almost perfect accuracy by the sixth session. In contrast, when learning sessions were separated by 30 days, forgetting was much greater across sessions, and initial test performance did not reach the level observed in the other two conditions, even after six sessions (see triangles in Fig. 10). The key point for our present purposes is that the pattern reversed on the final test 30 days later, such that the best retention of the translations was observed in the condition in which relearning sessions had been separated by 30 days. That is, the condition with the most intersession forgetting yielded the greatest long-term retention. Spaced practice (1 day or 30 days) was superior to massed practice (0 days), and the benefit was greater following a longer lag (30 days) than a shorter lag (1 day).

Many theories of distributed-practice effects have been proposed and tested. Consider some of the accounts currently under debate (for in-depth reviews, see Benjamin & Tullis, 2010; Cepeda et al., 2006). One theory invokes the idea of deficient processing, arguing that the processing of material during a second learning opportunity suffers when it is close in time to the original learning episode. Basically, students do not have to work very hard to reread notes or retrieve something from memory when they have just completed this same activity, and furthermore, they may be misled by the ease of this second task and think they know the material better than they really do (e.g., Bahrick & Hall, 2005). Another theory involves reminding; namely, the second presentation of to-be-learned material serves to remind the learner of the first learning opportunity, leading it to be retrieved, a process well known to enhance memory (see the Practice Testing section above). Some researchers also draw on consolidation in their explanations, positing that the second learning episode benefits from any consolidation of the first trace that has already happened. Given the relatively large magnitude of distributed-practice effects, it is plausible that multiple mechanisms may contribute to them; hence, particular theories often invoke different combinations of mechanisms to explain the effects.

9.2 How general are the effects of distributed practice? The distributed-practice effect is robust. Cepeda et al. (2006) reviewed 254 studies involving more than 14,000 participants altogether; overall, students recalled more after spaced study (47%) than after massed study (37%). In both Donovan and Radesvich’s (1999) and Janiszewski et al.’s (2003) meta-analyses, distributed practice was associated with moderate effect sizes for recall of verbal stimuli. As we describe below, the distributed-practice effect generalizes across many of the categories of variables listed in Table 2.

9.2a Learning conditions. Distributed practice refers to a particular schedule of learning episodes, as opposed to a particular kind of learning episode. That is, the distributed-practice effect refers to better learning when learning episodes are spread out in time than when they occur in close succession, but those learning episodes could involve restudying material, retrieving information from memory, or practicing skills. Because our emphasis is on educational applications, we will not
draw heavily on the skill literature, given that tasks such as ball tossing, gymnastics, and music memorization are less relevant to our purposes. Because much theory on the distributed-practice effect is derived from research on the spacing of study episodes, we focus on that research, but we also discuss relevant studies on distributed retrieval practice. In general, distributed practice testing is better than distributed study (e.g., Carpenter et al., 2009), as would be expected from the large literature on the benefits of practice testing.

One of the most important questions about distributed practice involves how to space the learning episodes—that is, how should the multiple encoding opportunities be arranged? Cepeda et al. (2006) noted that most studies have used relatively short intervals (less than 1 day), whereas we would expect the typical interval between educational learning opportunities (e.g., lecture and studying) to be longer. Recall that the classic investigation by Bahrick (1979) showed a larger distributed-practice effect with 30-day lags between sessions than with 1-day lags (Fig. 10); Cepeda et al. (2006) noted that “every study examined here with a retention interval longer than 1 month demonstrated a benefit from distribution of learning across weeks or months” (p. 370; “retention interval” here refers to the time between the last study opportunity and the final test).

However, the answer is not as simple as “longer lags are better”—the answer depends on how long the learner wants to retain information. Impressive data come from Cepeda, Vul, Rohrer, Wixted, and Pashler (2008), who examined people’s learning of trivia facts in an internet study that had 26 different conditions, which combined different between-session intervals (from no lag to a lag of 105 days) with different retention intervals (up to 350 days). In brief, criterion performance was best when the lag between sessions was approximately 10-20% of the desired retention interval. For example, to remember something for 1 week, learning episodes should be spaced 12 to 24 hours apart; to remember something for 5 years, the learning episodes should be spaced 6 to 12 months apart. Of course, when students are preparing for examinations, the degree to which they can space their study sessions may be limited, but the longest intervals (e.g., intervals of 1 month or more) may be ideal for studying core content that needs to be retained for cumulative examinations or achievement tests that assess the knowledge students have gained across several years of education.

Finally, the distributed-practice effect may depend on the type of processing evoked across learning episodes. In the meta-analysis by Janiszewski et al. (2003), intentional processing was associated with a larger effect size ($M = .35$) than was incidental processing ($M = .24$). Several things should be noted. First, the distributed-practice effect is sometimes observed with incidental processing (e.g., R. L. Greene, 1989; Toppino, Fearnow-Kenney, Kiepert, & Teremula, 2009); it is not eliminated across the board, but the average effect size is slightly (albeit significantly) smaller. Second, the type of processing learners engage in may covary with the intentionality of their learning, with students being more likely to extract meaning from materials when they are deliberately trying to learn them. In at least two studies, deeper processing yielded a distributed-practice effect whereas more shallow processing did not (e.g., Challis, 1993; Delaney & Knowles, 2005). Whereas understanding how distributed-practice effects change with strategy has important theoretical implications, this issue is less important when considering applications to education, because when students are studying, they presumably are intentionally trying to learn.

9.2b Student characteristics. The majority of distributed-practice experiments have tested undergraduates, but effects have also been demonstrated in other populations. In at least some situations, even clinical populations can benefit from distributed practice, including individuals with multiple sclerosis (Goverover, Hillyar, Chiavarotti, Arango-Lasprilla, & DeLuca, 2009), traumatic brain injuries (Goverover, Arango-Lasprilla, Hillyar, Chiavarotti, & DeLuca, 2009), and amnesia (Cermak, Verhaef, Lanzoni, Mather, & Chase, 1996). In general, children of all ages benefit from distributed study. For example, when learning pictures, children as young as preschoolers recognize and recall more items studied after longer lags than after shorter lags (Toppino, 1991; Toppino, Kasser, & Mracek, 1991). Similarly, 3-year-olds are better able to classify new exemplars of a category if the category was originally learned through spaced rather than massed study (Vlach, Sandhofer, & Kornell, 2008). Even 2-year-olds show benefits of distributed practice, such that it increases their later ability to produce studied words (Childers & Tomasello, 2002). These benefits of spacing for language learning also occur for children with specific language impairment (Riches, Tomasello, & Conti-Ramsden, 2005).

At the other end of the life span, older adults learning paired associates benefit from distributed practice as much as young adults do (e.g., Balota, Duchek, & Paullin, 1989). Similar conclusions are reached when spacing involves practice tests rather than study opportunities (e.g., Balota et al., 2006; Logan & Balota, 2008) and when older adults are learning to classify exemplars of a category (as opposed to paired associates; Kornell, Castel, Eich, & Bjork, 2010). In summary, learners of different ages benefit from distributed practice, but an open issue is the degree to which the distributed-practice effect may be moderated by other individual characteristics, such as prior knowledge and motivation.

9.2c Materials. Distributed-practice effects have been observed with many types of to-be-learned materials, including definitions (e.g., Dempster, 1987), face-name pairs (e.g., Carpenter & DeLosh, 2005), translations of foreign vocabulary words (e.g., Bahrick & Hall, 2005), trivia facts (e.g., Cepeda et al., 2008), texts (e.g., Rawson & Kintsch, 2005), lectures (e.g., Glover & Corkill, 1987), and pictures (e.g., Hintzman & Rogers, 1973). Distributed study has also yielded improved performance in a range of domains, including biology (Reynolds & Glaser, 1964) and advertising (e.g., Appleton-Knapp, Bjork, & Wickens, 2005). If we include practice testing and practice of
skills, then the list of domains in which benefits of distributed practice have been successfully demonstrated can be expanded to include mathematics (e.g., Rickard, Lau, & Pashler, 2008; Rohrer, 2009), history (Carpenter et al., 2009), music (e.g., Simmons, 2011), and surgery (e.g., Moulton et al., 2006), among others.

Not all tasks yield comparably large distributed-practice effects. For instance, distributed-practice effects are large for free recall but are smaller (or even nonexistent) for tasks that are very complex, such as airplane control (Donovan & Radesveich, 1999). It is not clear how to map these kinds of complex tasks, which tend to have a large motor component, onto the types of complex tasks seen in education. The U.S. Institute of Education Sciences guide on organizing study to improve learning explicitly notes that “one limitation of the literature is that few studies have examined acquisition of complex bodies of structured information” (Pashler et al., 2007, p. 6). The data that exist (which are reviewed below) come from classroom studies and are promising.

9.2d Criterion tasks. We alluded earlier to the fact that distributed-practice effects are robust over long retention intervals, with Cepeda and colleagues (2008) arguing that the ideal lag between practice sessions would be approximately 10–20% of the desired retention interval. They examined learning up to 350 days after study; other studies have shown benefits of distributed testing after intervals lasting for months (e.g., Cepeda et al., 2009) and even years (e.g., Bahrick et al., 1993; Bahrick & Phelps, 1987). In fact, the distributed-practice effect is often stronger on delayed tests than immediate ones, with massed practice (cramming) actually benefitting performance on immediate tests (e.g., Rawson & Kintsch, 2005).

Much research has established the durability of distributed-practice effects over time, but much less attention has been devoted to other kinds of criterion tasks used in educational contexts. The Cepeda et al. (2009) meta-analysis, for example, focused on studies in which the dependent measure was verbal free recall. The distributed-practice effect has been generalized to dependent measures beyond free recall, including multiple-choice questions, cued-recall and short-answer questions (e.g., Reynolds & Glaser, 1964), frequency judgments (e.g., Hintzman & Rogers, 1973), and, sometimes, implicit memory (e.g., R. L. Greene, 1990; Jacoby & Dallas, 1981). More generally, although studies using these basic measures of memory can inform the field by advancing theory, the effects of distributed practice on these measures will not necessarily generalize to other educationally relevant measures. Given that students are often expected to go beyond the basic retention of materials, this gap is perhaps the largest and most important to fill for the literature on distributed practice. With that said, some relevant data from classroom studies are available; we turn to these in the next section.

9.3 Effects in representative educational contexts. Most of the classroom studies that have demonstrated distributed-practice effects have involved spacing of more than just study opportunities. It is not surprising that real classroom exercises would use a variety of techniques, given that the goal of educators is to maximize learning rather than to isolate the contributions of individual techniques. Consider a study by Sobel, Cepeda, and Kaplan (2011) in which fifth graders learned vocabulary words. Each learning session had multiple steps: A teacher read and defined words; the students wrote down the definitions; the teacher repeated the definitions and used them in sentences, and students reread the definitions; finally, the students wrote down the definitions again and created sentences using the words. Several different kinds of study (including reading from booklets and overheads, as well as teacher instruction) and practice tests (e.g., generating definitions and sentences) were spaced in this research. The criterion test was administered 5 weeks after the second learning session, and students successfully defined a greater proportion of GRE vocabulary words (e.g., accolade) learned in sessions spaced a week apart than vocabulary words learned in sessions spaced a minute apart (Sobel et al., 2011). A mix of teacher instruction and student practice was also involved in a demonstration of the benefits of distributed practice for learning phonics in first graders (Seabrook, Brown, & Solity, 2005).

Another study examined learning of statistics across two sections of the same course, one of which was taught over a 6-month period and the other of which covered the same material in an 8-week period (Budé, Imbos, van de Wiel, & Berger, 2011). The authors took advantage of a curriculum change at their university that allowed them to compare learning in a class taught before the university reduced the length of the course with learning in a class taught after the change. The curriculum change meant that lectures, problem-based group meetings, and lab sessions (as well as student-driven study, assignments, etc.) were implemented within a much shorter time period; in other words, a variety of study and retrieval activities were more spaced out in time in one class than in the other. Students whose course lasted 6 months outperformed students in the 8-week course both on an open-ended test tapping conceptual understanding (see Fig. 11) and on the final exam (Fig. 12). Critically, the two groups performed similarly on a control exam from another course (Fig. 12), suggesting that the effects of distributed practice were not due to ability differences across classes.

Finally, a number of classroom studies have examined the benefits of distributed practice tests. Distributed practice testing helps students in actual classrooms learn history facts (Carpenter et al., 2009), foreign language vocabulary (K. C. Bloom & Shuell, 1981), and spelling (Fishman et al., 1968).

9.4 Issues for implementation. Several obstacles may arise when implementing distributed practice in the classroom. Dempster and Farris (1990) made the interesting point that many textbooks do not encourage distributed learning, in that they lump related material together and do not review previously covered material in subsequent units. At least one formal content analysis of actual textbooks (specifically, elementary-school mathematics textbooks; Stigler, Fuson, Ham, & Kim, 1986) supported this claim, showing that American textbooks
grouped to-be-worked problems together (presumably at the end of chapters) as opposed to distributing them throughout the pages. These textbooks also contained less variability in sets of problems than did comparable textbooks from the former Soviet Union. Thus, one issue students face is that their study materials may not be set up in a way that encourages distributed practice.

A second issue involves how students naturally study. Michael (1991) used the term procrastination scallop to describe the typical study pattern—namely, that time spent studying increases as an exam approaches. Mawhinney, Bostow, Laws, Blumenfield, and Hopkins (1971) documented this pattern using volunteers who agreed to study in an observation room that allowed their time spent studying to be recorded. With daily testing, students studied for a consistent amount of time across sessions. But when testing occurred only once every 3 weeks, time spent studying increased across the interval, peaking right before the exam (Mawhinney et al., 1971). In other words, less frequent testing led to massed study immediately before the test, whereas daily testing effectively led to study that was distributed over time. The implication is that students will not necessarily engage in distributed study unless the situation forces them to do so; it is unclear whether this is because of practical constraints or because students do not understand the memorial benefits of distributed practice.

With regard to the issue of whether students understand the benefits of distributed practice, the data are not entirely definitive. Several laboratory studies have investigated students’ choices about whether to mass or space repeated studying of paired associates (e.g., GRE vocabulary words paired with their definitions). In such studies, students typically choose between restudying an item almost immediately after learning (massing) or restudying the item later in the same session (spacing). Although students do choose to mass their study under some conditions (e.g., Benjamin & Bird, 2006; Son, 2004), they typically choose to space their study of items (Pyc & Dunlosky, 2010; Tippino, Cohen, Davis, & Moors, 2009). This bias toward spacing does not necessarily mean that students understand the benefits of distributed practice per se (e.g., they may put off restudying a pair because they do not want to see it again immediately), and one study has shown that students rate their overall level of learning as higher after massed study than after spaced study, even when the students had experienced the benefits of spacing (e.g., Kornell & Bjork, 2008). Other recent studies have provided evidence that students are unaware of the benefits of practicing with longer, as opposed to shorter, lags (Pyc & Rawson, 2012b; Wiseman et al., 2012).

In sum, because of practical constraints and students’ potential lack of awareness of the benefits of this technique, students may need some training and some convincing that distributed practice is a good way to learn and retain information. Simply experiencing the distributed-practice effect may not always be sufficient, but a demonstration paired with instruction about the effect may be more convincing to students (e.g., Balch, 2006).

9.5 Distributed practice: Overall assessment. On the basis of the available evidence, we rate distributed practice as having high utility: It works across students of different ages, with a wide variety of materials, on the majority of standard laboratory measures, and over long delays. It is easy to implement
(although it may require some training) and has been used successfully in a number of classroom studies. Although less research has examined distributed-practice effects using complex materials, the existing classroom studies have suggested that distributed practice should work for complex materials as well. Future research should examine this issue, as well as possible individual differences beyond age and criterion tasks that require higher-level cognition. Finally, future work should isolate the contributions of distributed study from those of distributed retrieval in educational contexts.

10 Interleaved practice

In virtually every kind of class at every grade level, students are expected to learn content from many different subtopics or problems of many different kinds. For example, students in a neuroanatomy course would learn about several different divisions of the nervous system, and students in a geometry course would learn various formulas for computing properties of objects such as surface area and volume. Given that the goal is to learn all of the material, how should a student schedule his or her studying of the different materials? An intuitive approach, and one we suspect is adopted by most students, involves blocking study or practice, such that all content from one subtopic is studied or all problems of one type are practiced before the student moves on to the next set of material. In contrast, recent research has begun to explore interleaved practice, in which students alternate their practice of different kinds of items or problems. Our focus here is on whether interleaved practice benefits students' learning of educationally relevant material.

Before we present evidence of the efficacy of this technique, we should point out that, in contrast to the other techniques we have reviewed in this monograph, many fewer studies have investigated the benefits of interleaved practice on measures relevant to student achievement. Nonetheless, we elected to include this technique in our review because (a) plenty of evidence indicates that interleaving can improve motor learning under some conditions (for reviews, see Brady, 1998; R. A. Schmidt & Bjork, 1992; Wulf & Shea, 2002) and (b) the growing literature on interleaving and performance on cognitive tasks is demonstrating the same kind of promise.

10.1 General description of interleaved practice and why it should work. Interleaved practice, as opposed to blocked practice, is easily understood by considering a method used by Rohrer and Taylor (2007), which involved teaching college students to compute the volumes of different geometric solids. Students had two practice sessions, which were separated by 1 week. During each practice session, students were given tutorials on how to find the volume for four different kinds of geometric solids and completed 16 practice problems (4 for each solid). After the completion of each practice problem, the correct solution was shown for 10 seconds. Students in a blocked-practice condition first read a tutorial on finding the volume of a given solid, which was immediately followed by the four practice problems for that kind of solid. Practice solving volumes for a given solid was then followed by the tutorial and practice problems for the next kind of solid, and so on. Students in an interleaved-practice group first read all four tutorials and then completed all the practice problems, with the constraint that every set of four consecutive problems included one problem for each of the four kinds of solids. One week after the second practice session, all students took a criterion test in which they solved two novel problems for each of the four kinds of solids. Students' percentages of correct responses during the practice sessions and during the criterion test are presented in Figure 13, which illustrates a typical interleaving effect: During practice, performance was better with blocked practice than interleaved practice, but this advantage dramatically reversed on the criterion test, such that interleaved practice boosted accuracy by 43%.

One explanation for this impressive effect is that interleaving gave students practice at identifying which solution method (i.e., which of several different formulas) should be used for a given solid (see also, Mayfield & Chase, 2002). Put differently, interleaved practice helps students to discriminate between the different kinds of problems so that they will be more likely to use the correct solution method for each one. Compelling evidence for this possibility was provided by Taylor and Rohrer (2010). Fourth graders learned to solve mathematical problems involving prisms. For a prism with a given number of base sides (b), students learned to solve for the number of faces (b + 2), edges (b × 3), corners (b × 2), or angles (b × 6). Students first practiced partial problems: A term for a single component of a prism was presented (e.g., corners), the student had to produce the correct formula (i.e., for corners, the correct response would be "b × 2"), and then

![Graph showing practice performance and test performance](image-url)

**Fig. 13.** Percentage of correct responses on sets of problems completed in practice sessions and on a delayed criterion test in Rohrer and Taylor (2007). Error bars represent standard errors.
feedback (the correct answer) was provided. After practicing partial problems, students practiced full problems, in which they were shown a prism with a number of base sides (e.g., 14 sides) and a term for a single component (e.g., edges). Students had to produce the correct formula \((b \times 3)\) and solve the problem by substituting the appropriate value of \(b\) \((14 \times 3)\). Most important, students in a blocked-practice group completed all partial- and full-practice problems for one prism feature (e.g., angles) before moving onto the next. For students in an interleaved-practice group, each block of four practice problems included one problem for each of the four prism features. One day after practice, a criterion test was administered in which students were asked to solve full problems that had not appeared during practice.

Accuracy during practice was greater for students who had received blocked practice than for students who had received interleaved practice, both for partial problems (99% vs. 68%, respectively) and for full problems (98% vs. 79%). By contrast, accuracy 1 day later was substantially higher for students who had received interleaved practice (77%) than for students who had received blocked practice (38%). As with Rohrer and Taylor (2006), a plausible explanation for this pattern is that interleaved practice helped students to discriminate between various kinds of problems and to learn the appropriate formula to apply for each one. This explanation was supported by a detailed analysis of errors the fourth graders made when solving the full problems during the criterion task. Fabrication errors involved cases in which students used a formula that was not originally trained (e.g., \(b \times 8\)), whereas discrimination errors involved cases in which students used one of the four formulas that had been practiced but was not appropriate for a given problem. As shown in Figure 14, the two groups did not differ in fabrication errors, but discrimination errors were more common after blocked practice than after interleaved practiced. Students who received interleaved practice apparently were better at discriminating among the kinds of problems and consistently applied the correct formula to each one.

How does interleaving produce these benefits? One explanation is that interleaved practice promotes organizational processing and item-specific processing because it allows students to more readily compare different kinds of problems. For instance, in Rohrer and Taylor (2007), it is possible that when students were solving for the volume of one kind of solid (e.g., a wedge) during interleaved practice, the solution method used for the immediately prior problem involving a different kind of solid (e.g., a spheroid) was still in working memory and hence encouraged a comparison of the two problems and their different formulas. Another possible explanation is based on the distributed retrieval from long-term memory that is afforded by interleaved practice. In particular, for blocked practice, the information relevant to completing a task (whether it be a solution to a problem or memory for a set of related items) should reside in working memory; hence, participants should not have to retrieve the solution. So, if a student completes a block of problems solving for volumes of wedges, the solution to each new problem will be readily available from working memory. By contrast, for interleaved practice, when the next type of problem is presented, the solution method for it must be retrieved from long-term memory. So, if a student has just solved for the volume of a wedge and then must solve for the volume of a spheroid, he or she must retrieve the formula for spheroids from memory. Such delayed practice testing would boost memory for the retrieved information (for details, see the Practice Testing section above). This retrieval-practice hypothesis and the discriminative-contrast hypothesis are not mutually exclusive, and other mechanisms may also contribute to the benefits of interleaved practice.

### 10.2 How general are the effects of interleaved practice?

#### 10.2a Learning conditions

Interleaved practice itself represents a learning condition, and it naturally covaries with distributed practice. For instance, if the practice trials for tasks of a given kind are blocked, the practice for the task is massed. By contrast, by interleaving practice across tasks of different kinds, any two instances of a task from a given set (e.g., solving for the volume of a given type of geometrical solid) would be separated by practice of instances from other tasks. Thus, at least some of the benefits of interleaved practice may reflect the benefits of distributed practice. However, some researchers have investigated the benefits of interleaved practice with spacing held constant (e.g., Kang & Pashler, 2012; Mitchell, Nash, & Hall, 2008), and the results suggested that spacing is not responsible for interleaving effects. For instance, Kang and Pashler (2012) had college students study paintings by various artists with the goal of developing a concept of each artists’ style, so that the students could later correctly identify the artists who had produced paintings that had not been

---

**Fig. 14.** Types of errors made by fourth graders while solving mathematical problems on a delayed criterion test in Taylor and Rohrer (2010). Error bars represent standard errors.
presented during practice. During practice, the presentation of paintings was either blocked by artist (e.g., all paintings by Jan Blencowe were presented first, followed by all paintings by Richard Lindenberg, and so on) or interleaved. Most important, a third group received blocked practice, but instead of viewing the paintings one right after another in a massed fashion, a cartoon drawing was presented in between the presentation of each painting (the cartoons were presented so that the temporal spacing in this spaced-block-practice group was the same as that for the interleaved group). Criterion performance was best after interleaved practice and was significantly better than after either standard or temporally spaced blocked practice. No differences occurred in performance between the two blocked-practice groups, which indicates that spacing alone will not consistently benefit concept formation.

This outcome is more consistent with the discriminative-contrast hypothesis than the retrieval-practice hypothesis. In particular, on each trial, the group receiving temporally spaced blocked practice presumably needed to retrieve (from long-term memory) what they had already learned about a painter's style, yet doing so did not boost their performance. That is, interleaved practice encouraged students to identify the critical differences among the various artists' styles, which in turn helped students discriminate among the artists' paintings on the criterion test. According to this hypothesis, interleaved practice may further enhance students' ability to develop accurate concepts (e.g., a concept of an artist's style) when exemplars of different concepts are presented simultaneously. For instance, instead of paintings being presented separately but in an interleaved fashion, a set of paintings could be presented at the same time. In this case, a student could more readily scan the paintings of the various artists to identify differences among them. Kang and Pashler (2012) found that simultaneous presentation of paintings from different artists yielded about the same level of criterion performance (68%) as standard interleaving did (65%), and that both types of interleaved practice were superior to blocked practice (58%; for a similar finding involving students learning to classify birds, see Wahlheim, Dunlosky, & Jacoby, 2011).

Finally, the amount of instruction and practice that students initially receive with each task may influence the degree to which interleaving all tasks enhances performance. In fact, in educational contexts, introducing a new concept or problem type (e.g., how to find the volume of a sphereoid) would naturally begin with initial instruction and blocked practice with that concept or problem type, and most of the studies reported in this section involved an introduction to all tasks before interleaving began. The question is how much initial practice is enough, and whether students with low skill levels (or students learning to solve more difficult tasks) will require more practice before interleaving begins. Given that skill level and task difficulty have been shown to moderate the benefits of interleaving in the literature on motor learning (e.g., Brady, 1998; Wulf & Shea, 2002), it seems likely that they do the same for cognitive tasks. If so, the dosage of initial instruction and blocked practice should interact with the benefits of interleaving, such that more pretraining should be required for younger and less skilled students, as well as for more complex tasks.

Consistent with this possibility are findings from Rau, Aleven, and Rummel (2010), who used various practice schedules to help teach fifth and sixth graders about fractions. During practice, students were presented with different ways to represent fractions, such as with pie charts, line segments, and set representations. Practice was either blocked (e.g., students worked with pie charts first, then line segments, and so on), interleaved, or first blocked and then interleaved. The prepractice and postpractice criterion tests involved fractions. Increases in accuracy from the prepractice test to the postpractice test occurred only after blocked and blocked-plus-interleaved practice (students in these two groups tended to perform similarly), and then, these benefits were largely shown only for students with low prior knowledge. This outcome provides partial support for the hypothesis that interleaved practice may be most beneficial only after a certain level of competency has been achieved using blocked practice with an individual concept or problem type.

10.2b Student characteristics. The majority of studies on interleaved practice have included college-aged students, and across these studies, sometimes interleaved practice has boosted performance, and sometimes it has not. Even so, differences in the effectiveness of interleaved practice for this age group are likely more relevant to the kind of task employed or, perhaps, to the dosage of practice, factors that we discuss in other sections. Some studies have included college students who were learning tasks relevant to their career goals—for instance, engineering students who were learning to diagnose system failures (e.g., de Croock, van Merrienboer, & Paas, 1998) and medical students who were learning to interpret electrocardiograms (Hatala, Brooks, & Norman, 2003). We highlight outcomes from these studies in the Materials subsection (10.2c) below. Finally, Mayfield and Chase (2002) conducted an extensive intervention to train algebra to college students with poor math skills; interleaving was largely successful, and we describe this experiment in detail in the Effects in Representative Educational Contexts subsection (10.3) below.

Concerning younger students, as reported above, Taylor and Rohrer (2010) reported that fourth graders benefited from interleaved practice when they were learning how to solve mathematical problems. In contrast, Rau et al. (2010) used various practice schedules to help teach fifth and sixth graders about fractions and found that interleaved practice did not boost performance. Finally, Olina, Reiser, Huang, Lim, and Park (2006) had high school students learn various rules for comma usage with interleaved or blocked practice; higher-skilled students appeared to be hurt by interleaving (although pretests scores favored those in the blocked group, and that advantage may have carried through to the criterion test), and interleaving did not help lower-skill students.
10.2c Materials. The benefits of interleaved practice have been explored using a variety of cognitive tasks and materials, from the simple (e.g., paired associate learning) to the relatively complex (e.g., diagnosing failures of a complicated piece of machinery). Outcomes have been mixed. Schneider, Healy, and Bourne (1998, 2002) had college students learn French vocabulary words from different categories, such as body parts, dinnerware, and foods. Across multiple studies, translation equivalents from the same category were blocked during practice or were interleaved. Immediately after practice, students who had received blocked practice recalled more translations than did students who had received interleaved practice (Schneider et al., 2002). One week after practice, correct recall was essentially the same in the blocked-practice group as in the interleaved-practice group. In another study (Schneider et al., 1998, Experiment 2), interleaved practice led to somewhat better performance than blocked practice on a delayed test, but this benefit was largely due to a slightly lower error rate. Based on these two studies, it does not appear that interleaved practice of vocabulary boosts retention.

More promising are results from studies that have investigated students' learning of mathematics. We have already described some of these studies above (Roehler & Taylor, 2007; Taylor & Roehler, 2010; but see Rau et al., 2010). Other math skills that have been trained include the use of Boolean functions (Carlson & Shin, 1996; Carlson & Yure, 1990) and algebraic skills (Mayfield & Chase, 2002). For the former, interleaved practice improved students' speed in solving multistep Boolean problems, especially when students could preview the entire multistep problem during solution (Carlson & Shin, 1996). For the latter, interleaving substantially boosted students' ability to solve novel algebra problems (as we discuss in detail below).

Van Merriënboer and colleagues (de Croock & van Merriënboer, 2007; de Croock et al., 1998; van Merriënboer, de Croock, & Jelsma, 1997; van Merriënboer, Schuurman, de Croock, & Paas, 2002) trained students to diagnose problems that occurred in a distiller system in which different components could fail; practice at diagnosing failures involving each component was either blocked or interleaved during practice. Across their studies, interleaved practice sometimes led to better performance on transfer tasks (which involved new combinations of system failures), but it did not always boost performance, leading the authors to suggest that perhaps more practice was needed to demonstrate the superiority of interleaved practice (de Croock & van Merriënboer, 2007). Blocked and interleaved practice have also been used to train students to make complex multidimensional judgments (Helsdingen, van Gog, & van Merriënboer, 2011a, 2011b), with results showing that decision making on criterion tests was better after interleaved than blocked practice. One impressive outcome was reported by Hatala et al. (2003), who trained medical students to make electrocardiogram diagnoses for myocardial infarction, ventricular hypertrophy, bundle branch blocks, and ischemia. The criterion test was on novel diagnoses, and accuracy was substantially greater after interleaved practice (47%) than after blocked practice (30%).

Finally, interleaved practice has been shown to improve the formation of concepts about artists' painting styles (Kang & Pashler, 2012; Kornell & Bjork, 2008) and about bird classifications (Wahland et al., 2011). The degree to which the benefits of interleaving improve concept formation across different kinds of concepts (and for students of different abilities) is currently unknown, but research and theory by Goldstone (1996) suggest that interleaving will not always be better. In particular, when exemplars within a category are dissimilar, blocking may be superior, because it will help learners identify what the members of a category have in common. By contrast, when exemplars from different categories are similar (as with the styles of artists and the classifications of birds used in the prior interleaving studies on concept formation), interleaving may work best because of discriminative contrast (e.g., Carvalho & Goldstone, 2011). These possibilities should be thoroughly explored with naturalistic materials before any general recommendations can be offered concerning the use of interleaved practice for concept formation.

10.2d Criterion tasks. In the literature on interleaving, the materials that are the focus of instruction and practice are used as the criterion task. Thus, if students practice solving problems of a certain kind, the criterion task will involve solving different versions of that kind of problem. For this reason, the current section largely reflects the analysis of the preceding section on materials (10.2c). One remaining issue, however, concerns the degree to which the benefits of interleaved practice are maintained across time. Although the delay between practice and criterion tests for many of the studies described above was minimal, several studies have used retention intervals as long as 1 to 2 weeks. In some of these cases, interleaved practice benefited performance (e.g., Mayfield & Chase, 2002; Roehler & Taylor, 2007), but in others, the potential benefits of interleaving did not manifest after the longer retention interval (e.g., de Croock & van Merriënboer, 2007; Rau et al., 2010). In the latter cases, interleaved practice may not have been potent at any retention interval. For instance, interleaved practice may not be potent for learning foreign-language vocabulary (Schneider et al., 1998) or for students who have not received enough practice with a complex task (de Croock & van Merriënboer, 2007).

10.3 Effects in representative educational contexts. It seems plausible that motivated students could easily use interleaving without help. Moreover, several studies have used procedures for instruction that could be used in the classroom (e.g., Hatala et al., 2003; Mayfield & Chase, 2002; Olina et al., 2006; Rau et al., 2010). We highlight one exemplary study here. Mayfield and Chase (2002) taught algebra rules to college students with poor math skills across 25 sessions. In different sessions, either a single algebra rule was introduced or previously introduced rules were reviewed. For review
sessions, either the rule learned in the immediately previous session was reviewed (which was analogous to blocking) or the rule learned in the previous session was reviewed along with the rules from earlier sessions (which was analogous to interleaved practice). Tests were administered prior to training, during the session after each review, and then 4 to 9 weeks after practice ended. On the tests, students had to apply the rules they had learned as well as solve problems by using novel combinations of the trained rules. The groups performed similarly at the beginning of training, but by the final tests, performance on both application and problem-solving items was substantially better for the interleaved group, and these benefits were still evident (albeit no longer statistically significant) on the delayed retention test.

10.4 Issues for implementation. Not only is the result from Mayfield and Chase (2002) promising, their procedure offers a tactic for the implementation of interleaved practice, both by teachers in the classroom and by students regulating their study (for a detailed discussion of implementation, see Rohrer, 2009). In particular, after a given kind of problem (or topic) has been introduced, practice should first focus on that particular problem. After the next kind of problem is introduced (e.g., during another lecture or study session), that problem should first be practiced, but it should be followed by extra practice that involves interleaving the current type of problem with others introduced during previous sessions. As each new type of problem is introduced, practice should be interleaved with practice for problems from other sessions that students will be expected to discriminate between (e.g., if the criterion test will involve a mixture of several types of problems, then these should be practiced in an interleaved manner during class or study sessions). Interleaved practice may take a bit more time to use than blocked practice, because solution times often slow during interleaved practice; even so, such slowing likely indicates the recruitment of other processes—such as discriminative contrast—that boost performance. Thus, teachers and students could integrate interleaved practice into their schedules without too much modification.

10.5 Interleaved practice: Overall recommendations. On the basis of the available evidence, we rate interleaved practice as having moderate utility. On the positive side, interleaved practice has been shown to have relatively dramatic effects on students’ learning and retention of mathematical skills, and teachers and students should consider adopting it in the appropriate contexts. Also, interleaving does help (and rarely hinders) other kinds of cognitive skills. On the negative side, the literature on interleaved practice is currently small, but it contains enough null effects to raise concern. Although the null effects may indicate that the technique does not consistently work well, they may instead reflect that we do not fully understand the mechanisms underlying the effects of interleaving and therefore do not always use it appropriately. For instance, in some cases, students may not have had enough instruction or practice with individual tasks to reap the benefits of interleaved practice. Given the promise of interleaved practice for improving student achievement, there is a great need for research that systematically evaluates how its benefits are moderated by dosage during training, student abilities, and the difficulty of materials.

Closing Remarks

Relative utility of the learning techniques

Our goal was to provide reviews that were extensive enough to allow anyone interested in using a particular technique to judge its utility for his or her own instructional or learning goals. We also realized that offering some general ratings (and the reasons behind them) might be useful to readers interested in quickly obtaining an overview on what technique may work best. To do so, we have provided an assessment of how each technique fared with respect to the generalizability of its benefits across the four categories of variables listed in Table 2, issues for implementation, and evidence for its effectiveness from work in representative educational contexts (see Table 4). Our goal for these assessments was to indicate both (a) whether sufficient evidence is available to support conclusions about the generalizability of a technique, issues for its implementation, or its efficacy in educational contexts, and, if sufficient evidence does exist, (b) whether it indicates that the technique works. For instance, practice testing received an assessment of Positive (P) for criterion tasks; this rating indicates that we found enough evidence to conclude that practice testing benefits students across a wide range of criterion tasks and retention intervals. Of course, it does not mean that further work in this area (i.e., testing with different criterion tasks) would not be valuable, but the extent of the evidence is promising enough to recommend it to teachers and students.

A Negative (N) rating indicates that the available evidence shows that the learning technique does not benefit performance for the particular category or issue. For instance, despite its popularity, highlighting did not boost performance across a variety of criterion tasks, so it received a rating of N for this variable.

A Qualified (Q) rating indicates that both positive and negative evidence has been reported with respect to a particular category or issue. For instance, the keyword mnemonic received a Q rating for materials, because evidence indicates that this technique does work for learning materials that are imagery friendly but does not work well for materials that cannot be easily imagined.

A rating of Insufficient (I) indicates that insufficient evidence is available to draw conclusions about the effects of a given technique for a particular category or issue. For instance, elaborative interrogation received an I rating for criterion tasks because we currently do not know whether its effects are durable across educationally relevant retention intervals. Any cell
Table 4. Utility Assessment and Ratings of Generalizability for Each of the Learning Techniques

<table>
<thead>
<tr>
<th>Technique</th>
<th>Utility</th>
<th>Learners</th>
<th>Materials</th>
<th>Criterion tasks</th>
<th>Issues for implementation</th>
<th>Educational contexts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elaborative interrogation</td>
<td>Moderate</td>
<td>P-I</td>
<td>P</td>
<td>I</td>
<td>P</td>
<td>I</td>
</tr>
<tr>
<td>Self-explanation</td>
<td>Moderate</td>
<td>P-I</td>
<td>P</td>
<td>P-I</td>
<td>Q</td>
<td>1</td>
</tr>
<tr>
<td>Summarization</td>
<td>Low</td>
<td>Q</td>
<td>P-I</td>
<td>Q</td>
<td>Q</td>
<td>Q-I</td>
</tr>
<tr>
<td>Highlighting</td>
<td>Low</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>P</td>
<td>N</td>
</tr>
<tr>
<td>The keyword mnemonic</td>
<td>Low</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q-I</td>
</tr>
<tr>
<td>Imagery use for text learning</td>
<td>Low</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>P</td>
<td>I</td>
</tr>
<tr>
<td>Rereading</td>
<td>Low</td>
<td>I</td>
<td>P</td>
<td>Q-I</td>
<td>P</td>
<td>1</td>
</tr>
<tr>
<td>Practice testing</td>
<td>High</td>
<td>P-I</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P-P-I</td>
</tr>
<tr>
<td>Distributed practice</td>
<td>High</td>
<td>P-I</td>
<td>P</td>
<td>P-I</td>
<td>P-P-I</td>
<td>P-P-I</td>
</tr>
<tr>
<td>Interleaved practice</td>
<td>Moderate</td>
<td>I</td>
<td>Q</td>
<td>P-I</td>
<td>P-P-I</td>
<td>P-P-I</td>
</tr>
</tbody>
</table>

Note: A positive (P) rating indicates that available evidence demonstrates efficacy of a learning technique with respect to a given variable or issue. A negative (N) rating indicates that a technique is largely ineffective for a given variable. A qualified (Q) rating indicates that the technique yielded positive effects under some conditions (or in some groups) but not others. An insufficient (I) rating indicates that there is insufficient evidence to support a definitive assessment for one or more factors for a given variable or issue.

in Table 4 with an I rating highlights the need for further systematic research.

Finally, some cells include more than one rating. In these cases, enough evidence exists to evaluate a technique on one dimension of a category or issue, yet insufficient evidence is available for some other dimension. For instance, self-explanation received a P-I rating for criterion tasks because the available evidence is positive on one dimension (generalizability across a range of criterion tasks) but is insufficient on another key dimension (whether the benefit of self-explanation generalizes across longer retention intervals). As another example, rereading received a Q-I rating for criterion tasks because evidence for the effectiveness of this technique over long retention intervals is qualified (i.e., under some learning conditions, it does not produce an effect for longer retention intervals), and insufficient evidence is available that is relevant to its effectiveness across different kinds of criterion tasks (e.g., rereading does boost performance on recall tasks, but little is known as to its benefits for comprehension). When techniques have multiple ratings for one or more variables, readers will need to consult the reviews for details.

Finally, we used these ratings to develop an overall utility assessment for each of the learning techniques. The utility assessments reflect how the benefits of each learning technique generalize across the different categories of variables (e.g., for how many variables the technique received a P rating). For example, the keyword mnemonic and imagery use for text learning were rated low in utility in part because their effects are limited to materials that are amenable to imagery and because they may not work well for students of all ages. Even so, some teachers may decide that the benefits of techniques with low-utility ratings match their instructional goals for their students. Thus, although we do offer these easy-to-use assessments of each learning technique, we also encourage interested teachers and students to carefully read each review to make informed decisions about which techniques will best meet their instructional and learning goals.

Implications for research on learning techniques

A main goal of this monograph was to develop evidence-based recommendations for teachers and students about the relative utility of various learning techniques. A related goal was to identify areas that have been underinvestigated and that will require further research before evidence-based recommendations for their use in education can be made. A number of these gaps are immediately apparent upon inspection of Table 4. To highlight a few, we do not yet know the extent to which many of the learning techniques will benefit students of various ages, abilities, and levels of prior knowledge. Likewise, with a few exceptions (e.g., practice testing and distributed practice), the degree to which many of the techniques support durable learning (e.g., over a number of weeks) is largely unknown, partly because investigations of these techniques have typically involved a single session that included both practice and criterion tests (for a discussion of the limitations of such single-session research, see Rawson & Dunlosky, 2011). Finally, few techniques have been evaluated in representative educational contexts.

This appraisal (along with Table 4) suggests two directions for future research that could have immediate implications for education. First, more research is needed to fully explore the degree to which the benefits of some techniques generalize to the variables listed in Table 2. Particularly important will be investigations that evaluate the degree to which interactions among the variables limit or magnify the benefits of a given technique. Second, the benefit of most of the techniques in representative educational settings needs to be more fully explored. Easy-to-use versions of the most promising tech-
niques should be developed and evaluated in controlled investigations conducted in educationally representative contexts. Ideally, the criterion measures would include high-stakes tests, such as performance on in-class exams and on achievement tests. We realize that such research efforts can be time-consuming and costly, but conducting them will be crucial for recommending educational changes that will have a reasonable likelihood of improving student learning and achievement.

**Implications for students, teachers, and student achievement**

Pressley and colleagues (Pressley, 1986; Pressley, Goodchild, et al., 1989) developed a good-strategy-user model, according to which being a sophisticated strategy user involves "knowing the techniques that accomplish important life goals (i.e., strategies), knowing when and how to use those methods . . . and using those methods in combination with a rich network of nonstrategic knowledge that one possesses about the world" (p. 302). However, Pressley, Goodchild, et al. (1989) also noted that "many students are committed to ineffective strategies . . . moreover, there is not enough professional evaluation of techniques that are recommended in the literature, with many strategies oversold by proponents" (p. 301). We agree and hope that the current reviews will have a positive impact with respect to fostering further scientific evaluation of the techniques.

Concerning students' commitment to ineffective strategies, recent surveys have indicated that students most often endorse the use of rereading and highlighting, two strategies that we found to have relatively low utility. Nevertheless, some students do report using practice testing, and these students appear to benefit from its use. For instance, Gurung (2005) had college students describe the strategies they used in preparing for classroom examinations in an introductory psychology course. The frequency of students' reported use of practice testing was significantly correlated with their performance on a final exam (see also Hartwig & Dunlosky, 2012). Given that practice testing is relatively easy to use, students who do not currently use this technique should be able to incorporate it into their study routine.

Why don't many students consistently use effective techniques? One possibility is that students are not instructed about which techniques are effective or how to use them effectively during formal schooling. Part of the problem may be that teachers themselves are not told about the efficacy of various learning techniques. Given that teachers would most likely learn about these techniques in classes on educational psychology, it is revealing that most of the techniques do not receive sufficient coverage in educational-psychology textbooks. We surveyed six textbooks (cited in the Introduction), and, except for mnemonics based on imagery (e.g., the keyword mnemonic), none of the techniques was covered by all of the books. Moreover, in the subset of textbooks that did describe one or more of these techniques, the coverage in most cases was relatively minimal, with a brief description of a given technique and relatively little guidance on its use, effectiveness, and limitations. Thus, many teachers are unlikely getting a sufficient introduction to which techniques work best and how to train students to use them.

A second problem may be that a premium is placed on teaching students content and critical-thinking skills, whereas less time is spent teaching students to develop effective techniques and strategies to guide learning. As noted by McNamara (2010), "there is an overwhelming assumption in our educational system that the most important thing to deliver to students is content" (p. 341, italics in original). One concern here is that students who do well in earlier grades, in which learning is largely supervised, may struggle later, when they are expected to regulate much of their own learning, such as in high school or college. Teaching students to use these techniques would not take much time away from teaching content and would likely be most beneficial if the use of the techniques was consistently taught across multiple content areas, so that students could broadly experience their effects on learning and class grades. Even here, however, recommendations on how to train students to use the most effective techniques would benefit from further research. One key issue concerns the earliest age at which a given technique could (or should) be taught. Teachers can expect that upper elementary students should be capable of using many of the techniques, yet even these students may need some guidance on how to most effectively implement them. Certainly, identifying the age at which students have the self-regulatory capabilities to effectively use a technique (and how much training they would need to do so) is an important objective for future research. Another issue is how often students will need to be retrained or reminded to use the techniques to ensure that students will continue to use them when they are not instructed to do so. Given the promise of some of the learning techniques, research on professional development that involves training teachers to help students use the techniques would be valuable.

Beyond training students to use these techniques, teachers could also incorporate some of them into their lesson plans. For instance, when beginning a new section of a unit, a teacher could begin with a practice test (with feedback) on the most important ideas from the previous section. When students are practicing problems from a unit on mathematics, recently studied problems could be interleaved with related problems from previous units. Teachers could also harness distributed practice by re-presenting the most important concepts and activities over the course of several classes. When introducing key concepts or facts in class, teachers could engage students in explanatory questioning by prompting them to consider how the information is new to them, how it relates to what they already know, or why it might be true. Even homework assignments could be designed to take advantage of many of these techniques. In these examples (and in others provided in the Issues for Implementation subsections), teachers could
implement a technique to help students learn, regardless of whether students are themselves aware that a particular technique is being used.

We realize that many factors are responsible whenever any one student fails to achieve in school (Hattie, 2009) and hence that a change to any single factor may have a relatively limited effect on student learning and achievement. The learning techniques described in this monograph will not be a panacea for improving achievement for all students, and perhaps obviously, they will benefit only students who are motivated and capable of using them. Nevertheless, when used properly, we suspect that they will produce meaningful gains in performance in the classroom, on achievement tests, and on many tasks encountered across the life span. It is obvious that many students are not using effective learning techniques but could use the more effective techniques without much effort, so teachers should be encouraged to more consistently (and explicitly) train students to use learning techniques as they are engaged in pursuing various instructional and learning goals.

Acknowledgments

We thank Reed Hunt, Mark McDaniel, Roddy Roediger, and Keith Thiede for input about various aspects of this monograph. We appreciate constructive comments about a draft from Sean Kang, Alexa MacKendrick, Richard Mayer, Hal Pashler, Dan Robinson, and Doug Rohrer. Thanks also to Robert Goldstone, Detlev Leutner, and Doug Rohrer for providing details of their research, and to Melissa Bishop and Cindy Widack for technical support.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by a Bridging Brain, Mind and Behavior Collaborative Award through the James S. McDonnell Foundation’s 21st Century Science Initiative.

Notes

1. We also recommend a recent practice guide from the U.S. Institute of Education Sciences (Pashler et al., 2007), which discusses some of the techniques described here. The current monograph, however, provides more in-depth and up-to-date reviews of the techniques and also reviews some techniques not included in the practice guide.

2. Although this presentation mode does not involve reading per se, reading comprehension and listening comprehension processes are highly similar aside from differences at the level of decoding the perceptual input (Gernsbacher, Varner, & Faust, 1990).

3. We did not include learning conditions as a category of variable in this table because the techniques vary greatly with respect to relevant learning conditions. Please see the reviews for assessments of how well the techniques generalized across relevant learning conditions.

References


Learning Styles
Concepts and Evidence

Harold Pashler,1 Mark McDaniel,2 Doug Rohrer,3 and Robert Bjork4

1University of California, San Diego, 2Washington University in St. Louis, 3University of South Florida, and 4University of California, Los Angeles

SUMMARY—The term “learning styles” refers to the concept that individuals differ in regard to what mode of instruction or study is most effective for them. Proponents of learning-style assessment contend that optimal instruction requires diagnosing individuals’ learning style and tailoring instruction accordingly. Assessments of learning style typically ask people to evaluate what sort of information presentation they prefer (e.g., words versus pictures versus speech) and/or what kind of mental activity they find most engaging or congenial (e.g., analysis versus listening), although assessment instruments are extremely diverse. The most common—but not the only—hypothesis about the instructional relevance of learning styles is the meshing hypothesis, according to which instruction is best provided in a format that matches the preferences of the learner (e.g., for a “visual learner,” emphasizing visual presentation of information).

The learning-styles view has acquired great influence within the education field, and is frequently encountered at levels ranging from kindergarten to graduate school. There is a thriving industry devoted to publishing learning-styles tests and guidebooks for teachers, and many organizations offer professional development workshops for teachers and educators built around the concept of learning styles.

The authors of the present review were charged with determining whether these practices are supported by scientific evidence. We concluded that any credible validation of learning-styles-based instruction requires robust documentation of a very particular type of experimental finding with several necessary criteria. First, students must be divided into groups on the basis of their learning styles, and then students from each group must be randomly assigned to receive one of multiple instructional methods. Next, students must sit for a final test that is the same for all students. Finally, in order to demonstrate that optimal learning requires that students receive instruction tailored to their putative learning style, the experiment must reveal a specific type of interaction between learning style and instructional method: Students with one learning style achieve the best educational outcome when given an instructional method that differs from the instructional method producing the best outcome for students with a different learning style. In other words, the instructional method that proves most effective for students with one learning style is not the most effective method for students with a different learning style.

Our review of the literature disclosed ample evidence that children and adults will, if asked, express preferences about how they prefer information to be presented to them. There is also plentiful evidence arguing that people differ in the degree to which they have some fairly specific aptitudes for different kinds of thinking and for processing different types of information. However, we found virtually no evidence for the interaction pattern mentioned above, which was judged to be a precondition for validating the educational applications of learning styles. Although the literature on learning styles is enormous, very few studies have even used an experimental methodology capable of testing the validity of learning styles applied to education. Moreover, of those that did use an appropriate method, several found results that flatly contradict the popular meshing hypothesis.

We conclude therefore, that at present, there is no adequate evidence base to justify incorporating learning-styles assessments into general educational practice. Thus, limited education resources would better be devoted to adopting other educational practices that have a strong evidence base, of which there are an increasing number. However, given the lack of methodologically sound studies of learning styles, it would be an error to conclude that all possible versions of learning styles have been tested and found wanting; many have simply not been tested at all.

Address correspondence to Harold Pashler, Department of Psychology 0109, University of California, San Diego, La Jolla, CA 92093; e-mail: hpashler@ucsd.edu.
Further research on the use of learning-styles assessment in instruction may in some cases be warranted, but such research needs to be performed appropriately.

INTRODUCTION

The term learning styles refers to the view that different people learn information in different ways. In recent decades, the concept of learning styles has steadily gained influence. In this article, we describe the intense interest and discussion that the concept of learning styles has elicited among professional educators at all levels of the educational system. Moreover, the learning-styles concept appears to have wide acceptance not only among educators but also among parents and the general public. This acceptance is perhaps not surprising because the learning-styles idea is actively promoted by vendors offering many different tests, assessment devices, and online technologies to help educators identify their students' learning styles and adapt their instructional approaches accordingly (examples are cited later).

We are cognitive psychologists with an interest both in the basic science of learning and memory and in the ways that science can be developed to be more helpful to teachers and students. We were commissioned by Psychological Science in the Public Interest to assess, as dispassionately as we could, the scientific evidence underlying practical application of learning-style assessment in school contexts. This task involved two steps: (a) analyzing the concept of learning styles to determine what forms of evidence would be needed to justify basing pedagogical choices on assessments of students' learning styles and (b) reviewing the literature to see whether this evidence exists. Our team began this undertaking with differing—but not passionately held—opinions on learning styles as well as a shared desire to let the empirical evidence lead us where it would.

We start by offering the reader a brief overview of the learning-styles concept, including some of the publications and entrepreneurial ventures that have been developed around the idea. Next, we analyze the learning-styles concept from a more abstract point of view. Here, we grapple with some potentially confusing issues of definition and logic that in our opinion require more careful consideration in connection with learning styles than they have so far received. We argue that this analysis is a useful, and essential, prerequisite to organizing and appraising the evidence on learning styles. Finally, we describe the results of our search of published literature, draw some conclusions, and suggest lines of future research. We should emphasize, however, that the present article is not a review of the literature of learning styles; indeed, several such reviews have appeared recently (e.g., Coffield, Moseley, Hall, & Ecclestone, 2004; Kozhevnikov, 2007; Sternberg, Grigorenko, & Zhang, 2008). In brief, we sought to determine what kinds of findings would provide sufficient evidence for the learning-styles concept, as detailed in the following sections, and then we searched for evidence that satisfied this minimal criterion.

AN OVERVIEW OF LEARNING STYLES: DOCTRINES AND INDUSTRY

As described earlier, the concept of learning styles encompasses not only a large body of written materials but also what seems to be a thriving set of commercial activities. The writings that touch on the learning-styles concept in its broadest sense include several thousand articles and dozens of books. These figures may seem surprisingly large, but one should keep in mind the sheer number of different schemes or models of learning styles that have been proposed over the years. For example, in a relatively comprehensive review, Coyle et al. (2004) described 71 different schemes, and they did not claim that their list was exhaustive.

The commercial activity related to learning styles is largely centered around the publishing and selling of measurement devices to help teachers assess individual learning styles; typically, although not always, these devices classify the learner into different style categories. Testing has been recommended by organizations at all levels of education that might be presumed to base their recommendations on evidence. For example, the National Association of Secondary School Principals commissioned the construction of a learning-styles test that it distributed widely (Keefe, 1988). Similarly, the Yale Graduate School of Arts and Sciences (2009) currently maintains a Web site that offers advice for Yale instructors; the site informs visitors that “college students enter our classrooms with a wide variety of learning styles.” The site goes on to recommend that instructors determine their own “modality of learning” as well as assess their students’ learning styles and make their instructional choices accordingly.

Furthermore, the learning-styles concept is embraced in a number of current educational psychology textbooks. For instance, Omrod (2008) wrote, “Some cognitive styles and dispositions do seem to influence how and what students learn.... Some students seem to learn better when information is presented through words (verbal learners), whereas others seem to learn better when it's presented through pictures (visual learners)” (p. 160, italics in original). Thus, educational psychology students and aspiring teachers are being taught that students have particular learning styles and that these styles should be accommodated by instruction tailored to those learning styles.

Some of the most popular learning-style schemes include the Dunn and Dunn learning-styles model (e.g., Dunn, 1990), Kolb’s (1984, 1985) Learning Styles Inventory, and Honey and Mumford’s (1992) Learning Styles Questionnaire. The assessment devices that have been developed in relation to the model of Dunn and Dunn are particularly popular and extensive.
Customers visiting the Web site of the International Learning Styles Network (www.learningstyles.net) are advised that

Learning style is the way in which each learner begins to concentrate on, process, absorb, and retain new and difficult information (Dunn and Dunn, 1992; 1993; 1999). The interaction of these elements occurs differently in everyone. Therefore, it is necessary to determine what is most likely to trigger each student’s concentration, how to maintain it, and how to respond to his or her natural processing style to produce long term memory and retention. To reveal these natural tendencies and styles, it is important to use a comprehensive model of learning style that identifies each individual’s strengths and preferences across the full spectrum of physiological, sociological, psychological, emotional, and environmental elements. (International Learning Styles Network, 2008)

As of June 2008, the company sells five different assessment tools for different age groups—ranging from the Observational Primary Assessment of Learning Style (OPAL) for ages 3 to 6 to Building Excellence (BE) for ages 17 and older (at a cost of approximately $5.00 per student for the classification instrument). The vendor claims these assessments “measure the patterns through which learning occurs in individual students; they summarize the environmental, emotional, sociological, physiological, and global/analytic processing preferences that a student has for learning” (International Learning Styles Network, 2008). A summer certification program is also offered in connection with this approach (the basic certification program costs $1,225 per trainee, excluding meals and lodging, with a higher level certification for conducting research on learning styles also offered for an additional $1,000). The Dunn and Dunn assessment instrument for adults asks respondents to indicate, for example, whether they learn best when they hear a person talk about something, whether their desk is typically disorganized and messy, whether they would say that they normally think in words as opposed to mental images, and whether they would characterize themselves as someone who thinks intuitively or objectively (Rundle & Dunn, 2007).

Kolb’s (1984, 1985) Learning Styles Inventory is another very popular scheme, particularly within the United States. It conceives of individuals’ learning processes as differing along two dimensions: preferred mode of perception (concrete to abstract) and preferred mode of processing (active experimentation to reflective observations). The Learning Styles Inventory classifies individuals into four types on the basis of their position along these two dimensions: dividers (concrete, reflective), assimilators (abstract, reflective), convergers (abstract, active), and accommodators (concrete, active). The self-assessment requires people to agree or disagree (on a 4-point scale) with, for example, the idea that they learn best when they listen and watch carefully, or that when they learn they like to analyze things and to break them down into parts.

The Learning Styles Inventory is distributed by the Hay Group (http://www.haygroup.com) and sold in packs of 10 booklets for approximately $100.00 (as of June 2008). The Hay Group also distributes an informational booklet called “One Style Doesn’t Fit All: The Different Ways People Learn and Why It Matters” (Hay Group, n.d.). According to the booklet, the practical benefits of classifying individuals’ learning styles include “placing them in learning and work situations with people whose learning styles are different from their own,” “improving the fit between their learning style and the kind of learning experience they face,” and “practicing skills in areas that are the opposite of their present strengths” (Hay Group, n.d., p. 11).

These three examples are merely some of the more popular and well-advertised products within the learning-styles movement. Readers interested in a more comprehensive view should consult Coffield et al. (2004).

HOW DID THE LEARNING-STYLES APPROACH BECOME SO WIDESPREAD AND APPEALING?

Origin and Popularity

The popularity and prevalence of the learning-styles approach may, of course, be a product of its success in fostering learning and instruction. Assessing the extent to which there is evidence that the approach does indeed foster learning is the primary goal of this review. However, there are reasons to suspect that other factors—in addition to, or instead of, actual effectiveness—may play a role in the popularity of the learning-styles approach.

Most learning-styles taxonomies are “type” theories: That is, they classify people into supposedly distinct groups, rather than assigning people graded scores on different dimensions. One can trace the lineage of these theories back to the first modern typological theorizing in the personality field, which was undertaken by the psychiatrist and psychoanalyst C.G. Jung (1964). Jung’s ideas were explicitly incorporated into a psychological test developed in the United States, the Myers–Briggs Type Indicator test. This test became very popular starting in the 1940s and remains widely used to this day. The Myers–Briggs categorizes people into a number of groups, providing information that is said to be helpful in making occupational decisions. The assumption that people actually cluster into distinct groups as measured by this test has received little support from objective studies (e.g., Druckman & Porter, 1991; Stricker & Ross, 1964), but this lack of support has done nothing to dampen its popularity. It seems that the idea of finding out “what type of person one is” has some eternal and deep appeal, and the success of the Myers–Briggs test promoted the development of type-based learning-style assessments.

Another, very understandable, part of the appeal of the learning-styles idea may reflect the fact that people are concerned that they, and their children, be seen and treated by educators as unique individuals. It is also natural and appealing to think that all people have the potential to learn effectively and easily if only instruction is tailored to their individual learning styles. Another related factor that may play a role in the popu-
larity of the learning-styles approach has to do with responsibility. If a person or a person's child is not succeeding or excelling in school, it may be more comfortable for the person to think that the educational system, not the person or the child himself or herself, is responsible. That is, rather than attribute one's lack of success to any lack of ability or effort on one's part, it may be more appealing to think that the fault lies with instruction being inadequately tailored to one's learning style. In that respect, there may be linkages to the self-esteem movement that became so influential, internationally, starting in the 1970s (Twenge, 2006).

Interactions of Individual Differences and Instructional Methods
As we argue in the next section, credible evidence in support of practices based on learning styles needs to document a specific type of interaction between instructional method and assessments of an individual's learning style. Basically, evidence for a learning-styles intervention needs to consist of finding that a given student's learning is enhanced by instruction that is tailored in some way to that student's learning style.

Naturally, it is undeniable that the optimal instructional method will often differ between individuals in some respects. In particular, differences in educational backgrounds can be a critical consideration in the optimization of instruction. New learning builds on old learning, for example, so an individual student's prior knowledge is bound to determine what level and type of instructional activities are optimal for that student. Many research studies (see, e.g., McNamara, Kintsch, Butler-Songer, & Kintsch, 1996) have demonstrated that the conditions of instruction that are optimal differ depending on students' prior knowledge. Later in this review, we summarize some of the evidence suggesting that aptitude measures can help predict what instructional methods are most effective.

WHAT EVIDENCE IS NECESSARY TO VALIDATE INTERVENTIONS BASED ON LEARNING STYLES?

We turn now to the core of the learning-styles idea: an assessment of the degree to which it has been validated.

Existence of Study Preferences
In reviewing the literature on learning styles and examining the different ways in which this term is frequently used, we make a basic distinction between what we call the existence of study preferences and what we call the learning-styles hypothesis. The existence of preferences, as we interpret it, amounts simply to the fact that people will, if asked, volunteer preferences about their preferred mode of taking in new information and studying. Given that learning-style questionnaires focusing on preferences have at least some psychometric reliability (i.e., a person's score on one day predicts their score on another day; e.g., Henson & Hwang, 2002; Veres, Sims, & Shake, 1987), the existence of preferences with some coherence and stability is not in dispute. A study by Massa and Mayer (2006), which is discussed in more detail later, provides further evidence on this point. Massa and Mayer developed three instruments to assess people's preferences for receiving instruction verbally versus accompanied by pictorial illustrations. Responses on these instruments were significantly correlated with the degree to which college students chose to receive verbal elaboration versus pictorial elaboration of technical terms in an electronics lesson. Massa and Mayer also found significant correlations between the instruments they used to assess people's preference for certain kinds of representations and the mode of elaboration people elected to receive in the electronics lesson. (As discussed at more length later, however, the preference for visual versus verbal information intake had little, if any, relationship to an individual's objectively measured specific-aptitude profile.)

Having noted the reality of these preferences, we emphasize that the implications of such preferences for educational practices and policies are minimal. The existence of preferences says nothing about what these preferences might mean or imply for anything else, much less whether it is sensible for educators to take account of these preferences. Most critically, the reality of these preferences does not demonstrate that assessing a student's learning style would be helpful in providing effective instruction for that student. That is, a particular student's having a particular preference does not, by itself, imply that optimal instruction for the student would need to take this preference into account. In brief, the existence of study preferences would not by itself suggest that buying and administering learning-styles tests would be a sensible use of educators' limited time and money.

The Learning-Styles Hypothesis
What, then, is the version of the learning-styles hypothesis that has practical implications for educational contexts? It is the claim that learning will be ineffective, or at least less efficient than it could be, if learners receive instruction that does not take account of their learning style, or conversely, it is the claim that individualizing instruction to the learner's style can allow people to achieve a better learning outcome.

It is important to note that there is a specific version of the learning-styles hypothesis that evidently looms largest both within the educational literature and within the minds of most people writing about learning styles: the idea that instruction should be provided in the mode that matches the learner's style. For example, if the learner is a "visual learner," information should, when possible, be presented visually. We refer to this specific instance of the learning-styles hypothesis as the meshing hypothesis—the claim that presentation should mesh with the learner's own proclivities.

Most proponents of the learning-styles idea subscribe to some form of the meshing hypothesis, and most accounts of how in-
struction should be optimized assume the meshing hypothesis: For example, they speak of (a) tailoring teaching to "the way in which each learner begins to concentrate on, process, absorb, and retain new and difficult information" (Dunn & Dunn’s framework; International Learning Styles Network, 2008), (b) the learner’s preferred modes of perception and processing (Kolb’s, 1984, 1985, framework), or (c) “the fit between people’s learning style and the kind of learning experience they face” (Hay Group, n.d., p. 11). Note that the learning-styles hypothesis, as defined here, could be true without the meshing hypothesis being true—if, for example, individuals classified as visual learners profited more from verbal instruction in some situations or if individuals classified as verbal learners profited more from visual instruction. In our review, we searched for evidence for both this broad version of the learning-styles hypothesis and the more specific meshing hypothesis.

Interactions as the Key Test of the Learning-Styles Hypothesis
To provide evidence for the learning-styles hypothesis—whether it incorporates the meshing hypothesis or not—a study must satisfy several criteria. First, on the basis of some measure or measures of learning style, learners must be divided into two or more groups (e.g., putative visual learners and auditory learners). Second, subjects within each learning-style group must be randomly assigned to one of at least two different learning methods (e.g., visual versus auditory presentation of some material). Third, all subjects must be given the same test of achievement (if the tests are different, no support can be provided for the learning-styles hypothesis). Fourth, the results need to show that the learning method that optimizes test performance of one learning-style group is different than the learning method that optimizes the test performance of a second learning-style group.

Thus, the learning-styles hypothesis (and particular instructional interventions based on learning styles) receives support if and only if an experiment reveals what is commonly known as a crossover interaction between learning style and method when learning style is plotted on the horizontal axis. Three such findings are illustrated in Figures 1A to 1C. For each of these types of findings, the method that proves more effective for Group A is not the same as the method that proves more effective for Group B. One important thing to notice about such a crossover interaction is that it can be obtained even if every subject within one learning-style group outscores every subject within the other learning-style group (see Fig. 1B). Thus, it is possible to obtain strong evidence for the utility of learning-style assessments even if learning style is correlated with what might, for some purposes, be described as ability differences. Moreover, the necessary crossover interaction allows for the possibility that both learning-style groups could do equally well with one of the learning methods (see Fig. 1C).

Figures 1D to 1I show some hypothetical interactions that would not provide support for the learning-styles hypothesis because, in each case, the same learning method provides optimal learning for every learner. Note that these findings are insufficient even though it is assumed that every interaction in Figure 1 is statistically significant. It is interesting to note that the data shown in Figures 1D and 1G do produce a crossover interaction when the data are plotted so that the horizontal axis represents learning method, as shown in Figure 2, but this mere rearrangement of the data does not alter the fact that the same learning method maximizes performance of all subjects. Thus, as noted earlier, a style-by-method crossover interaction constitutes sufficient evidence for the learning-styles hypothesis if and only if the horizontal axis represents learning style, as in Figures 1A to 1C.

To provide the most liberal criterion in our search for evidence supporting the learning-styles hypothesis, we cast the hypothesis so that it requires only the style-by-method crossover interaction described previously. It does not require that the optimal method for each group would somehow match or conform to each group’s learning style (the meshing hypothesis referred to earlier).

Primary Mental Abilities: Relation to Learning Styles
In our discussion of styles thus far, we have focused on preferences for how information would be presented to a person rather than on the notion of the person having different ability to process one kind of information or another. This focus is in conformity with the dictionary definition of style and matches at least the most typical usage of the term learning style within the education field. However, the notion of learning style as a set of preferences and the notion of learning style as a specific aptitude are very closely intertwined in many discussions of learning styles. Moreover, it is our impression that among the general public, the notion of learning styles and the notion of differential abilities are scarcely distinguished at all. There is, after all, a commonsense reason why the two concepts could be conflated: Namely, different modes of instruction might be optimal for different people because different modes of presentation exploit the specific perceptual and cognitive strengths of different individuals, as suggested by the meshing hypothesis.

Similar to the learning-styles hypothesis, the idea of specific abilities also implies a special form of crossover interaction. However, the interaction is different in kind from what was outlined earlier as the key test of the learning-styles hypothesis.

1 A reviewer of an earlier version of this article noted that the interactions shown in Figures 1H and 1I might have potential practical importance, even in the absence of a true crossover. If one could sort people into two groups, one of which would benefit from an instructional manipulation and the other of which was completely unaffected by it, it might (on some assumptions) be worthwhile doing the sorting and selectively offering the manipulation. We agree. However, we show later, the general conclusions reached here do not depend on this issue because we have not found any actual interactions of the types in Figures 1H and 1I in the learning-styles literature.
Acceptable Evidence
In examples A, B, and C, the learning method that optimized the mean test score of one kind of learner is different from the learning method that optimized the mean test score of the other kind of learner.

![Graph A](image)

![Graph B](image)

![Graph C](image)

Unacceptable Evidence
In examples D through I, the same learning method optimized the mean test score of both kinds of learners, thereby precluding the need to customize instruction.

![Graph D](image)

![Graph E](image)

![Graph F](image)

![Graph G](image)

![Graph H](image)

![Graph I](image)

Fig. 1. Acceptable and unacceptable evidence for the learning-styles hypothesis. In each of the hypothetical experiments, subjects have been first classified as having Learning Style A or B and then randomly assigned to Learning Method 1 or 2. Later, all subjects have taken the same test. The learning-styles hypothesis is supported if and only if the learning method that optimized the mean test score of one group is different from the learning method that optimized the mean test score of the other group, as in A, B, and C. By contrast, if the same learning method optimized the mean test score of both groups, as in D through I, the result does not provide evidence. (Note that all nine interactions are assumed to be statistically significant.) In general, the learning-styles hypothesis is supported if and only if a study finds a crossover interaction between learning method and learning style, assuming that the horizontal axis represents the learning-style variable. See the text for more details.

If the notion of specific aptitudes or skills is valid, one ought to be able to divide subjects into two or more groups (e.g., Group A of learners with high auditory ability and Group B of learners with high visual ability). There should then be two tests such that Group A outscores Group B on one test, whereas Group B outscores Group A on the other test.

There is little doubt that specific-ability differences of this kind exist. The first psychologist to provide strong empirical evidence for the idea of specific-ability differences was Louis Thurstone (e.g., Thurstone, 1938). Thurstone proposed seven "primary mental abilities": verbal comprehension, word fluency, number facility, spatial visualization, associative memory,
perceptual speed, and reasoning. Although these abilities are not completely uncorrelated (implying, to some, the idea of general mental ability or "g"; see Jensen, 1998; Spearman, 1927), they do show a moderate degree of independence (Thurstone, 1938). Although this provides evidence for specific aptitudes, it does not show that one needs to provide different groups with different forms of instruction to maximize their performance on any single outcome test. Thus, evidence for specific aptitudes does not, by itself, validate the learning-styles hypothesis.

There are few data on the relationship between preferences and specific aptitudes. However, one recent and well-executed study, which we discuss at more length later, discloses that preference for visual versus verbal information intake shows hardly any relationship to an individual's objectively measured specific-aptitude profile (Massa & Mayer, 2006). Thus, the common assumption that preferences and abilities are closely tied is open to challenge. But as we have defined the learning-styles hypothesis, one could find evidence for the hypothesis regardless of whether the style measure involved a specific aptitude, a preference, or both.

EVALUATION OF LEARNING-styles LITERATURE

Style-by-Treatment Interactions: The Core Evidence Is Missing

For the reasons described earlier, it is our judgment that a validation of an intervention based on learning styles would need to offer one kind of evidence, and one kind of evidence alone: a crossover interaction of the form illustrated in Figures 1A to 1C. On the basis of this analysis, we scoured the literature to identify studies that provided such evidence. Remarkably, despite the vast size of the literature on learning styles and classroom instruction, we found only one study that could be described as even potentially meeting the criteria described earlier, and as we report in the following text, even that study provided less than compelling evidence.

The study in question was reported by Sternberg, Grigorenko, Ferrari, and Clinkenbeard (1999). In this study, 324 "gifted and talented" high school students were given the Sternberg Triarchic Abilities Test, which provided a rating of each student's analytical, creative, and practical ability. On the basis of this test, the authors selected a subset of 112 subjects (35%) for whom one of these three abilities was much higher than the other two, and depending on their area of strength, these subjects were assigned to the high-analytical, high-creative, or high-practical groups. (Another 87 students were assigned to two additional groups not described here, and the remaining 125 students were excluded from the study.) The participating subjects enrolled in an introductory psychology summer course at Yale University, and each student was randomly assigned to class meetings that emphasized analytical instruction, creative instruction, practi-

---

Fig. 2. Examples of crossover interactions that would not validate the learning-styles hypothesis. The two hypothetical outcomes in A are identical to the outcomes in B, and these examples demonstrate that the choice of variable for the horizontal axis can affect whether an interaction appears to "cross over." Regardless of appearance, though, each of the graphs above demonstrates that the same learning method (Method 1) proved superior for all subjects. Thus, the data above do not provide evidence for the learning-styles hypothesis. However, if the horizontal axis depicts the learning-style variable, a crossover interaction is both sufficient and necessary to show evidence for the learning-styles hypothesis, as in Figures 1A–1C. Note that the above two results are identical to those in Figures 1D and 1G.
their course performance was assessed by raters, and the ratings were “subjected to principal-component analyses” (Sternberg et al., 1999, p. 7). The authors reported several analyses, and, for the analysis of the interaction of interest, they compared the course performance of matched subjects (i.e., students who received instruction that matched their strongest ability) to mismatched subjects. The article states that after the data were “screened for deviant scores” (Sternberg et al., 1999, p. 10), matched subjects reliably outscored mismatched subjects on two of the three kinds of assessments.

Thus, the authors reported a style-by-treatment interaction. Although suggestive of an interaction of the type we have been looking for, the study has peculiar features that make us view it as providing only tenuous evidence. For one thing, the reported interaction was found only with highly derived measures (as noted above), and the untransformed outcome measures (e.g., the mean score on each final assessment) were not reported for the different conditions. Furthermore, and as noted previously, only about one third of the subjects were classified into the groups that produced the interaction. Finally, the interaction was achieved only after the outliers were excluded for unspecified reasons. In brief, although the article presents data that may be worth following up, it has serious methodological issues. Even for those who might disagree with this judgment, the potential support that this study could provide for any of the particular interventions based on learning styles that are being marketed at the present time is extremely limited because the instructional manipulation does not seem to correspond to any of the more widely promoted and used learning-styles interventions.

In summary, our efforts revealed at most one arguable piece of evidence for the learning-styles hypothesis in general. For the many specific assessment devices and interventions being actively marketed to teachers, as described earlier in this article, we were unable to find any evidence that would meet the key criteria discussed earlier (i.e., interactions of the form shown in Figs. 1A–1C). Moreover, we found a number of published studies that used what we have described as the appropriate research design for testing the learning-styles hypothesis and found results that contradict widely held versions of the learning-styles hypothesis; we turn to these studies now.

Learning-Styles Studies With Appropriate Methods and Negative Results

Massa and Mayer (2006) reported a particularly informative and well-designed study of learning styles with a set of three experiments. They constructed a reasonably realistic computer-based electronics lesson. Two different sorts of help screens were customized for verbal or visual learners, providing either supplementary printed text or carefully developed diagrams and illustrations, respectively. A wide variety of preference-based and ability-based individual-difference measures were administered to sort visual from verbal learners in various ways. In general, the results, which the researchers replicated, showed no tendency for better performance for those who received help screens matched to their preferences. Critically, Massa and Mayer found no support for any of these interactions despite exhaustive analysis of nearly 20 individual-difference measures that spanned their three proposed facets of verbalizer–visualizer learning styles. The authors concluded that their results provided no support for “the idea that different instructional methods should be used for visualizers and verbalizers” (Massa & Mayer, 2006, pp. 333–334).

Within a medical-education context, a recent study by Cook, Thompson, Thomas, and Thomas (2009) examined the hypothesis that learners with a “sensing learning style” would do better when given instruction in which the problem was presented prior to the content information used to solve the problem, whereas “intuitive learners” would do better with the reverse. The authors noted that this learning-styles taxonomy is similar to Kolb’s (1984, 1985) concrete–abstract dimension. Studying a sample of 123 internal medicine residents and presenting modules on four ambulatory medicine topics, they found no support for this prediction.

Another study reaching a similar conclusion, albeit using tasks with less direct correspondence to real educational activities, was reported by Constantinidou and Baker (2002). These investigators used a laboratory task to ask whether self-reported preferences in information uptake predicted ability to perceive and store information in different modalities. They examined the relationship between adults’ scores on the Visualizer–Verbalizer Questionnaire (VrVQ; Richardson, 1977) to their verbal free-recall performance on a task that presented words through the auditory modality, the visual modality (as line drawings of the corresponding object), or both. The VrVQ asks people a series of questions about their relative preference for taking in information through verbal versus visual means. VrVQ scores were not related in any strong or clearly interpretable way to relative levels of free-recall performance for different input modalities. Visual presentations produced better free recall than did purely verbal presentations, and the authors reported finding “no relationship between a visual learning style and the actual learning of verbal items that are presented visually or auditorily” (Constantinidou & Baker, 2002, p. 306).

These studies, which we believe are methodologically strong, provide no support for the learning-styles hypothesis (or its popular specific version, the meshing hypothesis). As mentioned previously, however, it would clearly be a mistake to label these negative results as a conclusive refutation of the learning-styles hypothesis in general. Further research modeled on the work of Massa and Mayer (2006) may bring to light assessments paired with interventions that do meet our criteria. But at present, these negative results, in conjunction with the virtual absence of positive findings, lead us to conclude that any application of learning styles in classrooms is unwarranted.
RELATED LITERATURES WITH APPROPRIATE METHODOLOGIES

Aptitude-by-Treatment Interactions
Although the literature on learning styles per se has paid scarce attention to the need for group-by-treatment interactions, there has been a clear recognition of the importance of such interactions within an older educational psychology literature, going back to Cronbach’s (1957) appeal for research to uncover interactions between aptitude and aspects of the instructional context (termed treatments). Although the validity of aptitude-by-treatment interactions (ATIs) is a separate issue from the validity of learning-style measures, which is the primary focus of the current article, we describe several ATIs so that the reader may gain an appreciation of a literature that recognizes the need to demonstrate the necessary interaction.

Initial attempts to demonstrate so-called ATIs were reviewed in a classic work by Cronbach and Snow (1977). According to Cronbach and Snow, these attempts were not highly successful because treatment durations were too brief, and aspects of the methodologies were inadequate. After that review, significant improvements were made in methodologies, with a number of studies examining treatments implemented in classroom settings for relatively long durations.

The kind of potential interaction that has received the most attention within the ATI tradition involves the degree to which the teaching approach provides ample structure or guidance for the learner. The primary hypothesis that has stimulated much of the work in this area is the idea that students with high ability tend to fare better in less structured learning environments than in highly structured learning environments. By contrast, students with low ability are hypothesized to fare better with instruction that is highly structured and provides explicit guidance than with instruction that is less structured and provides little guidance (see, e.g., Snow, 1977). A variant of this theme that also sparked interest is the idea that highly structured situations might reduce performance differences between students with high and low abilities (Freebody & Tirre, 1985).

As detailed in the next paragraph, two key difficulties in evaluating this hypothesis are as follows: (a) The implementation of instructional methods that differ in structure (guidance) has been quite variable, and (b) the measures used to assess student abilities have varied considerably.

Freebody and Tirre (1985) reported an ATI in line with the above hypothesis that involved two competing reading-instruction approaches. One approach, the Matteson program (see Schlenker, 1978), provides a list of behavioral objectives in major reading-skills areas (e.g., word recognition, vocabulary development, literal and interpretive comprehension) combined with individualized learning packages that cover these areas, following precisely defined sequences. The other approach, the Scott Foresman (1972) program, is not strictly sequenced and monitored. Instead, the emphasis is on frequent discussions focusing on the literal and inferential aspects of discourse. This approach is assumed to place a greater burden on the student for acquiring specific reading skills (see Freebody & Tirre, 1985).

All of the sixth-grade students in a large school district who had been in one of the two reading programs for 2 years or longer served as subjects (N = 180, nearly equally distributed across reading programs). Their aptitudes were assessed with a standardized test that included nonverbal and verbal measures of ability. The outcome measure was the reading test score achieved at the conclusion of the sixth-grade year. Multiple regression analyses produced a significant ATI. The interpretation of the interaction was based on predicted outcomes (from the regression equations) for particular low-ability values and particular high-ability values. These predicted outcomes indeed showed that students with low ability would generally perform better on the structured reading program (Matteson) than on the less structured reading program (Scott Foresman). The reverse would be predicted for the students with high ability: better performance on the less structured than on the more structured reading instruction method. Although suggestive, these data do not establish that students at a particular ability level (either low or high) fared significantly better (in terms of reading outcomes) as a function of the reading program in which the students were enrolled.

Additional direct support for the idea that learning outcomes for students with high and low abilities might reverse with a greater degree of structure embedded in instruction was reported in the domain of elementary school mathematics (Cramer, Post, & Behr, 1989). Fourth graders being taught fractions were given four lessons (in six 40-minute class periods) on completing rational numbers tasks that involved shading a particular fractional area (two thirds) of different kinds of visual figures (e.g., a rectangle divided into three columns). In the high-structured condition, instruction was teacher centered with little student choice. The teacher paced through each example in large-group lecture fashion. In the low-structured condition, the teacher provided an initial introduction to the problems and then students worked through examples at their own pace. The materials involved leading questions to guide the learner to discovery of the key concepts. Both instructional conditions used identical examples, and both contained a 10-minute practice phase that completed each 40-minute class period. Students in the higher and lower ranges of cognitive restructuring ability, as measured by the Group Embedded Figures test (see Witkin & Goodenough, 1981), were assigned to each instructional condition. At the conclusion of the lessons, the students completed a final test containing problems (rational numbers tasks) of the type taught in the lessons. For the more difficult problems—those requiring physical restructuring of the diagrams—a crossover interaction between ability and the degree of instructional structure emerged. The students with high ability performed better following low- than high-structured instruction; by contrast, students with low ability performed better following high-
low-structured instruction. Particularly notable is that the students with low ability outscored the students with high ability (at least nominally) after both received the more highly structured instruction. This pattern thus provides evidence that learning is optimized when students with low ability are provided with structured instruction and students with high ability are provided with less structured instruction.

However, other studies that examined different content domains and used different assessment instruments did not always support the idea that high-ability students are better off with less structured instruction, whereas low-ability students profit more from higher-structured instruction. In Janicki and Peterson (1981), 117 grade school students completed a 2-week fractions unit in a "direct" instructional fashion (which involved homework assignments that students completed in class on their own) or in a less structured fashion (this involved mixed-ability four-student group seatwork with choice of homework or math games). Aptitude, as determined by a composite measure that included Ravens Progressive Matrices, did not interact with instructional method.

Greene (1980) similarly failed to find an interaction when fifth and sixth graders with high and low aptitude (as determined by Lorge-Thorndike verbal and nonverbal tests) were given high-structured instruction (specified sequence of workbook assignments and performance standards) or low-structured instruction (choice and pacing of which exercises to do in the workbook) on a letter-series task. The letter-series task was chosen to reflect general problem-solving goals in education. The basic result was that the students with higher aptitude performed better than the students with lower aptitude regardless of instruction.

In a well-conducted experiment, Peterson, Janicki, and Swing (1980, Study 2) manipulated instruction for a 2-week ninth-grade social studies unit across six classes (146 students). Two teachers taught each of three classes with one of three teaching methods. One teaching method was a standard lecture-recitation approach. In the second method, termed inquiry, students researched a historical question using primary sources. The third method, public issues discussion, required students to support a position on a current public issue using primary material. Aptitude was defined as verbal ability. The outcome measure was a test that included multiple-choice questions on historical facts and short essay questions requiring integration and evaluation of material. Critically, the test targeted readings and content common to all three instructional approaches. In line with the previous findings, there was no interaction between teaching method and ability for the essay performances, with students with higher ability performing better on the essay questions in general.

It is interesting to note that for the multiple-choice questions, there was a significant ATI such that students with high ability performed better with the lecture-recitation teaching method than with the inquiry or public issues discussion methods, whereas students with low ability performed better when receiving the inquiry or public issues methods than with the lecture-recitation method. This pattern would appear to counter the main hypothesis being considered in reviewing this body of ATI work, because the inquiry and public discussion methods encouraged learner self-direction (less structure). However, Peterson et al. (1980) offered an interpretation based on the underlying cognitive demands placed on the students by the different instructional methods. They suggested that the lecture-recitation approach implemented in the study placed a heavier burden on students’ cognitive skills than did the other approaches. Specifically, students had to comprehend and attend to the lectures, take careful notes, and memorize target information. The idea is that students with high ability would have the requisite skills to accomplish these challenges. Of course, this interpretation does not clarify why the students with high ability would fare less well with the other instructional methods, relative to the lecture-recitation method.

One study activity that appears to be sensitive to individual ability differences is concept mapping (creation of diagrams that show the relationship among concepts), with students with low verbal ability profiting more from concept maps (in a chemistry learning activity) than students with high verbal ability (Stensvold & Wilson, 1990). Not surprisingly, in most studies the students with higher ability outperformed the students with lower ability in both instructional conditions.

However, complete crossovers have recently been reported with embedded-question techniques for learning from textbook chapters. In Callender and McDaniel (2007), the ability of interest was the degree to which learners can construct a coherent representation of presented content (either through text or lectures). Poor structure builders are assumed to perform relatively poorly at constructing a coherent representation of connected discourse that is either read or spoken (Gernsbacher, 1990). Such comprehenders appear to construct too many substructures to accommodate incoming information, rather than constructing a unified integrated representation of the target material. By contrast, good structure builders are able to extract coherent, well-organized mental representation of the text. Accordingly, Callender and McDaniel reasoned that embedding questions into a textbook chapter would orient poor structure builders to anchoring information around which to build a coherent representation and therefore improve learning for students at this level of comprehension ability. Embedded questions might be superfluous for good structure builders, however, because they are already able to construct coherent representations.

To test these predictions, Callender and McDaniel (2007) had college-age subjects read a chapter from an introductory psychology textbook with or without embedded questions. Afterward, the subjects were given a multiple-choice test consisting of questions targeting the information featured by the embedded questions and questions on information not targeted by the embedded questions. For poor structure builders, embedded questions significantly improved performance on target questions (relative to reading without embedded questions) but not performance on nontarget questions. Good structure builders
did not profit from embedded questions, and indeed their performance for nontarget information was better without embedded questions. Note that these patterns could be considered evidence for the general notion that more guided study activities are preferable for comprehenders of lower ability, whereas less guided presentations (no embedded questions) are preferred for comprehenders of higher ability. These patterns clearly require replication, as only one chapter was considered and the subjects were in a laboratory experiment and not an actual course. Yet, this finding illustrates the potential fruitfulness of attempting to link more specific cognitive processing abilities to instructional techniques designed to dovetail with those abilities.

In summary, ATIs evidently do occur, but it has not been easy to determine exactly when they occur. This diversity of outcomes is perhaps not surprising given that available studies vary on a number of potential critical dimensions, including target content, particular implementations of variations in instructional structure, assessments used to index ability, and the kinds of criterial (outcome) tests used. In some studies, the ATIs can be reported for one type of criterial measure but not another (e.g., see Cramer et al., 1989; Peterson, 1979; Peterson et al., 1980). At best, then, the ATI literature provides a mixed picture. A few studies are consistent with the idea that structured instruction produces better learning outcomes for students of lower ability (relative to less structured instruction), whereas less structured instruction produces better learning outcomes for students with higher abilities (relative to structured instruction). But other studies either did not obtain significant ATIs involving general ability and the degree of structure in instruction or in some cases indicated that students with lower ability fared worse with structured instruction than with less structured instruction. The greater coherence of the literature assessing structure building suggests that a more fine-grained approach that focuses on individual differences in underlying cognitive processes, rather than general aptitudes, and implements instructional methods that target those processes may be more fruitful in producing robust interactions between learner ability and learner-directed activities.

**Personality-by-Treatment Interactions**

There are also some more fragmentary but methodologically sophisticated studies documenting personality-by-instructional treatment effects, though these findings, like the aptitude–treatment interactions described just above, do not speak to the validity of the learning-styles hypothesis. Several studies have looked at a personality measure called locus of control, which refers to an individual’s belief about whether his or her successes or failures are a consequence of internal or external factors (Rotter, 1966). An internal locus of control indicates a belief that outcomes are a consequence of one’s own actions. An external locus of control reflects the belief that outcomes are unrelated to one’s own actions. One hypothesis that has received consideration is that learners with an internal locus of control may fare better with less structured than with highly structured instruction, whereas learners with an external locus of control will achieve more with highly structured than with less structured instruction.

Several studies have examined this hypothesis in college mathematics classes for prospective elementary school teachers. Horak and Horak (1982) examined two instructional methods during a 2-week unit on transformational geometry, with each method randomly assigned to a particular class section (total number of students was 102). In the highly guided instruction ("deductive"), students were given rules or principles and then proceeded to apply the rules to examples. In the less guided instruction ("inductive"), students were given examples, with no rule or principle stated for the students or expected from them. The criterial test included questions designed to test lower levels of understanding (knowledge of terminology and reproduction of material presented) and higher levels of understanding (e.g., problem solving). Marginally significant support for the predicted interaction was found for the questions testing lower levels of understanding; Students with an external locus of control performed better after the highly guided instruction than after the less guided instruction. The reverse was observed for students with an internal locus of control, with performance after less guided instruction exceeding performance after highly guided instruction (this also occurred with the questions tapping higher levels of understanding).

Parallel findings of marginal magnitude were reported in similar mathematics classes for elementary school teachers with shorter treatment periods (McLeod & Adams, 1980/1981). Three experiments were conducted using somewhat different instantiations of amount of guidance given during instruction and somewhat different target content. In only one experiment was the interaction significant (although a second experiment showed the same pattern); In this experiment, all students spent 1 week learning about networks with an inductive set of materials (see earlier). The amount of guidance was manipulated by having students work individually on problems and encouraging help from the instructor (high guidance, here students asked may questions) or by having students work in groups of 4 (low guidance, very few questions were posed to the instructor). On an immediate but not a delayed (given several weeks after instruction) criterial test, students with an internal locus of control performed better with low guidance than with high guidance; the reverse was found for the students with an external locus of control. The absence of significant interactions in the other two experiments may have been a consequence of shorter treatments (75-min lesson in one experiment) or small sample size (just under 60 students in each experiment), as the authors suggested.

It is interesting to note that Janicki and Peterson's (1981) study that failed to find an interaction with general ability (reviewed in the preceding section) did observe a significant personality-by-treatment interaction with a composite factor of
locus of control and attitudes toward math (in teaching fractions to grade-school students). This composite factor of accountability for learning interacted with instructional method such that those students with higher accountability (more internal locus of control; 36% of the students) performed better on immediate and delayed computation and story-problem tests when in the less guided small group setting than when in the highly guided direct-instruction setting. Instructional setting did not produce differences for the students with lower accountability (external locus of control).

In summary, there is modest evidence for the idea that students with an internal locus of control benefit more from less guided or structured instruction than from more guided instruction, whereas students with an external locus of control might benefit more from guided (structured) instruction than from less guided (structured) instruction. Previous studies reinforce those reviewed herein with similar patterns (Daniels & Stevens, 1976; Horak & Slobodzian, 1980; Parent, Forward, Cantor, & Mohling, 1975; Yeany, Dost, & Mattews, 1980). The reliability and generalizability of these findings to other content areas and to longer instructional treatments remain to be demonstrated. A clear uncertainty is specifying the exact aspects of instruction (group vs. individual work; density of questions directed at the instructor; homework choice vs. no choice) that are interacting with locus of control.

CONCLUSIONS AND RECOMMENDATIONS

Our evaluation of the learning-styles concept led us to identify the form of evidence needed to validate the use of learning-style assessments in instructional settings (i.e., Figures 1A–1C). As described earlier, our search of the learning-styles literature has revealed only a few fragmentary and unconvincing pieces of evidence that meet this standard, and we therefore conclude that the literature fails to provide adequate support for applying learning-style assessments in school settings. Moreover, several studies that used appropriate research designs found evidence that contradicted the learning-styles hypothesis (Massa & Mayer, 2006; Constantinou & Baker, 2002). Finally, even if a study of a particular learning-style classification and its corresponding instructional methods was to reveal the necessary evidence, such a finding would provide support for that particular learning-style classification only—and only then if its benefits surpass the high costs of student assessments and tailored instruction.

Our conclusions have particularly clear-cut implications for educational researchers, in our opinion. We urge investigators examining learning-styles concepts to embrace the factorial randomized research designs described in the earlier “Interactions as the Key Test of the Learning-Styles Hypothesis” section, because these alone have the potential to provide action-relevant conclusions. The kind of research that is needed must begin by classifying learners into categories based on clearly specified measures and then randomize learners to receive one of several different instructional treatments. Equally crucial, the interventions must be followed by a common prespecified learning assessment given to all the participants in the study. The paucity of studies using this methodology is the main factor that renders the learning-styles literature so weak and unconvincing, despite its large size.

Points of Clarification

Although we have argued that the extant data do not provide support for the learning-styles hypothesis, it should be emphasized that we do not claim that the same kind of instruction is most useful in all contexts and with all learners. An obvious point is that the optimal instructional method is likely to vary across disciplines. For instance, the optimal curriculum for a writing course probably includes a heavy verbal emphasis, whereas the most efficient and effective method of teaching geometry obviously requires visual–spatial materials. Of course, identifying the optimal approach for each discipline is an empirical question, and we espouse research using strong research methods to identify the optimal approach for each kind of subject matter.

Furthermore, it is undoubtedly the case that a particular student will sometimes benefit from having a particular kind of course content presented in one way versus another. One suspects that educators’ attraction to the idea of learning styles partly reflects their (correctly) noticing how often one student may achieve enlightenment from an approach that seems useless for another student. There is, however, a great gap from such heterogeneous responses to instructional manipulations—whose reality we do not dispute—to the notion that presently available taxonomies of student types offer any valid help in deciding what kind of instruction to offer each individual. Perhaps future research may demonstrate such linkages, but at present, we find no evidence for it.

Costs and Benefits of Educational Interventions

It should also be noted that even if the evidence had convincingly documented style-by-method interactions—which we have concluded is scarcely the case—the interactions would need to be large and robust, and not just statistically significant, before the concomitant educational interventions could be recommended as cost-effective. After all, there is no doubt that interventions built around learning styles will be costly. Students must be assessed and grouped by learning style and then given some sort of customized instruction, which, in turn, requires additional teacher training as well as the creation and validation of instructional activities for each learning style. Moreover, if one is to partition the children within a given classroom and teach each subset differently, this may require increasing the number of teachers. Ultimately, the practical question will be whether the benefits of learning-styles inter-
ventions exceed other ways of using the time and money needed to incorporate these interventions.

Beliefs Versus Evidence as a Foundation for Educational Practices and Policies

Basic research on human learning and memory, especially research on human metacognition, much of it carried out in the last 20 years or so, has demonstrated that our intuitions and beliefs about how we learn are often wrong in serious ways. We do not, apparently, gain an understanding of the complexities of human learning and memory from the trials and errors of everyday living and learning. Many demonstrations have shown that participants who are asked to predict their own future performance following conditions of instruction that researchers know to be ineffective will often predict better performance under poorer conditions of instruction than will participants provided with better conditions of instruction (for a review, see Schmidt & Bjork, 1992). Part of the problem is that conditions that make performance improve rapidly during instruction or training, such as blocking or temporal massing of practice, can fail to support long-term retention and transfer, whereas conditions that introduce difficulties for learners and appear to slow the learning process, such as interleaving different types of problems, or employing temporal spacing of practice on what is to be learned, often enhance long-term retention and transfer. As learners, we can also be fooled by subjective impressions, such as the ease or sense of familiarity we gain on reading expository text or how readily some information comes to mind, both of which can be products of factors unrelated to actual comprehension or understanding.

There is growing evidence that people hold beliefs about how they learn that are faulty in various ways, which frequently lead people to manage their own learning and teach others in non-optimal ways. This fact makes it clear that research—not intuition or standard practices—needs to be the foundation for upgrading teaching and learning. If education is to be transformed into an evidenced-based field, it is important not only to identify teaching techniques that have experimental support but also to identify widely held beliefs that affect the choices made by educational practitioners but that lack empirical support. On the basis of our review, the belief that learning-style assessments are useful in educational contexts appears to be just that—a belief. Our conclusion reinforces other recent skeptical commentary on the topic (e.g., Coffield et al., 2004; Curry, 1990; Willingham, 2005, 2009). Future research may develop learning-style measures and targeted interventions that can be shown to work in combination, with the measures sorting individuals into groups for which genuine group-by-treatment interactions can be demonstrated. At present, however, such validation is lacking, and therefore, we feel that the widespread use of learning-style measures in educational settings is unwise and a wasteful use of limited resources.

Everybody's Potential to Learn

As a final comment, we feel the need to emphasize that all humans, short of being afflicted with certain types of organic damage, are born with an astounding capacity to learn, both in the amount that can be learned in one domain and in the variety and range of what can be learned. Children, unless stifled in some way, are usually virtuosos as learners.

As we asserted earlier, it is undeniable that the instruction that is optimal for a given student will often need to be guided by the aptitude, prior knowledge, and cultural assumptions that student brings to a learning task. However, assuming that people are enormously heterogeneous in their instructional needs may draw attention away from the body of basic and applied research on learning that provides a foundation of principles and practices that can upgrade everybody's learning. For example, the finding that learners' memory for information or procedures can be directly enhanced through testing (Roediger & Karpicke, 2006) is not something that applies to only a small subset of learners but (as far as can be told) applies to all. Although performance of a student on a test will typically depend on that student's existing knowledge, testing (when carried out appropriately, which sometimes requires providing feedback) appears to enhance learning at every level of prior knowledge.

Given the capacity of humans to learn, it seems especially important to keep all avenues, options, and aspirations open for our students, our children, and ourselves. Toward that end, we think the primary focus should be on identifying and introducing the experiences, activities, and challenges that enhance everybody's learning.

SUMMARY

Our review of the learning-styles literature led us to define a particular type of evidence that we see as a minimum precondition for validating the use of a learning-style assessment in an instructional setting. As described earlier, we have been unable to find any evidence that clearly meets this standard. Moreover, several studies that used the appropriate type of research design found results that contradict the most widely held version of the learning-styles hypothesis, namely, what we have referred to as the meshing hypothesis (Constantinidou & Baker, 2002; Massa & Mayer, 2006). The contrast between the enormous popularity of the learning-styles approach within education and the lack of credible evidence for its utility is, in our opinion, striking and disturbing. If classification of students' learning styles has practical utility, it remains to be demonstrated.

Acknowledgments—We are grateful to Maxwell Moholy for assistance with the literature review. We also thank Elena L. Grigorenko, Roddy Roediger, and two anonymous reviewers for their comments on an earlier version of this article.
REFERENCES


Progressive mastery through deliberate practice: A promising approach for improving writing

Article in Improving Schools - April 2016
DOI: 10.1177/1365480216643209

1 author:
Nell Scharff Panero
City University of New York - Bernard M. Baruch College
3 PUBLICATIONS 4 CITATIONS

All content following this page was uploaded by Nell Scharff Panero on 13 June 2016.
The user has requested enhancement of the downloaded file.
Progressive mastery through deliberate practice: A promising approach for improving writing

Nell Scharff Panero
Baruch College, City University of New York, USA

Abstract
This study is an analysis of the curriculum used to teach writing at one US high school in which outcomes for students were extremely strong. The study surfaces what was different in the approach used from what is typically understood and promoted as best practice in the teaching of writing. It does so in order to surface what elements of writing instruction are fruitful for further research. Findings identify an approach to teaching writing called progressive mastery through deliberate practice. This article articulates the five elements that constitute this approach (understanding the final product, breaking down its critical elements, creating a hierarchy among the elements, layering and progression of elements, and gradual release of scaffolds), as well as the ways they work in concert to bring about improved student writing and thinking.

Keywords
High school reform, high school writing, writing and thinking, writing for content learning, writing instruction

There is virtual agreement across all sectors (policy-makers, researchers, educators, parents, employers) that there is a crisis in literacy internationally, not only in reading — which has been the focus of policy and research in recent decades — but also — a more recent focus of attention — in writing. A 2000 report by the US National Reading Panel crystallized the scientifically understood elements of effective reading instruction in English, and there is general agreement among reading researchers that if these elements were taught effectively, reading failure in English would be ‘largely avoidable’ – that the current 20–30 percent of reading failure in the United States, for example, could be reduced to between 2 and 10 percent (Walsh, Glaser, & Wilcox, 2006, p. 6). Current efforts to improve reading in the United States, therefore, focus on ensuring that teachers are trained in known strategies rather than on trying to determine the nature of effective strategies themselves.1

This same approach would not make sense in writing, however, which rests on a much less robust research base (Graham & Perin, 2007). Writing performance is arguably more dismal than

Corresponding author:
Nell Scharff Panero, School of Public Affairs, Baruch College, City University of New York, 1 Bernard Baruch Way, New York NY 10010, USA.
Email: helen.panero@baruch.cuny.edu
reading. According to the National Assessment of Educational Progress (US Department of Education, 2015), widely regarded as the gold standard of US educational research and referred to as the Nation’s Report Card, 75 percent of US 4th and 8th grade students write below grade-level standards, with a mere 3 percent of US students from across all demographics deemed ‘advanced’ – indicating that more than poverty is at play, that in fact US schools are not effectively teaching writing to any population, if they are teaching writing at all. While it is hard to get a handle on international comparisons (Programme for International Student Assessment (PISA), the widely respected international survey of 15-year-olds across 64 countries every 3 years, does not assess writing specifically), 2012 results indicate that more than one out of every four students in the nations studied – over 13 million students in 64 countries – has failed to achieve the ‘very basic’ level of academic proficiency needed to secure school and economic success (Organisation for Economic Co-operation and Development (OECD, 2016). The absence of writing as a distinct measure speaks to its relative neglect in education reform efforts generally (Graham, Capizzi, Harris, Hebert, & Morphy, 2014), although this may be changing.

Revisions to the National Curriculum in England since 1999 have increased the emphasis on non-fiction writing and language-based instruction in an attempt to shift the ‘personal growth/literary model’ that has and may still dominate English instruction (Andrews, Torgerson, Graham, & McGuinn, 2009). In the United States, the Common Core State Standards (CCSS), released in 2010 and currently the major focus of US school reform, prioritize expository writing as a lever for improving life chances for individuals, closing achievement gaps between populations, and improving US competitiveness internationally. High expectations and frequent expository writing in English, social studies, and science are core pillars of the standards. One problem is that research-based understandings of effective writing practices are insufficiently captured in national/state standards (Soiferman, Boyd, & Straw, 2010; Troia et al., 2015). A second problem is that current research-based understandings of what it would take to raise student performance in writing to meet the new standards are inadequate. Graham and Perin (2007), gurus of writing research, write, ‘Research is clearly needed not only to identify additional effective practices that already exist but to develop new ones’ (p. 26).

This study seeks to expand the notion of what elements of writing instruction show promise for future research. It does so by looking closely at a particular approach to teaching writing as implemented in one specific US high school in which results (improved writing, graduation rates, and measures of college readiness) were dramatic and in which for the most part methods used were different from those currently understood and promoted as ‘best practice’ in the teaching of writing. The school studied – New Dorp High School in Staten Island, NY – and its turnaround were documented in the widely read and acclaimed ‘The Writing Revolution’ (Tyre, 2012). The article describes, in brief, the role of writing instruction across subjects and grades in improving student writing. This study compares New Dorp’s approach to writing instruction with what is generally accepted as true, assuming that differences in this school’s approach account for differential outcomes. The hope is that doing so will point to promising directions for future writing research.

Effective writing instruction: what do we know from research?

Most of what is generally understood and accepted as true about effective writing instruction in English for students in grades 4–12 is captured and most concisely expressed in two reports of meta-analyses of existing research. The first, Writing Next: Effective Strategies to Improve Writing of Adolescents in Middle and High Schools (Graham & Perin, 2007), draws from prior meta-analyses and subsequent research (176 experimental and quasi-experimental studies of grades 4–12) to identify and recommend 11 instructional practices shown to improve the quality
of student writing in varied contexts (see Table 1 for the recommendations along with their effect sizes and numbers of studies each was based upon). The second, *Writing to Read: Evidence for How Writing Can Improve Reading* (Graham & Hebert, 2010), drew on 93 experimental and quasi-experimental studies and found three general categories of writing instruction and some specific sub-components that improve reading comprehension (see Table 2). Notably, some frequently studied strategies were not found to improve student writing. Traditional grammar instruction, for example, revealed a small but negative impact on the quality of student writing, and the simple act, often promoted, of increasing students’ time spent writing yielded no significant improvement. Graham and Perin (2007) cite Braddock and Jones’ (1969) earlier claim ‘that providing more opportunities to write without effective instruction and motivation is not enough to improve writing quality’ (p. 26).

Policy, research, and practitioner audiences regard these recommendations as the best of research-based knowledge to inform writing programming and instruction in English. The recommendations provide some guidance. The first and strongest recommendation, with an effect size of .82, for example, calls for direct, explicit, and systematic instruction in planning, revising, and editing. Beyond naming effective strategies and providing varied examples within a category, however, the recommendations offer little clear direction. First, strategies were studied in isolation and need to be put together in flexible ways that constitute a curriculum or approach, but the interaction of elements has not been sufficiently studied to recommend how. Second, while impact in some

---

**Table 1.** Strategies found to improve writing quality (Graham & Perin, 2007).

<table>
<thead>
<tr>
<th>Recommended strategy</th>
<th>Effect size</th>
<th>No. of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy instruction (teaching strategies for planning, revising and editing)</td>
<td>.82</td>
<td>20</td>
</tr>
<tr>
<td>Summarization</td>
<td>.82</td>
<td>4</td>
</tr>
<tr>
<td>Collaborative writing</td>
<td>.75</td>
<td>7</td>
</tr>
<tr>
<td>Setting specific product goals</td>
<td>.70</td>
<td>5</td>
</tr>
<tr>
<td>Word processing</td>
<td>.55</td>
<td>18</td>
</tr>
<tr>
<td>Sentence combining</td>
<td>.50</td>
<td>5</td>
</tr>
<tr>
<td>Prewriting</td>
<td>.32</td>
<td>5</td>
</tr>
<tr>
<td>Process writing approach</td>
<td>.32</td>
<td>5</td>
</tr>
<tr>
<td>Study of models</td>
<td>.25</td>
<td>6</td>
</tr>
<tr>
<td>Writing for content learning</td>
<td>.23</td>
<td>26</td>
</tr>
</tbody>
</table>

---

**Table 2.** Writing strategies found to improve reading comprehension (Graham & Hebert, 2010).

<table>
<thead>
<tr>
<th>Recommended strategy</th>
<th>Effect size</th>
<th>No. of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Having students write about text they read</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing reactions</td>
<td>.77</td>
<td>9</td>
</tr>
<tr>
<td>Writing summaries</td>
<td>.52</td>
<td>19</td>
</tr>
<tr>
<td>Writing notes</td>
<td>.47</td>
<td>23</td>
</tr>
<tr>
<td>Writing and/or answering questions</td>
<td>.27</td>
<td>8</td>
</tr>
<tr>
<td>II. Teaching skills that go into creating a text (process of writing, text structures for writing, paragraph or sentence construction skills)</td>
<td>.18&lt;sup&gt;a&lt;/sup&gt;</td>
<td>12</td>
</tr>
<tr>
<td>III. Increase how much students write</td>
<td>.27&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5</td>
</tr>
</tbody>
</table>

<sup>a</sup>For standardized norm-referenced tests.  
<sup>b</sup>For researcher-designed tests.
categories relies on direct, systematic instruction, impact in others (summarizing, for instance) is said to result equally from either ‘a rule-governed or a more intuitive approach’ (Graham & Perin, 2007, p. 16). In other words, the nature and number of the studies drawn upon are, according to the authors, sufficient for making strong recommendations about strategies that work but insufficient for dictating overall approaches or implementation details.

While writing research has grown tremendously (Hillock’s 1986 meta-analysis included only 29 studies compared to Graham & Perin’s 176), there is much work to be done. Specific limitations of current research include the lack of numbers of studies overall (only five recommended elements across both reports yielded 10 or more studies that met the criteria for inclusion) and the particular dearth of studies about low-achieving writers, especially about adolescents in urban settings, although in the United States these are the students with the lowest writing proficiency and most in need.

What’s typically done?

Although research-based understanding of how writing is taught is severely limited, especially in middle and high schools (Applebee & Langer, 2009), the recommendations put forth in Writing Next (Graham & Perin, 2007) reflect known trends and tensions in writing instruction (Gilbert & Graham, 2010), specifically the shift in recent decades from a more traditional, teacher-centered model to the ‘caught, not taught’ approach thought to dominate elementary and middle school writing instruction (Graham as cited in Tyre, 2012) and to influence high school writing instruction as well.2 In this approach, students are immersed in rich literacy environments and expected to ‘catch’ most of what is needed to become effective writers. Most students, according to Graham, do not catch what they need – and those coming from poverty, with learning disabilities and/or limited English proficiency, and/or those with weak instruction may not catch much at all (as quoted in Tyre, 2012). High schools are less ideological, perhaps, although the recent emphasis on holistic approaches, such as collaborative writing, is evident there as well. While the emphasis in elementary schools has been on creative and narrative writing (although a shift to including more expository writing is occurring now, as a result of the common core), high school teachers are expected to teach mostly expository writing. The issues here are that high school writing is infrequent, of insufficient length, relegated primarily to English class, and not reflective of evidence-based practices (Applebee & Langer, 2009; Graham & Hebert, 2010).

Writing Next recommends strategies typically associated with both camps. Writing process, for example, is typically associated with a more wholistic approach, whereas direct instruction in sentence combining might be seen as ‘traditional’. These polarities may be less relevant to classroom teachers looking to integrate the best of what works for their specific students regardless of ideology of approach than to policy-makers or education schools. The tension is alive and palpable in US schools, however, along with the perception, not validated by systematic research, that holistic approaches along with those ‘favoring self-expression and emotion over lucid communication’ (Tyre, 2012) currently hold sway.

New Dorp High School

In 2005, New Dorp High School was a typical large, comprehensive US struggling high school with a population comparable to other urban schools and struggling for survival.3 In that year, they launched an intensive, multi-year improvement process coupling re-organization into smaller, theme-based learning communities with a strategy for instructional improvement in which teacher teams were supported to identify and close skill gaps for struggling students and in which what was
learned from this process informed systemic, school-wide changes so that future students would not get stuck in the same places. By 2009, 4 years into their improvement process, teachers across the building, of various grades and subjects, had identified and come to a common understanding of fundamental skill gaps in student writing – gaps that they had not previously known to exist. They had identified, for example, that many students could not use the conjunctions because, but, and so correctly in a sentence.

At this point, there was no explicit or systematic approach to teaching writing. As part of the improvement process, teachers had observed, transcribed, and analyzed classroom instruction – leading them to realize that they were not actually teaching writing in ways that addressed the newly identified needs. As one English teacher said after analyzing transcripts of her own classes, I used to think I was teaching writing. Now I realize that I’m having students write – a little at the beginning of the period, or in a journal, or even for an essay assignment. But I’m not actually teaching writing.

Starting in 2009, New Dorp selected and implemented Teaching Basic Writing Skills (Hochman, 2009) first in the Social Studies department and then, starting in 2011, school wide. Over the next few years, student engagement and on-track measures improved. Teachers and administrators reported that special education students and English language learners who had been disengaged in the past, reportedly putting their heads on desks during exams that they barely attempted, were now drafting outlines and using all the time allotted to complete and succeed on essay-based exams. Most dramatically, in 2009, one social studies teacher of upper level students who take a prestigious exam associated with a college-level US History course had all 23 of her students take the associated exam, passage of which affords college credit. Only two of her students passed, both with a score of 3. In 2012, the same teacher had students in her course who had experienced 3 years of the school’s new approach to systematic writing instruction. In her class of 28 students, 26 now passed the exam with a score of 4 or 5. The teacher, an experienced lead teacher who had initially resisted the program fearing it would dilute time spent on content, directly attributed the improved outcomes to the impact of the writing program – which improved student writing along with their content knowledge, thinking, and speaking. ‘Sometimes I can’t even sleep at night thinking about how big this is and what’s coming. I can feel it’, she said, marveling at the transformation in her own teaching and the sea change she imagined a broader knowledge of the practices could bring.

Teaching basic writing skills

New Dorp’s strategies are based primarily on those in Teaching Basic Writing Skills (Hochman, 2009), an approach to the teaching of basic expository writing skills developed by Judith C. Hochman. The approach aligns with some but not all recommendations by Graham and colleagues. It aligns in its overall emphasis on direct, systematic instruction and the explicit teaching of strategies for planning, revising, and editing (recommendation #1). However, the strategies used are more scaffolded and directly controlled than is suggested in the recommendations. Originally designed for special education students and most famously implemented at The Windward School, an independent school for special education students in a suburb of New York City, this was the first time the author knows that the approach went full scale (school wide) in a general education setting.

The approach has three elements – a focus on sentences, paragraphs, and essays. Most counter-intuitive, perhaps, is its robust focus on sentence-level strategies – the explicit teaching of sentence-level strategies to improve the quality of sentences, to reinforce content, and to develop the
elemental building blocks of expression and thinking. The paragraph work is highly structured as well, primarily with two core outlines – one for a single paragraph and the other for a multi-paragraph essay. At each stage, students apply what they learned earlier so that the work deepens and builds as students become increasingly independent.

This approach is not a program or a curriculum per se. It involves embedding sentence, paragraph, and essay level strategies in content instruction to develop written and oral expression simultaneously with content knowledge. What is most notable and atypical perhaps, especially for high schools, is that the foundational elements are taught discretely and slowly. New Dorp’s students focus almost exclusively on foundational note-taking and sentence-level strategies for their entire first semester. At the end of 9th grade, they are expected to have mastered only the paragraph. They begin writing full essays at the beginning of 10th grade, their second year. Social studies and English classes take the lead, introducing each strategy and providing daily practice. Other subjects select strategies that have been introduced and that support their content objectives, providing weekly practice. Foundational strategies are never dropped. Even when writing full essays, students continue to practice foundational sentence and paragraph skills.

This atypical practice to writing instruction used at New Dorp had strong results. What has not been understood, however, is how it worked to bring about these results. This study analyzes the key pieces of New Dorp’s approach and the interactions among these pieces to posit a theory of progressive mastery as a beginning explanation.

Method

Data and rationale

This study involves an analysis of the curriculum used to teach writing at New Dorp High School in order to deduce the approach used. Data include writing curriculum in the subject area in which implementation and outcomes were strongest in the school – Social Studies – and for the grade levels and courses (9th- and 10th-grade Global History and 11th-grade US History) that culminated in the first evidence of direct impact on a trusted, national measure (the advanced placement history exam taken at the end of 11th grade). Progress for 12th-grade students is measured primarily by course passage and graduation rates, providing less persuasive evidence of direct impact. Therefore, 12th-grade curriculum was not included in this study.

Specific data included worksheets created by members of the social studies department and presented to students at the start of each year in bound workbooks – one workbook per year-long course with one to two worksheets per instructional day. These materials mirrored the scope and sequence of writing skills determined by the department and that later guided roll-out of the work school-wide, allowing students daily practice of specific, predetermined writing skills embedded in content and for the department (and later school) to have a unified, systematic approach for teaching writing. They reveal a focus primarily on sentence and note-taking skills in the first semester of freshman year, the continuation of these skills and addition of paragraph level skills for the second semester of freshman year, and the continuation of these skills with the addition of essay level skills in the 10th and 11th grades.

The workbooks were collected for this research in the fall of 2013 (for the 9th grade) and the fall of 2014 (for 10th and 11th grades). Some activities, therefore, had been revised since the years leading to the US History exam results achieved in 2012. According to the principal, however, revisions were tweaks rather than substantial. For example, they involved adding or deleting the number of days on a particular strategy. Taken together, the three workbooks provide a clear picture of the curriculum and instruction that were provided to students in their social studies courses from 2009 to 2012.
Data analysis

In order to identify areas of overlap and disconnect with existing writing research, the data were coded using an open coding system with a start list drawn from the existing literature (Birks & Mills, 2015; Charmaz, 2014; Corbin & Strauss, 2014; Glaser & Strauss, 1967; Grove, 1968; Martin & Gynnild, 2011). The recommended strategies in Writing Next (Graham & Perin, 2007) and Writing to Read (Graham & Hebert, 2010) became preliminary codes, with new codes emerging from the data itself. A few codes overlapped. Writing for content learning, for example, appeared frequently in the New Dorp data since many activities were designed to improve writing and content mastery simultaneously. Summarization and sentence combining also appeared in both.

Most new codes were sub-categories of those in the start list. Most notably, there were 93 emergent sub-codes falling under Graham et al.’s first recommendation, writing strategies, which calls for direct instruction in strategies for planning, revising, and editing. This difference is to be expected, since the current study is of a particular approach, whereas meta-analyses by definition identify larger patterns. Some new codes emerged that did not fall under those from the start list. Most notably, note-taking and previewing emerged as important elements not found by the meta-analyses to improve the quality of writing. (Note-taking is identified in the meta-analyses as improving reading, however.) Of note as well were codes from the start list (found to improve writing and/or reading according to the meta-analyses) that did not show up at all in the New Dorp curriculum – for which there were no data coded. These include collaborative writing, word processing, and inquiry activities.

After initial coding, materials were grouped and sorted to identify patterns, and six overarching categories were identified – the large ‘buckets’ (Ely, Anzul, Friedman, Garner, & Steinmetz, 1991) into which the sub-codes fell – as follows: what is (1) taught, (2) previewed, (3) practiced, (4) applied, (5) modeled, and (6) teaches content. Based on these categories, the piles of codes were stabilized. Then, the researcher sought to understand the relationships among the elements/categories. To do so, she created a schema – an overarching sense of how the six umbrella categories work together to capture, in essence, how the approach operates as a system (Lewis, 1992).

Findings

Progressive mastery through deliberate practice

Progressive mastery differs from typical writing instruction primarily in terms of the role of the teacher in providing a controlled, deliberate, and highly scaffolded instructional experience for students. The approach – progressive mastery through deliberate practice – requires that the teacher have a clear sense not only of the whole which the student must come to master independently and flexibly but also of the smaller pieces that comprise it. The teacher must understand how the particular small pieces function not only to develop writing but also to strengthen content learning, oral expression, and reading (how it is, therefore, that they are high-leverage), along with how best to sequence and layer them to gradually develop student mastery.

The teacher’s role in a sense is to protect the student from feeling overwhelmed by what he or she cannot yet master. The teacher deliberately selects and sequences small pieces of instruction based on knowledge of how they spiral and build in response to the evidence of student learning needs that each instructional activity yields. Students experience mastery of the small pieces along the way, which cultivates engagement and self-efficacy (for teacher and student) and spurs student motivation. There is no script. Rather there is ongoing strategic decision-making by the teacher based on expert knowledge (of the components of writing instruction and how they layer) and
expert practice (responsiveness to formative evidence). In this approach, the students achieve progressive mastery through the teacher’s deliberate practice.

Notably, the small pieces of instruction are in fact very small – smaller than those typically taught in high schools. The sentence-level strategies that form the heart of the approach and that demand substantial instructional time include, for example, completing given sentence stems using the conjunctions because, but, and so, as in the following example (note that the stem given by the teacher is in italics; only the underlined portion is provided by the student):

The Neolithic Revolution led to new occupations because there was a growth in trade.

The Neolithic Revolution led to new occupations, but it led to reduced food production.

The Neolithic Revolution led to new occupations, so new towns and cities developed.

Critics may argue that this activity and others like it are too low-level, especially for high school. What is critical, however, is the teacher’s understanding of why and when it is worth spending the time on something so small. First, it must be the right small thing – something carefully chosen because it targets and develops an essential element of thinking that students lack. In fact, recent research of public school students in New York City indicates that its high school students do not, for the most part, know how because, but, and so function in a sentence and that when they are explicitly taught the function of these and other sentence-level components, their learning of content and expression about and beyond that content accelerate (Panero & Talbert, 2013). Second, the strategies chosen must teach and/or reinforce core content.

While this study draws on curriculum from social studies in particular, the analysis reveals the role of the strategies in supporting content learning in general. The purpose of the progressive mastery approach may be best understood not as teaching writing per se but rather as drawing upon writing strategies to develop the underpinnings of communication and thinking.

What are the specifics of how it works?

The overall framework described above can be understood in terms of the following five distinct and overlapping elements that comprise it: the teacher’s (1) understanding of the final product, (2) breaking down its critical elements, (3) creating a hierarchy among the elements, (4) layering and progression of elements, and (5) gradual release of scaffolds.

Understanding the final product

In progressive mastery, end goals are understood in the precise detail which then drives the instruction needed to achieve them. In order to teach argumentative and persuasive writing, for example, the teacher must know exactly what the final product for each will look like and require, including the distinct requirements of a thesis statement for each. In the curriculum studied, the argumentative essay is defined as presenting both sides of an argument without taking a stand, whereas in the persuasive essay, the author acknowledges both sides but sides with one. Students are expected to distinguish between and create thesis statements like the following, the first argumentative and the second persuasive:

While some claim that Shakespeare was the most important writer of the Renaissance, others argue that Machiavelli was the most influential.
While some claim that Shakespeare was the most important writer of the Renaissance, clearly Machiavelli was the most influential in his time.

Similarly, the end product for a body paragraph specifies more than a statement of belief followed by reasons. The body paragraph in the curriculum studied requires a general statement for the topic sentence, three to four detail sentences linked by appropriate transitions, and a concluding sentence, also a general statement, that comprises a concluding transition and that paraphrases the topic sentence.

The purpose of these examples is not to suggest that these are the only or best ways to write or teach thesis statements and paragraphs. Rather, they are intended to illustrate how the teacher’s understanding of the final product works in the progressive mastery approach. The clarity of what is expected in the final product (overall and at any moment in time) dictates and drives the resulting focus and grain size of instruction. This ensures mastery of the small pieces that build to the desired end.

Breaking down its critical elements

A related but distinct element of progressive mastery is the teacher’s ability to see the pieces that make up the final product – whatever it may be – separately, as pieces. The single paragraph, for example, is a piece of a larger product (the essay). Critical smaller pieces of the body paragraph as described above (including those as small as a sub-component of a concluding sentence, such as paraphrasing) must at times be broken down even further.

To teach paraphrasing, a critical smaller element of a concluding sentence as described in the prior example, the teacher must understand what more and less sophisticated examples of paraphrasing look like, the particular challenges involved in learning to do it, and the steps (scaffolds) that allow for gradual mastery. Drilling down within paraphrasing begins to make clear the complexity involved in doing and teaching it. How is a student to know what words to paraphrase? A student must be able to identify the most important words in a sentence and to supply words to replace them. To do so, students are taught to distinguish a synonym from a word that isn’t necessarily a synonym but that may lend specificity or nuance. Critical as well is the teacher’s understanding of the developmental arc of learning to paraphrase. In the curriculum for this study, students underline and restate first on a word and then on a phrase level. Gradually, they learn to paraphrase ideas in longer and increasingly complex texts.

The example below illustrates how the word-level strategy is modeled and scaffolded at New Dorp and how paraphrasing is taught concurrently with content and the skill of producing concluding sentences. In the directions, students are asked to produce a concluding sentence ‘for a given topic sentence and details’. First, they are provided with a list of concluding transitions to draw on and then the following model in which the concluding sentence is provided for them:

T.S. Alexander the Great, a powerful military leader, had a huge impact on Greece.

1. fought battles → many victories
2. built empire = Europe + Middle East + Asia
3. expansion → cultural diffusion
4. Hellenism = blending of Greek + Indian + Egyptian + Persian culture

C.S. Clearly, Alexander the Great had a lasting effect on Greece.
The example illustrates how closed (rather than open) assignments work within progressive mastery. The closed design focuses students on particular skills in order to ensure progressive mastery. It simultaneously embeds them in a larger context so as to reinforce previously taught skills (note-taking with key words and symbols) and to prepare students for future skills (writing a single paragraph outline).

Creating a hierarchy among the elements

In addition, the progressive mastery teacher must identify those elements that are most foundational and elemental and therefore most critical to teach and worthy of instructional time. These are the pieces without which students will not be able to progress—although a teacher might not know why—and that when taught explicitly and reinforced can spur jumps in student performance in the many areas they undergird: oral language, knowledge acquisition, reading comprehension, and writing (Tyre, 2012). They are also the pieces that high school teachers have not typically been trained to teach or to assess. In the curriculum studied, they include distinguishing general from specific statements and expressing complex ideas through use of subordinating conjunctions.

Typically, when teaching the paragraph, high school teachers begin by teaching the topic sentence followed by details. This seems logical since the topic sentence comes first and dictates the details. In the curriculum studied, however, the key distinction between topic and detail sentences is that the first is a general statement, whereas the second is specific. By beginning instruction by having students distinguish between the two, the teacher assesses (and reinforces) the students’ knowledge of the underlying, foundational skill that lies at the heart of all categorization. Many students, New Dorp teachers found, were not clear about this distinction until it was taught to them directly, as in the following activity in which students identify related statements as either general (G) or specific (S):

___ The disease caused widespread labor shortages because skilled workers died.
___ There were many economic effects of the Black Plague.

If a student grasps this distinction, the teacher moves on. If not, the teacher continues to teach it. If not thinking consciously about teaching a foundational, transferable skill, a teacher of topic sentences may be open to a range of what students produce, including one that contains a fair amount of detail. If the purpose of teaching the topic sentence, however, is not only the paragraph at hand but (more importantly) to assess and consolidate student knowledge of the underlying relationship between general and specific—one that will be transferable to a broad range of academic learning and thought including and beyond writing—teachers will require a clear general statement like the one in the example above until they are certain that students have mastered the core distinction.

Similarly, completing sentences that begin with a subordinating conjunction (since, while, although, etc.) is a useful tool for developing the underlying skills needed to express relationships, first within the bounds of a single sentence. Subordinating conjunctions introduce clauses that cannot stand alone—that are subordinate to another clause or clauses. Having students write sentences that start with one, therefore, pushes them to write complex rather than simple sentences. In the studied curriculum, students build this muscle first by being asked to complete given sentence stems that begin with a subordinating conjunction, as follows:

Since the Tigris and Euphrates rivers were unpredictable,

While Sumerians did create the first schools.
Once they have demonstrated mastery of the more scaffolded version (above), then they are provided with only the topic and the first word, as follows:

**Topic: Sparta**

**While**

**If**

What is critical is the teacher’s understanding that not all elements are equally important. Those that provide needed foundational underpinnings and that transfer to a range of academic tasks are prioritized. The coordinating conjunction ‘so’ and the arrow symbol (→) in note-taking, as modeled to students in the concluding sentence example shown earlier, may seem very tiny but support students’ expression of the bigger, critical concept of cause and effect and, in the hands of a strong teacher, build students’ capacity for increasingly complex expressions of that relationship. Learning to start a single sentence with a subordinating conjunction might at first seem too circumscribed to be effective in developing the underlying capacity to express complexity. Subsequently, however, students apply this strategy to writing argumentative and persuasive thesis statements that begin with a subordinating conjunction. This allows them to conceptualize and lay out their entire essay structure up front for the reader and reflect application and extension of the thinking that was the teacher’s primary target.

**Layering and progression of elements**

An analysis of the relationship between two discrete yet connected activities – expansion and revision of simple, given paragraphs – illustrates the importance of the proper layering and progression of elements within progressive mastery.

The primary purpose of sentence expansion is to support students to provide more information than they typically do within the context of a single sentence. In this activity, the teacher provides a simple, unelaborated sentence and two to five related question words (who, where, when, how, and/or why). The student then writes one expanded sentence containing information from the simple sentence and answers to the questions posed, as required by the following example:

Athens and Sparta fought.  
When? ........................................................................................................
Why? ........................................................................................................
Expanded Sentence: ................................................................................

The expansion strategy looks backward in that students must draw upon some of what they have learned previously to complete it correctly. Before trying expansion, students learn that dotted lines require notes only (solid lines call for complete sentences); the basics of note-taking using key words, symbols, and abbreviations; and the function of the coordinating conjunctions because, but, and so, which they draw upon to answer why in the example above.

The activity is forward-looking as well, preparing students to apply what is learned in a contained fashion more broadly and with increasing independence in future tasks. In a subsequent activity, for example, students apply knowledge of how the question words function in expansion to improving simple, given paragraphs. Students are provided with a paragraph made up of three to six simple, unelaborated sentences, such as ‘Caesar was a leader.’ ‘He
ruled.’ and ‘He gave land.’ Then they are asked to interrogate the paragraph – to draw upon knowledge of the question words that the teacher provided in the prior activity (along with a few other previously taught sentence strategies) and to annotate the paragraph with those question words to indicate what specific information in what places, if provided, would best improve it.

The following model indicates for students how their annotation might look:

\[
(\text{Appositive}) \quad (\text{When? Where?}) \quad (\text{How?}) \quad (\text{Why? to Who?})
\]

- Caesar was a leader.
- He ruled.
- He gave land.

\[
(\text{What jobs?}) \quad (\text{How?}) \quad (\text{Closing transition and what impact?})
\]

- He created jobs. He reformed.
- He had an impact.

Once introduced, most strategies within progressive mastery are taught concurrently. Sentence activities, for example, are never dropped, even when students write full essays. Complexity depends less on the length or nature of a specific strategy drawn upon than on the complexity of the reading material and/or ideas filtered through it. There is, however, a logical order and progression to when many activities are introduced, guided by knowledge of what the current layer depends upon and enables.

**Gradual release of scaffolds**

Scaffolds are, by definition, temporary supportive structures which must, as students become increasingly able, be released. The approach to teaching the introductory paragraph for a multiple paragraph essay in this study illustrates how gradual release of scaffolds operates within progressive mastery.

In the curriculum studied, students learn a formula for a three-sentence introduction: a general statement, followed by a specific statement, followed by the thesis statement (GST). When students are first exposed to the multiple paragraph outline with which they plan an essay, they are not asked to plan an introduction at all. Instead, a thesis statement is provided for them. Their task is to elicit the categories embedded within the thesis statement, place them on the outline tool, and complete the outline using notes – all skills they have learned previously. The first exposure to the thesis statement, then, is to work with one that has been provided for them.

Next, the teacher breaks down and allows for progressive mastery of each element of GST. Students learn, for example, that adding a date, place, or definition to discussion of a topic will yield a specific (rather than a general) statement and are guided to generate ideas in a chart before writing their own specific statement for given general and thesis statements, as follows:

<table>
<thead>
<tr>
<th>Date</th>
<th>Place</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Directions: Write a Specific Statement (Sp. St.) for the given General Statement (G. S.) and Thesis Statement (Th. St.)

1. G. S. Many leaders have experienced periods of greatness.


3. Th. St. Napoleon had many social and political achievements.

To consolidate students’ skill in creating each type of sentence and the paragraph overall, the teacher gradually releases scaffolds. First, there is only exposure. Then, as in the example above, students are supported to write one type of statement when provided with the other two. Next, the student is asked to generate two when provided with one. Finally, the student must create the entire paragraph independently, as in the following subsequent activity in which students are asked only to ‘write a GST introduction’ for the following prompt:

Many controversial domestic issues have divided the American people. The United States government has taken action to address these issues.

Select two controversial domestic issues that have divided the American people and for each

- Discuss the historical background of the controversy
- Explain the points of view of those who supported and those who opposed this issue
- Discuss one United States government action that was taken to address this issue

For a time, a simple outline tool with only solid and dotted lines is all that remains. Eventually, by necessity, even that is removed, since students are unable to bring an outline with them when taking a standardized exam. By then, the hope is that students have internalized the core outline structures and foundational pre-requisites and that they can and will build upon what’s been internalized for the rest of their writing lives.

Discussion

Findings from this study suggest that progressive mastery is different from what is typical and/or generally accepted as best practice in writing instruction in terms of the role that writing plays in the curriculum as well as the roles of teachers and students in relation to each other and the curriculum. In addition, it reveals differences in the assumptions and beliefs that drive all aspects of writing instruction (see Table 3 for a distillation of these differences).

In the typical approach, the role of the writing teacher (usually the English or humanities teacher, especially at the high school level) is to develop writing. Other subject teachers see themselves primarily as content experts. In the progressive mastery approach, these two notions are not in opposition. Since the role of writing in progressive mastery is not only to develop writing but also to develop content knowledge, expression, and thinking, all teachers play a key role. This does not mean that it is everyone’s job to teach the expository essay. It means that all subject teachers draw upon a similar language and set of writing strategies to the extent that these serve their specific, content-based instructional goals and with the knowledge that doing so simultaneously improves students’ writing.

The typical approach assumes that poor expository writing results from a lack of student interest and motivation, and that dry, formal, overly structured tasks are ineffective in cultivating the
Table 3. Different understandings of roles and assumptions/beliefs driving writing instruction.

<table>
<thead>
<tr>
<th>Understanding of role of...</th>
<th>Typical</th>
<th>Progressive Mastery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing in the curriculum</td>
<td>To develop writing</td>
<td>To develop thinking</td>
</tr>
<tr>
<td>The writing teacher</td>
<td>To foster discovery and self-expression through relatively open or unstructured tasks, allowing for struggle and exploration; to provide guidance primarily on the writing process.</td>
<td>To foster mastery through relatively closed or structured tasks, allowing for progressive mastery of small pieces that build and transfer; to provide direct instruction in specific skills</td>
</tr>
<tr>
<td>The student</td>
<td>To discover, experiment, find one's voice</td>
<td>To meet specific, clearly articulated goals</td>
</tr>
</tbody>
</table>

Assumptions and beliefs about...

<table>
<thead>
<tr>
<th>What cultivates student motivation</th>
<th>The opportunity to engage with rigorous, high-level tasks and/or those that students can relate to on a personal level; student choice about what to write about and how</th>
<th>Precise skill needs being addressed via tasks that are challenging but achievable; the teacher determining what students will write and how it should look (specific expectations and aligned feedback)</th>
</tr>
</thead>
<tbody>
<tr>
<td>How rigor is defined</td>
<td>Extent to which task requirements align with an objective standard at a moment in time; length of writing produced</td>
<td>Extent to which task requirements target precise student learning needs at a moment in time; quality of thinking developed</td>
</tr>
<tr>
<td>Who is responsible for teaching writing</td>
<td>English Language Arts and/or humanities teachers</td>
<td>All content teachers</td>
</tr>
</tbody>
</table>

needed motivation. The response is to hook students with topics that are more creative and/or rooted in their personal lives and to maximize student choice — with the idea that much of what students need to become strong writers is inside them already, needing free expression before it can then be shaped through practice and with help from the teacher.

Progressive mastery is rooted in a different set of assumptions about the underlying causes of poor writing and about what best cultivates engagement. It assumes that disinterest is a symptom rather than a cause, that students are missing fundamental skills without which they are unable to engage successfully in many of the writing tasks that are put before them, and that when the missing skills are explicitly taught in ways that can be rapidly mastered, students feel effective and, as a result, engaged. Choice is secondary not because proponents of progressive mastery dismiss its inherent value, but rather because the approach requires teacher-generated, highly controlled, content-based tasks that allow for gradual mastery. Creating, monitoring, and adjusting these tasks in an ongoing fashion to ensure that they hit the mark for both skill and content are labor-intensive and therefore much more difficult for multiple topics at once than for one.

Finally, the two approaches define rigor differently. In the typical approach, rigor is understood in terms of the extent to which a writing task (and the resulting student work) aligns with an objective standard at a specific moment in time. A rigorous essay adheres to grade-level expectations regarding the type and length of writing expected and the complexity of reading materials drawn
upon for evidence. If a student falls behind, the core strategy is practice of the complete form. Similarly, it is believed that complex texts from which students draw evidence should not generally be modified, even for students who read below grade level. Exposure to and grappling with complexity will, the thinking goes, foster mastery.

In progressive mastery, however, rigor is defined as the extent to which instruction and task requirements target the precise learning needs of specific students at a specific moment in time. Rigor is associated with a just-right instructional intervention that allows for sequential, progressive mastery of the critical small pieces that eventually allow students to meet grade-level standards.

Implications for future research

The current study is of writing curriculum at one high school using one version of progressive mastery – a particular set of strategies laid out in a specific sequence within one subject department. Future research should test the replicability of outcomes in other settings as well as the extent to which the strategies themselves and/or their repetition or sequencing accounts for the difference, as opposed to the progressive mastery approach itself. Anecdotal evidence suggests that progressive mastery improves writing for all student populations; but future research should seek to understand its differential impact, especially for those who have struggled most and about whom research is most lacking, as well as what adaptations work best in specific settings and for whom. In addition, some recommendations from the meta-analyses, such as peer collaboration and inquiry learning, were not present in this study. Further research should explore the extent to which these elements are in fact needed to improve writing at scale. Finally, this study suggests that progressive mastery holds promise for improving more than writing. The extent to which and the ways in which progressive mastery improves content knowledge, oral language, vocabulary acquisition, and reading comprehension are fruitful areas for future research.

Notes

1. See, for example, the National Council on Teacher Quality’s *What Education Schools Aren’t Teaching about Reading and What Elementary Teachers Aren’t Learning* – their 2006 report of a study that found that only 15 percent of US education schools provided future teachers with even ‘minimal exposure’ to the science of reading.
2. Not enough research exists to confirm this view (Applebee & Langer, 2009), but this is the opinion of the leading expert in writing research.
3. The demographics of New Dorp’s entering class have remained about the same since 2005: approximately 40 percent are poor, one-third are Hispanic, and 12 percent are Black. In 2006, 82 percent of freshmen entered the school reading below grade level. In 2007, 4 out of 10 students who began as freshmen had dropped out.
4. See *Strategic Inquiry* (Panero & Talbert, 2013) for a detailed description of this inquiry-based improvement process.
5. Note that the school’s ability to select and implement a school-wide writing approach that targeted specific, identified student needs resulted from culture and practice shifts brought about through a 4-year inquiry-based improvement process. The demands for teacher change and learning required to be willing to implement writing across the curriculum and to do so effectively are great, although not the focus of this study. See Panero and Talbert (2013) for detailed descriptions of the change process and the culture and practice shifts it brought about.
6. As a result of the writing focus, passing rates on high-stakes exams required for graduation jumped from 67 percent on the required English exam in 2009 to 89 percent in 2011 and from 67 percent on the required Global Studies exam in 2009 to 75 percent in 2011. The graduation rate rose from 63 percent in
2008 to 78.5 percent in 2011. By 2012, New Dorp brought more struggling students on track to college and graduation as measured by credit accumulation, high-stakes exam passage, and graduation rates than similar schools not using their improvement approach (see Panero & Talbert, 2013, Chapter 2, for a full description of the study and findings).

7. The exam is scored 1–5, with a 3 passing, 4 above standard, and 5 excellent. A 3 or above affords college credit.

8. This book documents findings from an evaluation of Strategic Inquiry, an inquiry-based school reform model, as implemented over 6 years in New York City. In this approach, teacher teams lead school improvement by identifying and closing student skill gaps, improving the instructional conditions that created them, and leading a growing number of colleagues to do the same. New Dorp High School was one of the schools studied up close (a case study site), but the report documents findings of inquiry teams in over 100 schools.

References


Getting the Big Idea: Concept-Based Teaching and Learning

“Transforming Learning Environments through Global and STEM Education” August 13, 2013

What is concept-based instruction?

Concept-based instruction is driven by “big ideas” rather than subject-specific content. By leading students to consider the context in which they will use their understanding, concept-based learning brings “real world” meaning to content knowledge and skills. Students become critical thinkers which is essential to their ability to creatively solve problems in the 21st century.

By introducing students to universal themes and engaging them in active learning, concept-based instruction:
- creates connections to students’ prior experience.
- brings relevance to student learning.
- facilitates deeper understanding of content knowledge.
- acts as a springboard for students to respond to their learning with action.

(Erickson 2008)

Why is it worth our time?

Concept-based instruction, by placing the learning process in the “big picture” context of a transdisciplinary theme, leads students to think about content and facts “at a much deeper level” and “as a practitioner would in that discipline” (Schill & Howell 2011). According to the International Baccalaureate Organization (IBO), teaching and learning that is driven by overarching concepts necessitates that students transfer their knowledge between personal experiences, learning from other disciplines, and the broader global community. Thus, concept-based instruction mandates more critical thinking at increasingly higher levels of Bloom’s taxonomy.

(Erickson 2012)

As we present it, concept-based instruction must begin with content skills and knowledge established by local standards and curriculum guides. To bring purpose to the content, the teacher plans learning activities that actively engage students in meaningful, “real world” concepts. These concepts could include skills, local issues, or values that might inspire students to act upon their learning. Relatively equal emphasis should be placed on both content and concepts throughout this process. A summative project or activity should be designed by the teacher to assess students’ mastery of the content as well as their ability to connect it to the “big picture” concept. This allows students to put their learning into action. An important reminder: concepts are not intended to replace content. Instead, concepts bring context and purpose to the content students are exploring.

- Josh and Joanne Edwards
What is a concept? (The concept-topic divide)

A common problem among teachers who want to bring concepts into their classroom is defining exactly what a concept is... and is not. An important distinction to note is the difference the topics that our curriculum mandates we include in our instruction, and the concepts that help connect that set of knowledge and skills to students’ lives.

According to Lynn Erickson, concepts are universal, timeless, abstract, and move students toward higher levels of thinking. Concepts are broad ideas that transcend the perspectives and limits of any specific subject-area. A concept is something that can be taught in any classroom, no matter what the content includes.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Human Body</td>
<td>Systems</td>
</tr>
<tr>
<td>Slavery</td>
<td>Oppression</td>
</tr>
<tr>
<td>Author’s Purpose</td>
<td>Perspective</td>
</tr>
<tr>
<td>Geometrical Translations</td>
<td>Change</td>
</tr>
<tr>
<td>Verb Conjugation</td>
<td>Relationships</td>
</tr>
<tr>
<td>Desktop Publishing</td>
<td>Communication</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Topics</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surrealism</td>
<td>Symbolism</td>
</tr>
<tr>
<td>Percussive Rhythms</td>
<td>Pattern</td>
</tr>
<tr>
<td>Stage Combat</td>
<td>Conflict</td>
</tr>
<tr>
<td>Team Sports</td>
<td>Communities</td>
</tr>
<tr>
<td>Sewing</td>
<td>Aesthetics</td>
</tr>
<tr>
<td>Horticulture</td>
<td>Sustainability</td>
</tr>
</tbody>
</table>

How do I choose concepts that are right for my teaching?

The concepts a teacher chooses to utilize will heavily depend on his or her content, the age, experiences, and diversity among the students, and personal goals and values for teaching. What is most important is that you, the teacher, are invested in helping students explore the “big ideas” you choose, and that the concepts you choose are relevant for your students.

There is also a variety of options you can consider in the kind of concepts you choose to use with your students. For example, you might choose content-specific concepts (still broader than topics) that easily connect to the information students are learning. This might be a great place to start, especially if you feel this approach to instruction will require some “getting used to.” Or, you might choose a set of thinking or learning skills that you want your students to master by the time they leave your class, like “intercultural awareness” or “persisting.” Another approach could be choosing broad, universal concepts that transcend all subject-areas. These universal concepts often have complex social implications that can lead to critical and reflective thinking among your students. (See the chart below for some specific examples.)

Whatever type of concepts you choose, consider ways that you can make them visible in your class’s physical space as well as the learning activities you plan.

<table>
<thead>
<tr>
<th>Content-centric concepts</th>
<th>Skill-centric concepts</th>
<th>Universal concepts</th>
<th>Course-long themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>In a Social Studies classroom, use the Five Themes of Geography as ongoing concepts that show up in each unit and get special attention in your teacher. Concepts would include Location, Place, Human-Environmental Interaction, Movement, and Region.</td>
<td>At the beginning of each week, introduce and discuss one of Costa and Kallick’s 16 Habits of Mind. Then, have your students reflect on how they utilized that Habit at the end of the week. The 16 Habits include Persisting, Listening with Understanding, and Empathy, Thinking about Thinking, and Applying Past Knowledge to New Situations, among others.</td>
<td>In each unit, use one of IBO’s newly published 16 MYP Key concepts to consider the broader impacts of the content students are learning. Do this by starting class discussions and using media to prompt debates about the meaning of a specific conceptual term, its repercussions in the “real world,” and how students think it connects to the content they are learning. The Key concepts include Change, Form, Identity, and Global Interactions, among others.</td>
<td>Developing the long-term theme of Global Issues, connect each unit’s content with a specific human rights issue around the world. Create a community service opportunity to give students the chance to act upon their discussions of these themes. You might choose to focus on any number of global or societal issues including access to sufficient drinkable water, human trafficking, labor conditions, and access to quality, affordable medical care.</td>
</tr>
</tbody>
</table>
How do I adapt my teaching strategies to include these concepts?

While each teacher will have a unique approach to implementing concept-based instruction, a 2011 article in *Science and Children* magazine outlines five basic steps to help teachers actually do concept-based learning with their students. Use this template, outfitted with the five steps, to start planning your concept-based unit.

**Teacher name:**

**Course title and grade level:**

**Dates for teaching this unit:**

1. **“Choose a topic of study”:** Start with the content your students need to learn. (Maybe choose a unit that you already want to change or re-develop?)

   **What content topic(s) will this unit include?**

2. **“Decide on a concept”:** Use the questions below to develop what might serve as a good concept for this unit.

   Thinking about this unit, what is the most important idea that you want your students to remember when they leave your class?

   IBO calls this the "enduring understanding."

   Try to summarize this “big idea” in one word.

   - Is this a concept, and not a topic?
   - Is this a concept that you value for your students to explore?

   Are there any other “big ideas” that fit this content well that you might also want your students to consider?

   **What concept(s) will this unit explore?**

   Did you find your “big idea” for this unit? If not, try choosing a “macroconcept” from a pre-established list that we’ve suggested, or create a concept map of the topic you are teaching and look for “big ideas” that emerge.

3. **“Develop essential understandings”:** What do you want your students to know by the end of the unit, on the factual and conceptual levels? By specifically writing down what you hope your students will learn about the content and the concept, you will clarify what you want to accomplish (making planning learning activities much easier!).

   Using “I can” statements, list the key learning objectives you have for your students during this unit. These should include content and concept.

   - I can...
   - I can...
   - I can...
4. **“Use inquiry-based investigations”**: Give your students the chance to use their prior knowledge to solve a problem or explore a new idea first. Often, the “big ideas” show up all on their own! Then, follow up by introducing the unit concept and the topic of study. Keep bringing class discussions back to this concept throughout the unit, where appropriate.

<table>
<thead>
<tr>
<th>What learning experiences will you use to help students connect the content and the concept in this unit?</th>
</tr>
</thead>
</table>

5. **Assess both content and concept**: By the end of the unit, give the students a chance to show you what they’ve learned. Either add a special activity that asks students to demonstrate their learning related to the unit concept, or blend this into your assessment of their content knowledge and skills by asking them to connect the two together.

<table>
<thead>
<tr>
<th>How will you assess student learning of the content, concept, and the connections between them?</th>
</tr>
</thead>
</table>

---

**“Unit Concepts”**

Divide the course up into specific units and choose one specific concept that fits well with that block of content. Plan lessons, activities, and class discussions that integrate this concept into students’ learning during this unit and find ways to connect the concept to assessment throughout this unit as well.

**“Core Concepts”**

Choose a core set of concepts that will be introduced early in the course and then addressed and re-addressed throughout the rest of the course. These reoccurring concepts can show up on a weekly or quarterly basis or even on an activity-by-activity basis.

**What will concept-based instruction look like in my classroom? How will my students and I use these concepts?**

**Blend these approaches!**

Find ways to use more than one of these methods to get your students thinking in context (or even, “in concept!”). Introduce your class’s core themes at the beginning of the course, then focus on one at a time during each major unit. As the unit concepts build, lead your students to a central, main concept that ties them all together at the end of the course!
THINKING BIG:
A Conceptual Framework for the Study of Everything

Our current fervor for highly specified standards for each academic discipline requires students to view reality as composed of fragmented and unrelated bits of information. Mr. Brady argues that what students really need is a system for organizing and integrating what they know so that they can understand the "big picture."

BY MARION BRADY

BUCKMINSTER Fuller once said, "American education has developed in such a way it will be the undoing of the society." Reading those words today, many may nod in agreement. Few, however, are likely to give the same reason as Fuller did for so bleak a prediction.

Fuller is most frequently remembered as the inventor of the geodesic dome — the lightest, strongest, most cost-effective enclosing structure ever devised. He was an inventive genius, but he was also a college professor, cartographer, philosopher, naval officer, mathematician, poet, researcher, cosmologist, industrialist, engineer, environmentalist, advisor to business and government, holder of 25 patents, author of 28 books, and recipient of 47 honorary degrees.

He aired his views on American education, including the judgment I quoted above, in the late 1980s in a speech delivered to a group that included college presidents. "What you fellows in the universities do," he continued, "is make all the bright students into experts in something. That has some usefulness, but the trouble is it leaves the ones with mediocre minds and the dunderheads to become generalists who must serve as college presidents . . . and presidents of the United States."

Generalists — people who strive to see the "big picture" — don't get

MARION BRADY is a retired teacher and school administrator and is currently a columnist for the Orlando (Fla.) Sentinel.

276 PHI DELTA KAPPAN 112
much respect in America. There is no listing for “Generalists” in the Yellow Pages, no places are reserved for them on the faculties of high schools and colleges, and no employment ads seek applications from them.

And what is the big picture today? Intensifying clashes on the “fault lines” between religions, societies, and civilizations; continuing threats of terrorism; a shrinking middle class and a widening gap between rich and poor; the confusing of national power with national greatness; dishonesty in boardrooms; violence accepted as entertainment; vast wealth plowed into no-return-on-investment armament and conflict; increasing environmental degradation; lobbyist-dominated legislatures; unwarranted confidence in the world-improving ability of force; official tolerance of tax evasion and a general decline in a sense of civic responsibility; and an education system beset by ideologically driven policies.

These related, big-picture issues are parts of a systemically integrated whole — a whole that the education establishment is not addressing. We send our graduates off with expertise in technology, banking, politics, medicine, law, and myriad other fields, staking our collective fate on their ability to manage crises as they pop up. But the old problems intensify and are joined by new ones.

That the education system we have created might, as Fuller said, actually be a major cause of those intensifying problems does not seem to have occurred to us. I can find little evidence of serious, ongoing dialogue among policy makers about the wisdom of continuing to educate students narrowly. Neither can I find evidence of concern about potential societal chaos when millions of narrowly educated experts pursue their professions with little or no understanding of how their actions interact.

Most of the college presidents and policy makers on the receiving end of Fuller’s blunt accusation probably went back to their respective institutions and did nothing. Those few who actually undertook instructional program changes probably played with course distribution requirements; organized or expanded interdisciplinary programs; focused instruction on projects, social problems, or themes; put a “capstone” course in place; or tried to skirt the issue by emphasizing “process” rather than “content.”

None of these actions do the job that needs doing. All these approaches to broadening students’ ability to deal with reality assume that the traditional academic disciplines are the basic organizers of knowledge and that the main task of educating is to introduce those disciplines to students. Wrong. The main task of educating is to help students make more sense of the world, themselves, and others.

Most educators no doubt share that aim and can provide myriad examples of their favorite discipline’s contribution to broad understanding. But that misses the point. The problem is not the failure of biology, economics, chemistry, psychology, and other school subjects to expand our understanding of reality. The problem is the failure of schooling to pull these disparate pieces together to address questions that few students articulate but that all, at some level, ask: What’s going on here? How did it come to be? How could it be different? How might I alter it?

For answers to questions like these, the disciplines are too narrow. They can be brought to bear on human experience, but neither individually nor collectively do they provide a coherent, holistic, comprehensive, hard-edged, and intellectually manageable conceptual framework that students can use to make more sense of daily life. The disciplines have different aims. They ignore much knowledge of great significance. Their vocabularies are arcane and often incompatible. Their methodologies are distinctive. They operate on different levels of generality and abstraction. They have long histories of competing for students and resources. Their practitioners have little interest in integrating and rarely even talk across disciplinary boundaries. Their support systems — bureaucracies, professional organizations, periodicals, funding sources, and so on — would oppose consolidation. Indeed, the trend is in the other direction: toward even greater fragmentation as disciplines split and split again into subdisciplines.

Up to a point, all of this is a good thing, for it is one way in which knowledge expands. But it is not the only way, and it is certainly not the way to build a comprehensive, intellectually manageable conceptual framework for organizing general education. What students need is a “master” organizer — a mental filing system or map they understand — that displays the general layout of the mind and its system for integrating knowledge. That organizer will recognize, encompass, and relate the disciplines, but it will not be fashioned from them.

AN ALTERNATIVE ORGANIZER OF KNOWLEDGE

ology. Zoology. We take the compartmentalization of knowledge for granted. So deeply ingrained is the idea that knowledge is best transmitted to the next generation in neat, discrete packages, that we can hardly imagine schooling organized in any other way.

This is a recent and curious assumption. Throughout recorded history, long before the academic disciplines took shape, great, noncompartmentalized minds were at work in ways still admired. We may retroactively place Socrates, Jesus, Leonardo, Galileo, Benjamin Franklin, de Tocqueville, and hundreds of other respected thinkers within today’s disciplinary categories, but that is not how they saw themselves. Indeed, if most of the great minds of the past had been forced to function within the arbitrary boundaries of our narrow academic disciplines, we might never have heard of them.

Nearer at hand and hardly less impressive examples of noncompartmentalized thinking are the intellectual accomplishments of small children. Their formal introduction to school subjects lies years in the future, and yet, starting in infancy, they learn to organize and integrate knowledge in situations whose complexity we fail to appreciate only because the situations are so familiar.

But for perhaps the most convincing evidence that complex thought processes can be independent of the disciplines, try introspection. Your moment-by-moment functioning requires you to continuously select, organize, and integrate enormous amounts of information. You do not do that by mentally moving between the various academic disciplines. You use a different, far more sophisticated, seamless approach.

We have created a way of life that makes discipline-based, specialized study essential. But for general education — for the task of managing our individual and collective affairs and exercising some control over the future — fragmented, specialized studies are not enough. There needs to be in place an overarching framework enclosing the disciplines, expanding them, filling the gaps between them, putting what appears to be random information in context, providing perspective, setting priorities, displaying the whole that the disciplines illumine in part.

Scholars have always insisted that the familiar curriculum of separate subjects is unacceptable. Daniel Tanner of Rutgers University put it this way: “All of our experience should have made it clear by now that faculty and students will not derive from a list of disjointed courses a coherent curriculum revealing the necessary interdependence of knowledge.”

The University of Washington’s John Goodlad said: “The division into subjects and periods encourages a segmented rather than an integrated view of knowledge. Consequently, what students are asked to relate to in schooling becomes increasingly artificial, cut off from the human experiences subject matter is supposed to reflect.”

And Stanford University’s Paul DeHart Hurd phrased it: “Beyond tradition there are neither philosophical nor psychological grounds for compartmentalizing knowledge into islands of information that are presumed to represent academic disciplines as school subjects are currently conceived.”

We have used the regularities and relationships we see in various parts of reality to create the academic disciplines. We should now use the larger regularities and relationships in reality to integrate disciplines, subjects, courses, topics, themes, social problems, projects, and other approaches to organizing instruction.

We all have a deeply embedded, probably hard-wired, system for “thinking big.” If that system is not hard-wired, it is certainly a cultural regularity with too much instructional potential to ignore, for it is a key to solving a fundamental curricular problem.

OUR ‘NATURAL’ ORGANIZER OF KNOWLEDGE

It is reasonable to suppose that a knowledge-organizing and integrating system used by everybody, every day, will be neither esoteric nor inaccessible. And indeed it is not.

Years ago, convinced of this fact, I began asking students to engage in a very simple activity designed to help them make explicit their implicitly understood system for organizing knowledge. I put them in small groups, told them to think about “an ordinary bit of reality — this room, right now,” and instructed them to write down as many facts about it as they could manage in five minutes. That done, I asked them to combine their lists and devise a system of categories and subcategories.

I have done this exercise with students from early adolescence through adulthood. Without fail, a five-part category system similar to that shown in Figure 1 eventually emerges as the major organizer. To make sense of experience — an event, situation, or condition — we set it off from other experience and lay our five-element template over it. We locate it in time and physical space, identify the participating actors, and examine
social patterns and assumptions that influence it.

These five concepts are the basic organizers of our mental models of reality. All knowledge “fits” within their collective boundaries. The traditional academic disciplines elaborate and explore random conceptual branchings of the five, but only familiarity with the systematically integrated roots from which those conceptual branchings spring enables students to capitalize on their near-infinite relational possibilities. This, not interdisciplinarity, is the key to the integration of everything the student knows.

There can be no acceptable general education without a discipline of general education, and there can be no coherent general education discipline without a framework of logically related organizing ideas. The scheme in Figure 1 or a conceptual framework very much like it is the foundation on which a general education curriculum can be built. It encompasses all knowledge. Unlike the disciplines, its five elements relate systemically. It calls attention to currently neglected areas of study and to future possibilities. It provides criteria for content selection and emphasis. It makes clear the mutually reinforcing nature of the traditional academic disciplines (but also their conceptual randomness). It is an extraordinary aid to memory. It facilitates the expansion of knowledge by helping students envision possible relationships between various aspects of reality. It automatically adjusts to student ability. Being conceptual rather than factual, it constantly adapts to change. Not inconsequentially, it raises the level of sophistication of the curriculum beyond the easy reach of well-meaning but ill-informed politicians and other education policy makers.

Moreover, this knowledge-organizing and integrating system does not even have to be taught. It is in constant use by every student. An instructor’s job is merely to raise it in the students’ consciousness from that which is implicitly known to that which is explicitly known. When students move from knowing to knowing what they know, they are able to perform at intellectual levels far beyond expectations.

TWO THEORIES

The present curriculum, made up as it is of separate, specialized studies, exerts considerable pressure on teachers to make major use of what could be called “Theory T.” Theory T dominates American education. It reflects the conventional wisdom about educating, underlies much of the education-related legislation handed down by federal and state governments, drives the current top-down “standards and accountability” fad, gives the publishers of textbooks and standardized tests a central role in the making of education policy, and is demonstrated daily in hundreds of thousands of classrooms — public, private, charter, parochial, and virtual — as well as home school settings.

T stands for “transfer.” Those who accept Theory T believe that knowledge is located in teachers’ heads, textbooks, reference materials, and on the Internet and that the instructional challenge is to transfer it from these locations into the empty space in students’ heads.

The degree of success of the transfer process can be measured with relative ease, which helps explains its broad appeal. At some agreed-upon interval after instruction — usually a few weeks or months — students are asked to recall what they have learned or at least to recognize a correct version of it when it is placed along-

---

**FIGURE 1.**
A Model of Reality

- **Time**
  - Point
  - Duration
  - Etc.
- **Setting**
  - Climate
  - Terrain
  - Etc.
- **Actors**
  - Number
  - Distribution
  - Etc.
  - Work
  - Worship
  - Etc.
  - Nature
  - The Supernatural
  - Etc.
- **Social Patterns**
- **Assumptions**

Further elaborated, in part, by the traditional academic disciplines

DECEMBER 2004    279

115
side other reasonable-sounding but false statements. Evaluating performance is simple enough to allow student responses to be scored by a machine.

In the traditional curriculum, the paraphernalia of the specialized disciplines encourages the use of Theory T. Each field has its own vocabulary, conceptual framework, methodologies, and so on. These must be mastered, at least to some degree, before the discipline can be used as a tool to make sense of the world. Since students are required to work with several disciplines at once, even gaining a minimal familiarity with them leaves little time for the kind of real-world applications that lead to a useful level of understanding. However, when students need only master a single conceptual framework for the whole of reality and when that framework is the one they are already using, they can immediately put the discipline to work.

Some educators — certainly a minority — reject Theory T, subscribing instead to what might be called “Theory R.” Socrates appears to have been among their number, although it is unlikely that all those who admire his teaching style buy its underlying premise.

Theory R and our five-element model of reality are highly compatible. Theory R assumes not that students’ heads are empty but that they are full. The primary instructional challenge, then, is not to transfer new knowledge but to help students reorganize existing knowledge to make it more useful, consistent, or true and to supplement it with insights and skills that will help explain more fully what they already know.

This rearranging process is complex, but far less so for students in Theory R than in Theory T classrooms. Students in Theory T classrooms are passive absorbers of information. Students in Theory R classrooms must be active processors of information. Theory T emphasizes recall; Theory R requires students to engage in every known thought process. Theory T feedback to teachers and peers is simple and straightforward, coming in the form of either correct or incorrect responses. Theory R feedback is often subtle and difficult to analyze, coming in the form of body language, dialogue, and other indicators from which student thought processes must be inferred.

Theory R requires students to make connections, to perceive relationships, and to synthesize ideas. It sends students searching the far corners of their minds without regard for the artificial, arbitrary boundaries imposed by academic disciplines. Therein, incidentally, lies a major and perhaps unexpected benefit of having in place a conceptual map like the one in Figure 1. As any attempt to remember today the details of what we once learned in school will demonstrate, the human brain is very poorly equipped to store and retrieve random information. Memory needs the help of logic, and logic requires the use of a system of organization.

To illustrate that general education deals with the kind of content and skills that Buckminster Fuller promoted, let me share an example of an assignment for middle or high school students. It can properly be called “general” because it deals with matters of potential significance for life as it is lived, requires the use of a full range of thought processes, and touches on matters that do not fit neatly into existing subjects or courses.

One of the many matters that has probably gotten less attention in your education thus far than it deserves has to do with why and how the world changes and the nature of the future. Here are several events, ideas, innovations, or policies that could have consequences beyond those which are immediately apparent. Choose one, and construct a flow chart showing its probable or possible consequences over time. Extend the chart at least to a point at which a valued local way of acting and/or thinking begins to be threatened.

a. A 3° Celsius increase in average annual temperature.

b. A five-gallon-a-day limit per individual on fresh water consumption.

c. A $10 VSA (Voice Stress Analyzer), portable, about the size of a pack of cigarettes. (Voice stress analyzers indicate by a light or another kind of signal when a speaker is probably lying. They work directly or by telephone, radio, or television.)

d. An antipollution law prohibiting the use of automobiles for commutes of less than one mile for all able-bodied persons.

e. A mandatory year of public service at age 18 or after completion of high school.

f. A heavy tax on automobiles, calculated according to the number of square feet or meters of road surface occupied.

g. Zoning regulations that encourage a return to the colonial-era arrangement of combined living and working spaces in urban areas.

Students in Theory T classrooms would not expect this assignment unless each of the items had been “covered” in class. However, students familiar with a conceptual framework for organizing and integrating knowledge would consider the assignment perfectly reasonable, even if the items had never been mentioned in
class. They would simply lay their five-element-model-of-reality template over the chosen item and begin to generate hypotheses about probable and possible causal relationships suggested by the template’s myriad “files and subfiles.”

The breadth of a true general education discipline and the specificity of the elements of such a discipline make it easy to endlessly generate instructional activities that even the most skeptical or jaded student will see as undeniably relevant and that all students, even the most brilliant, will find intellectually challenging.

**WHAT SHOULD WE DO?**

If Buckminster Fuller were alive today, he would surely accelerate his timetable for “the undoing of [American] society.” Today’s major education-related debates — about vouchers, choice, competition, merit pay, rewards, school shaming, discipline-based standards, high-stakes tests, accountability, privatization — do not even hint at the problem to which he was calling attention. No major participant in those debates is raising a single question about the aims of education, its proper scope, the validity or relative importance of particular standards, or the deeper meanings of “quality.” It is being assumed, wrongly, that the institution is basically sound, that it merely needs a tune-up, which can be provided by the play of market forces.

But bringing market forces to bear will not improve education. Indeed, present federally mandated “reforms” will do just the opposite; they will freeze even more rigidly in place the very curriculum — aimless and fragmented — that prompted Fuller’s warning. The mandates attempt to force teachers and students to do the wrong thing better.

For proof, one need only look at the so-called standards and the high-stakes tests being used to measure progress toward meeting those standards. Knowledge is “all of a piece.” Humans learn seamlessly. But the thousands of standards being put in place ignore this fact. Those who wrote them for various subjects obviously did not talk to one another, much less recognize the systemically integrated, mutually supportive nature of knowledge. The result is the perpetuation of an intellectually unmanageable, “mile wide and inch deep,” artificially compartmentalized curriculum, a curriculum acceptable not because it is theoretically sound, not because it is intellectually challenging, not because it meets individual or societal needs, but because its familiarity blocks recognition of its fundamental inadequacy.

Public education is imploding, at a rate accelerated by the reactionary education legislation coming from various levels of government. Even the best schools are squandering their students’ potential at a prodigious rate. It is far past time for leaders — principals, superintendents, college presidents, union officials, school board members, trustees — to realize that they must act, that teachers, curriculum coordinators, department heads, and others traditionally in charge of deciding what is taught are ill equipped by background, training, job description, and inclination to think about the whole of which their narrow areas of expertise are parts.

True leaders will hammer federal and state legislators until the innovation-stifling provisions of present legislation are repealed. They will then work with their faculties and other stakeholders to create a true general education curriculum, a curriculum that respects human nature and the brain’s holistic approach to making sense of experience.

Buckminster Fuller was dead right. American education has developed in a way that, left unchanged, will be the undoing of the society. But even educators and policy makers who do not accept Fuller’s bleak prediction are not off the hook. Schools are in the knowledge business. Any school that does not send its graduates off with a thorough understanding of the seamless, systemic nature of knowledge — and the ability to use that understanding to live life more fully and intelligently — is failing.

Rethinking Instructional Delivery for Diverse Student Populations:

Serving All Learners with Concept-Based Instruction

Jan D. McCoy and Leanne R. Ketterlin-Geller

When students arrive at content classes reading below grade level, teachers are challenged to deliver complex content. Also, students often study facts without reaching larger concepts. Research at the University of Oregon has concluded that if the teacher takes responsibility for identifying and elucidating the concept within course materials, both of these difficulties can be overcome. Overt identification of concepts and their characteristics and the deliberate use of graphic organizers reduce the reading comprehension demands placed on students with low abilities. Using the functional taxonomy presented, teachers can develop effective student exercises and assessments. An example shows that students provided with the concept-based approach outperformed students in a more traditional classroom on a problem-solving task.
Any students arrive at middle and secondary content classes reading significantly below grade level. Whether a result of learning disabilities, language barriers, or low skill levels, struggling readers often focus on decoding text at the expense of understanding material (Jitendra et al., 2001). Because textbooks are commonly used in content classes, students face serious difficulties attempting to understand the material. The diversity of student characteristics and learning needs can present seemingly innumerable challenges to general education teachers as they deliver complex content material in a diverse classroom. It follows that student learning is potentially jeopardized because of inappropriate instructional materials and overburdened teachers who are unable to serve the needs of all students.

To help students challenged by heavy reading requirements in content classes, text-related demands must be reduced without compromising the content. Addressing this issue has been the focus of ongoing work at the University of Oregon's Behavioral Research and Teaching (BRT) facility. We have been successful in redirecting curriculum and assessment practices to take into account not only the difficulties both skilled and struggling readers face in learning content material but also the instructional techniques teachers have employed for some time. Our goal is to increase student comprehension and retention through instructional modifications rather than tolerating the loss of instructional time by waiting for students’ reading abilities to improve.

**Basis for Research**

Teachers can address the needs of individual students with varied abilities and backgrounds by drawing from a smorgasbord of techniques and strategies, but this is taxing for teachers and may well result in confusion for the students. Thus, teachers need a strategy that supports students with cognitive disabilities and specialized learning needs while still advancing the learning of other students in the classroom.

Early research at BRT indicated that many students in general education classrooms, even those who test well in decoding and comprehension, have difficulty prioritizing information within the text (Tindal, Nolet, & Blake, 1992). This supports conclusions drawn by other researchers in the field (Bryant, Ugell, Thompson, & Hamill, 1999; Garner, Gillingham, & White, 1989) and highlights the critical need for content-area reading supports. Students are often distracted by alluring or seductive details that, although interesting, may not be as instructionally important as other information in the text (Garner et al., 1989).

The difficulty many students face in trying to isolate important information from the text is easily traced to the nature of textbooks. Both researchers and reviewers have noted that most information in textbooks is purely factual and fails to show how discrete facts can be linked together to form complex knowledge forms (Jitendra et al., 2001). Our research indicates that only high-achieving students ever arrive at a full understanding of the underlying concepts and principles from the content presented in texts. This shortcoming among textbooks leaves the responsibility for drawing these connections to either the teacher or the students themselves.

**A Concept-Based Model**

Because most students cannot link facts to concepts, classroom teachers must take responsibility for identifying concepts within the curriculum, an explicit and overt act. The following concept-based instructional model was designed to accommodate students with various learning needs while raising the expectations and performances of all students. For some time, other researchers have been investigating issues related to concept-based teaching with similarly positive results (Erickson, 1998; Roid & Haladyna, 1982). Our model draws on this earlier work but is characterized by three distinct components:

1. The teacher determines the concept that is the target of instruction.
2. A graphic organizer is developed to illuminate this concept for the students.
3. Students’ success in mastering the concept is measured by applying it across instances using increasingly complex critical thinking measures.

Under the model, we define a concept as consisting of the following:

- a broad class of objects or events bearing a unique label and spanning multiple instances and examples (e.g., revolution vs. the American Revolution);
- attributes (i.e., those elements of the concept that fully describe the concept and separate it from similar concepts); and
- examples and nonexamples that illustrate the attributes relative to a specific instance.

A key component in this effort is identifying the defining attributes for the concept. Other researchers have made use of concept-based models or graphic organizers and have reported some success. The model described here relies heavily on including attributes that can generalize to multiple instances of the concept; other models depend merely on defining the concept exclusively with a collection of example facts (c.f. Boudah, Lenz, Bulgren, Schumaker, & Deshler, 2000; Erickson, 1998; Kameenui
& Carnine, 1998). Generalizing concepts is the primary strength in our instructional model, and the attributes greatly enhance this generalizability across multiple instances.

The graphic organizer in Figure 1 for the concept "world war" illustrates this model. Instances of a world war include World War I, World War II, and the Cold War. The attributes of a world war used in this lesson sequence are technology, alliances, and dominance. These attributes and their associated examples (from the instance of the Cold War) are presented on the graphic organizer. This unit, presented in an 11th-grade American history class, helped all students to better understand the associations and relationships among World War I, World War II, and the Cold War (i.e., shared aims among alliances, striking improvements in technologies, and universal efforts at and reversals of domination) and dispelled many previously held misconceptions (see Figure 2, for an example of student work). Thus, isolating the critical concepts allows both the teacher and the students to focus on information that is necessary for understanding the domain of instruction (Voltz, Brazil, & Ford, 2001). All students will benefit from such explicit organization of important knowledge.

In addition to the teacher’s effort in identifying the essential concepts for the instructional unit or lesson, graphic organizers, such as the one presented in Figure 1, are essential elements of concept-based instruction because they explicitly illustrate the structure and organization of information (Hudson, Lignugaris-Kraft, & Miller, 1993; Tyrice, Fiore, & Cook, 1994). By constructing a visual display of relevant content material, students are able to link prior knowledge with new learning, thereby deepening their level of understanding the material (Bryant et al., 1999; Merkley & Jeffries, 2000/2001).

To design a graphic organizer, the teacher identifies the central concept that will be the focus of instruction. Associated attributes and their relationship to the concept are then determined. Next, a visual display of the information is constructed that explicitly links the attributes with the concept in a manner that highlights the relationships. Finally, space is provided for recording examples and nonexamples for the attributes. This design enables students to discern important facts from statements that may be interesting but that are not integral to the concept. Thus, all students are presented with a model for drawing meaning from the text, identifying concrete examples, and distinguishing between seductive details and critical facts. Ellis (1994) presented designs that showed cause-and-effect relationships, compare and contrast, and sequential organization of information. Students with disabilities or other reading difficulties will benefit substantially from a visual representation of the structure and organization of relevant information (Hudson et al., 1993; Jitendra et al., 2001).

These two elements—overt identification of subject-specific concepts and the creation and completion of graphic organizers—are central to the model. Carefully planned assessments aligned with the instructional content and goals effectively measure student acquisition of concept-related materials. This does not suggest that teachers abandon existing measures but rather that they consider alternatives appropriate to the concept addressed rather than the factual information presented in the textbook. Higher-order thinking skills can be measured across the student population when assessments are aligned to concepts rather than to facts.

**Aligning Assessments**

Assessment tasks in this model are based on a functional taxonomy similar to that proposed by Benjamin Bloom in the 1950s (University of Victoria, 1996). Here, however, the levels of intellectual engagement are operationalized to provide clear opportunities for students to demonstrate their levels of accomplishment (see Table 1). For example, in Bloom’s taxonomy, instructional objectives might require students to comprehend or analyze the ancient Greeks’ influence on the development of democratic models and institutions. The taxonomy presented in this model would ask that students summarize or illustrate instead. The latter acts are directly measurable compared to Bloom’s taxonomy, which asks the teacher to determine what sorts of tasks might best reflect comprehension or analysis.

Applying concept-based instruction, teachers can use the targeted intellectual operations to develop and implement student practice exercises and assessments. Generally, assessments for this model are open-ended in an effort to move students’ thinking beyond simple recall. It is difficult, although not impossible, to assess and evaluate students’ ability at the more advanced levels of this taxonomy using multiple-choice or short-answer questions. Therefore, we have focused our attention on open-ended essays scored using a scoring guide. Despite the greater time demand involved in scoring, most teachers are amenable to assessments of this type for higher-level intellectual operations; they believe that the significant improvement in the amount of information such assessments return on student ability is worth the extra effort at scoring.

**An Example**

This model of concept-based instruction has been widely implemented and studied in middle and secondary schools in a variety of content areas in our geographic region. To illustrate the potential benefits this model brings to teachers and students, we will briefly present our experience in two sixth-grade social studies classes studying the culture and history of Meso-Americans. A technical report on
Figure 1. Graphic organizer for world war (Cold War examples).
Prompt
1. Describe the role technology played on the German war plans and the blitzkrieg strategy.
2. How did the alliance of Great Britain, the U.S., and the U.S.S.R. help the allied powers take back France following the D-Day invasion?

Student Response (transcribed with errors)
1. A. The battle was affected by alliances. The US sent our left over war material to Great Britain.
   B. The battle was affected by the higher advanced radar system—and the British had better planes.
   C. It was also affected by the strategies of Britain. They blocked out towns at night, and the Germans couldn’t blitzkrieg!
2. A. Germany could’ve won if Italy would’ve surrounded Britain.
   B. If Germany had a better navy and radar system they could’ve won.
   C. If Germany had attacked great Britain and overwhelmed them sooner and faster Germany could’ve won.

Figure 2. Student work sample related to world war.

Table 1. Taxonomy of Intellectual Operations

<table>
<thead>
<tr>
<th>Intellectual Operation</th>
<th>Student Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reiteration</td>
<td>A verbatim reproduction of material that was previously taught.</td>
</tr>
<tr>
<td>Summarization</td>
<td>Generation or identification of a paraphrase, rewording, or condensation of content presented during instruction.</td>
</tr>
<tr>
<td>Illustration</td>
<td>Generation or identification of a previously unused example of a concept or principle.</td>
</tr>
<tr>
<td>Prediction</td>
<td>Description or selection of a likely outcome, given a set of antecedent circumstances or conditions that has not previously been encountered.</td>
</tr>
<tr>
<td>Evaluation</td>
<td>Careful analysis of a problem to identify and use appropriate criteria to make a decision in situations that require a judgment.</td>
</tr>
<tr>
<td>Explanation</td>
<td>Description of the antecedent circumstances or conditions that would be necessary to bring about a given outcome.</td>
</tr>
</tbody>
</table>

The purpose of the study was to explore the effects of using concept-based instruction on student understanding and retention of text material. The students of two classrooms, taught by Leroy and Daphne, were chosen to participate. Students in these two classes ranged from the 3rd to the 98th percentile rank on a statewide, standardized assessment of reading with an approximately equal distribution across classes.

Leroy structured his lesson in a traditional lecture format. Daphne, on the other hand, presented information using the concept-based approach, incorporating the graphic organizer shown in Figure 3 as an organizational strategy for the content of the textbook. Daphne identified civilizations as the key concept with religion, social groups, support activities, and writing as supporting attributes. Both teachers used the same textbook as the primary source of content information. At the end of the unit, students’ understanding of the content materials was assessed using a traditional fact-based reiteration test and two open-ended assessments. One transfer task required the students to create a map of an imaginary culture’s island habitat that comprised characteristics and features necessary to the functioning of civilizations. The second open-ended assessment asked students to write a description of their map, including an explanation of the features shown on the map and the significance of their relative placement. The scores from these various assessments were compared to determine the effectiveness of the instructional model.

Students presented with the concept-based approach performed much better on the open-ended assessments that required them to illustrate the concept in a new instance, demonstrating higher-order thinking. For example, Fred, a student ranked in the 14th percentile on the statewide reading assessment and identified as an English language learner, clearly demonstrated a thorough understanding of the concept and attributes. His response to the second open-ended assessment task is presented in Figure 4. The essay in Figure 5 was written by Daisy, a student ranked in the 90th percentile on the statewide reading assessment. Both essays illustrate that each student mastered the content.

Students who received instruction using the concept-based approach and who organized information from the text into a graphic organizer performed as well on the reiteration test as the students who received a more traditional instruction. This is an important result. A common concern among novice users of concept-based instruction is the potential loss of important factual instruction in favor of the more ethereal concept-based focus. Given these results, this appears not to be the case.

Throughout our research, our findings indicate that students across the spectrum of reading skills show consistent improvement toward mastery of higher-order thinking tasks when taught using the concept-based approach.
I found my island in the Pacific Ocean People all ready lived there. They
heard that I have read about the early civilization of amicas-omecs,
Mayas, Aztecs, and Incas. They think I know about four important feature
of civilization. Peoples belief was that the temples were religion. People
lived together by building a small village so they can support them-
selves. By the water the put a farmed land farmers go fishing a lot that's
also another thing they support them selfs they had trails to go from one
city to the other. People communicate with writing language that is
called hiyoglyphics that how they know how to get from one village to
the other by putting signs with hiyoglyphics.
SURGE ISLAND

I have been appointed King of Surge Island. The waves here surge into the coastline of the island giving it its name. The civilization here has not yet developed as they (the people) have asked me to come up with some advice for our undergoing religion, written language, how to live and how we can work together to stay alive. Here are my ideas:

RELIGION & GOVERNMENT

In religion, everyone within our empire will worship the God of Gods, the Sea God. Each year prisoners of war will be sacrificed to the Gods, their blood will keep the Gods happy. In every city there will be a temple where people will hold religious ceremonies, in honor of our God, the "Sea God".

The government will basically be the king. What ever the king says goes. New kings will be appointed dies or is no longer able to make wise decisions. Each king that dies will have a proper burial sight. This sight will be a small temple or pyramid which ever is suitable at the time. When the king is being put in his temple or pyramid a priest will lead a religious prayer for the king.

HOUSING

Since our land is some what marshy, I think it would be in our best interest to make our houses out of reeds. We would have to poke the reeds into the ground in the form of a round circle, then lay some reeds on top for a roof and tie it to the circular bottom, with the stalks left over from maize.

The people of our civilization will live in different cities or clusters of houses. The temple where people will hold religious ceremonies will be in the center of the city or the center of the cluster.

OCCUPATIONS

Some people will become peasants, and will begin farming or working on the farms. Their primary crops will be beans and maize. One eighth of all crops grown will become the king's, to serve him and the people within the kingdom.

We will probably have to dike up and build canals for the rivers. This will allow for better irrigation, and also help control the water, so we don't have to worry about flooding in our cities. On the map I have shown where dikes and canals are to be made.

Some people will need to take up a career in craftmaking. This is so when over-seas travelers want to trade, our craftworkers will be able to trade or sell sea shell necklaces, bracelets, and other items made from the island's shore! Also, some people will need to become scribes. That way each city within our empire can communicate through a system of writing.

Figure 5. Daisy's description of her island transcribed (with errors) from student's handwriting.

Conclusion

The outcome described in this example is typical of the results across our efforts to implement and test this model. Not only do students consistently perform as well as before on lower-level intellectual operations, they also reliably outperform students receiving traditional instruction on assessments evaluating more advanced intellectual operations. This result is not surprising given that most textbooks present information at the lowest level of intellectual operations, leaving advancement beyond this point up to the teacher.

Faithfully implemented, this model appears well able to accommodate the diversity of learning and reading skills found in contemporary classrooms without compromising expectations of student learning. Indeed, strong evidence has suggested that all students, including those identified as special needs and as talented and gifted, taught with this model can be expected to succeed not only in mastering the facts that fill their textbooks but also in demonstrating skill in applying a deeper understanding to unfamiliar instances of the same concept.

ABOUT THE AUTHORS

Jan D. McCoy, PhD, is an assistant professor of educational leadership and a research associate with Behavioral Research and Teaching at the University of Oregon. His current research interests include integration of students with special needs into general education classrooms through a more accessible curriculum and technology supports for student instruction. Leanne R. Ketterlin-Geller, PhD, is the director of research projects for Behavioral Research and Teaching and assistant professor of educational leadership at the University of Oregon. Her current research interests include universal design for assessment and instruction, curriculum-assessment alignment, and inclusion of students with disabilities in general education services. Address: Jan D. McCoy, 5262 University of Oregon, College of Education, Eugene, OR 97403; e-mail: jmccoy@darkwing.uoregon.edu

AUTHORS’ NOTES

1. Research reported here is part of an ongoing program of research at the University of Oregon's Behavioral Research and Teaching group under the direction of Dr. Gerald Tindal.
2. We thank the staff of Lincoln Middle School and Cottage Grove High School in the South Lake School District, Cottage Grove, Oregon, and Shasta Middle School in the Bethel School District, Eugene, Oregon, for their participation in and support of this research.
3. This research is supported under a grant from the U.S. Department of Education, Office of Special Education Programs (CFDA No. 84.3250, Award No. H325D990076).
REFERENCES


Concept-based teaching and learning
H Lynn Erickson
Introduction to IB position papers

This paper is part of a series of papers, written by IB practitioners and endorsed by the IB. Each paper addresses a topic or issue related to the IB’s philosophy or its educational practices.

Other papers in the series

Allan, M. May 2011. Thought, word and deed: The roles of cognition, language and culture in teaching and learning in IB World Schools.


Marshman, R. July 2010. Concurrency of learning in the IB Diploma Programme and Middle Years Programme.

Walker, G. October 2010. East is East and West is West.
Abstract

This paper examines the characteristics of concept-based curriculum and instruction models and identifies the International Baccalaureate (IB) programmes as a three-dimensional, concept-based model. A discussion of the benefits of concept-based instruction supports the majority of attributes in the IB learner profile. Concept-based instruction requires an understanding of synergistic thinking, transfer of knowledge and social construction of knowledge. This paper addresses these areas and discusses them in the context of the required IB pedagogy. It concludes with a review of the challenges in implementing a concept-based model and a summary of the rewards.

Introduction

The International Baccalaureate programmes offer a design for curriculum and instruction that is more challenging than traditional models, but which can produce deeper intellectual and emotional engagement in learning. The concept-based design is fully supported by cognitive and learning research. When information today is a click away on a computer keyboard, the use of classroom time must shift focus from covering and memorizing information to thinking with and applying knowledge at both the factual and conceptual levels. Thinking deeply with factual knowledge and concepts to communicate ideas and solve problems, transferring knowledge across distinct global contexts and situations, and seeing patterns and connections between concepts, ideas and situations are at the heart of concept-based teaching and learning. Less factual coverage can open the door to deeper thinking and understanding.

What is concept-based curriculum and instruction?

Concept-based curriculum and instruction is a three-dimensional design model that frames factual content and skills with disciplinary concepts, generalizations and principles. Concept-based curriculum is contrasted with the traditional two-dimensional model of topic-based curriculum which focuses on factual content and skills with assumed rather than deliberate attention to the development of conceptual understanding and the transfer of knowledge (see Figure 1).

![2D CURRICULUM/INSTRUCTION
 TOPIC/SKILL-BASED MODEL](image)

Concept-based teaching and learning
Two-dimensional curriculum models focus on facts and skills with the goals of content coverage, analysis and the memorization of information. Three-dimensional models focus on concepts, principles and generalizations, using related facts and skills as tools to gain deeper understanding of disciplinary content, transdisciplinary themes and interdisciplinary issues, and to facilitate conceptual transfer through time, across cultures and across situations. Three-dimensional models value a solid base of critical factual knowledge across the disciplines, but they raise the bar for curriculum and instruction by shifting the design focus to the conceptual level of understanding. This focus necessarily requires a supporting role for factual knowledge.

A corollary goal of concept-based instruction that is seldom stated overtly is development of the intellect. In a concept-based instruction model teachers use the facts in concert with concepts and generalizations to effect higher order, synergistic thinking. Facts provide the foundation and support for deeper, conceptual thinking and understanding. Three-dimensional concept-based curriculum models value student inquiry and constructivist learning to support personal meaning-making.

The research and agreement on the importance of conceptual understanding is undeniable. From the National Council of Teachers of Mathematics (NCTM) (2009) we hear the call:

Any national mathematics curriculum must emphasize depth over breadth and must focus on the essential ideas and processes of mathematics (p 1).

...research on the learning of complex subjects such as mathematics has solidly established the important role of conceptual understanding in the knowledge and activity of persons who are proficient (p 2).

One of the strongest research summaries supporting the importance of conceptual understanding can be found in How People Learn: Brain, Mind, Experience and School (Bransford et al 2000), published by the National Academy of Sciences and the National Research Council.

Experts' knowledge is connected and organized around important concepts (eg, Newton's second law of motion) (p 9).

To develop competence in an area of inquiry, students must: a) have a deep foundation of factual knowledge, b) understand facts and ideas in the context of a conceptual framework, and c) organize knowledge in ways that facilitate retrieval and application (p 16).
... organizing information into a _conceptual framework_ allows for greater _transfer_; that is, it allows the student to apply what was learned in new situations and to learn related information more quickly (p 17).

Anderson and Krathwohl’s book (2001) updated Benjamin Bloom’s _Taxonomy of Educational Objectives_ (1956), and further supports the need to teach for deeper conceptual understanding.

By separating factual knowledge from conceptual knowledge, we highlight the need for educators to teach for _deep understanding of conceptual knowledge_, not just for remembering isolated and small bits of factual knowledge (p 42).

Students _understand_ when they build connections between the “new” knowledge to be gained and their prior knowledge. More specifically, the incoming knowledge is integrated with existing schemas and cognitive frameworks. Since concepts are the building blocks for these schemas and frameworks, _conceptual knowledge_ provides a basis for understanding (p 70).

Beyond the research, the importance of conceptual structures for disciplinary content just makes logical sense.

**In what ways are IB programmes concept-based?**

The Primary Years Programme (PYP), Middle Years Programme (MYP) and Diploma Programme (DP) are three-dimensional and concept-based because, by deliberate design, they require students to process factual knowledge through the conceptual level of thinking.

The DP values deeper critical thinking and conceptual understanding. The required theory of knowledge course examines different ways of knowing and challenges students to think beyond the facts as they analyse complex questions and issues in interdisciplinary inquiries. The extended essay and a variety of internal assessment tasks also engage the critical thinking of students as they independently plan, research, write and defend a significant question drawn from one of the subject areas (IB 2009b).

DP teachers ensure that students know the attributes and meaning of the subject area concepts. At times, however, teachers express a concern over the tension between a heavy curriculum load and the time to teach for deeper conceptual understanding and the transfer of knowledge. This tension can sometimes result in implicit rather than explicit demonstrations of understanding. As the DP continues to develop, this tension might ease if key concepts and disciplinary related concepts were used to explicitly state the important conceptual relationships to guide instruction. This would focus the teaching and learning on the most significant conceptual understandings, and strengthen the bridge between the PYP, MYP and DP. I am sensitive to the curricular demands for university recognition but I also feel strongly that less is more when the student synergistically processes factual information through the conceptual level of thinking. Building the conceptual structures for deep understanding and the transfer of knowledge supports autonomous learners who maximize their learning by seeing patterns and connections between new knowledge and prior learning. New courses based on conceptual frameworks such as global politics will continue to emerge that will support teachers in helping students learn to do meta-analyses of complex systems, but traditional discipline-based courses also need a concept-based curriculum and instruction design for deep understanding.

The DP, like the PYP and MYP, supports international-mindedness and understanding of other cultures. Identifying key and related concepts and framing critical subject area content with a central idea and additional "supporting ideas" (which will be introduced later in this paper) can strengthen the transfer of knowledge across global contexts as new examples of previously learned concepts arise. Concept-based curriculums can support teachers in moving deliberately to idea-centred instruction. I know that DP teachers value deeper conceptual thinking and understanding. An idea-centred curriculum of important conceptual understandings supported by relevant content would help teachers meet these goals.
The PYP is transdisciplinary in nature. The identification of transdisciplinary themes (for example, who we are, how the world works) frame the concepts, skills, attitudes and actions linked to what is real and relevant in the world through the design of programmes of inquiry. The transdisciplinary themes ensure that curriculum and instruction move beyond factual coverage in discrete subject areas to an integrative synthesis of knowledge and conceptual understandings to better understand our world and our place within the world (IB 2010). The MYP is developing a set of “global contexts” to use in their unit designs that will facilitate transcendent thinking, similar to that driven by the transdisciplinary themes of the PYP (IB 2012).

In the PYP and MYP the key concepts draw thinking beyond the facts. This is significant for three main reasons.

1. The use of key concepts prevents an overreliance on memorization of facts as the end goal. In a concept-based model students must process the facts through their personal intellect—the conceptual mind. The key concept provides focus to the topic under study, acting as a conceptual draw for personal engagement and mental processing. The focus shifts from memorization—or a lower form of mental engagement—to deeper, personal inquiry as students consider connections between the facts and the key concept(s). Key concepts are macro-concepts that transcend disciplines such as change, interdependence, system and relationships. The PYP has identified eight macro-concepts as key concepts to use for the programme. The choice of just one or two key concepts for a unit planner can prevent the conceptual focus from becoming too diffuse. The MYP is also working to identify a set of key concepts. Key concepts integrate thinking at the conceptual level.

2. Key concepts facilitate the transfer of knowledge through time, across cultures and across situations. The IB position paper Thought, word and deed: The role of cognition, language and culture in teaching and learning in IB World Schools (Allan 2011) cites research undertaken in schools in the United States, Australia and Germany that found when students are exposed to abstract concepts apart from context, learning is difficult.

3. Intercultural understanding depends on the ability to see the commonalities and differences in terms of concepts and their expressions across global contexts, whether they be social, political, economic or geographical/environmental. When students develop understanding of key concepts and central ideas (statements of conceptual relationship) they become aware that these concepts and ideas can be applied across cultures.

I believe it is critical that all IB programmes attend to both the key concepts and the more discipline-specific related concepts to ensure that students develop breadth and depth of conceptual understanding. The transferability of key concepts such as system, change and order help students recognize the many permutations of each concept from body systems, to economic systems, to environmental systems. The related concepts, however, ensure that instruction builds depth of understanding by attending to, and adding to, the language of each subject area—the discipline-specific concepts and their important relationships from year to year. In the PYP, these related concepts can be taught in the context of the transdisciplinary units of instruction developed for the programmes of inquiry. Identifying the related concepts in these units ensures that disciplinary depth is included in the inquiry. When I use the term “related concepts” in my work with concept-based curriculum design, I am referring to the concepts related to specific disciplines within the unit, rather than specific concepts related to various key concepts. The reason for this is I want to identify the more specific concepts to build disciplinary depth.

At this time the PYP and the MYP ask teachers to use a key concept and a more discipline-specific related concept to state a central idea and concept statement respectively. To reinforce idea-centred teaching and conceptual thinking I recommend consideration of additional conceptual understandings crafted with the more discipline-specific related concepts to be added to each unit. I will call these understandings “supporting ideas” for the purposes of this discussion. In the MYP and PYP if a year-long course of instruction was framed under five or six units of instruction, I would think five to eight supporting ideas per unit—in addition to the central idea/concept statement—would be reasonable to guide the formative work.

Another reason I suggest that the PYP and MYP use the more specific related concepts to write additional supporting ideas for their unit planners is to continually build disciplinary schemata in the brain, so students are prepared for the conceptual rigour of the DP, as well as for lifelong learning and concept-based teaching and learning
work. It is through the conceptual structures of knowledge that the PYP, MYP and DP can be further aligned and articulated on the IB programme continuum.

Aligning the terminology through the different levels of the IB programmes and articulating central and supporting ideas using key and related concepts at all three levels of the IB would provide the structure for a continuous and coherent concept-based scheme of instruction. A concept-based curriculum is idea-centred. Central and supporting ideas facilitate a pedagogy requiring synergistic thinking which means guiding students, through inquiry, to realize the deeper conceptual understandings supported by factual content. In the PYP, a suggestion could be to write more specific disciplinary supporting ideas to serve as the lines of inquiry. Developing critical central and supporting ideas for the different subject areas (referred to as “subject groups” in the MYP) across all levels of the IB continuum would provide clear understanding of targets for the teachers, and would facilitate a truly idea-centred pedagogy. This would also allow the breadth of curriculum content to be compacted and focused.

How is a concept-based curriculum beneficial to student learning?

Concept-based curriculum and instruction is essential to the IB educational paradigm. The critical elements that require a concept-based model are intercultural understanding and international-mindedness, the ability to transfer knowledge and a rigorous intellectual model that is emotionally engaging and motivating.

There are many benefits to a concept-based model.

- **Thinking**—It requires thinking students who draw on critical, creative, reflective and conceptual thinking abilities.
  - Facilitates “synergistic thinking”—the cognitive interplay between the factual and conceptual levels of thinking.
  - Requires deeper intellectual processing as students relate the facts to key concepts and principles.
  - Develops conceptual structures in the brain (brain schemata) to relate new knowledge to prior knowledge, and to illuminate the patterns and connections of knowledge.
  - Facilitates the transfer of knowledge at the conceptual level.
  - Provides opportunities for personal meaning-making through processes of thinking, creating and reflecting.

- **Intercultural understanding**—It develops intercultural understanding and international-mindedness through conceptual transfer.
  - Facilitates the transfer of learning across global contexts as students engage with concepts and conceptual understandings as reflected across unique and varied cultures.
  - Encourages inquiry into global issues of concern that draw out the multiple perspectives and situations of different cultures and nations.

- **Motivation for learning**—It recognizes that intellectual and emotional engagement are essential to the motivation for learning.
  - Increases motivation for learning by inviting students to think about the facts through a relevant and personally engaging key concept. The unit topic and the key concept have an iterative relationship—each reinforces the other, for example, considering the facts about “Global conflicts in the 21st century” through the conceptual lens of perspectives, or considering facts about “Our land and people” through the lens of identity.
- Values and respects the thinking of the individual by “drawing understandings from” rather than “telling understandings to”.

- Encourages constructivist learning experiences that are relevant and important.

- Values collaborative thinking, discussions, and problem-solving with the belief that the social construction of meaning not only leads to a quality product, but is motivating to participants as well.

  - Fluency with language—it increases fluency with the languages of cultures and the disciplines.

  - Illuminates the conceptual structures of “meta-language” to facilitate multilingual learning and communication across cultures.

  - Builds increasing fluency with disciplinary language as students explain and support their conceptual understanding with relevant factual knowledge.

  - Reinforces a common conceptual vocabulary and set of critical conceptual understandings in the different disciplines which can help alleviate language barriers in global labour contexts when students enter the workforce.

These benefits are inherently supportive of the majority of the characteristics outlined in the IB learner profile (IB 2009a, p 5):

- inquirers

- knowledgeable

- thinkers

- communicators

- open-minded

- reflective.

What are the required pedagogical shifts?

Synergistic thinking

I believe that synergistic thinking (Erickson 2007; 2009b) is essential for intellectual development. Synergistic thinking is a cognitive interplay between the factual and conceptual levels of mental processing. Synergy can be defined as two interacting agents providing a greater effect than either agent acting alone. Thinking without this factual/conceptual interaction can be shallow. Without a deliberate curriculum design that mandates this intellectual interplay, we may confuse memorized knowledge with deeper understanding. Just knowing the definitions of concepts is not sufficient. Just knowing facts is not sufficient.

Transfer of knowledge and skills

Facts do not transfer. They are locked in time, place or situation. Knowledge transfers at the conceptual level as concepts, generalizations and principles are applied across global contexts and situations. The ability to use the conceptual level of thinking to relate new knowledge to prior knowledge, to see patterns and connections between different examples of the same concept or conceptual understanding, and to pattern and sort the expanding information base is a critical skill for the 21st century. The transfer of processes and skills across multiple disciplines and contexts to deepen understanding and enhance performance is another mandatory facet of IB programmes. The approaches to learning (ATL) in the MYP continue to be developed along with the transdisciplinary skills in the PYP. Work is underway to organize ATL skills across all levels of the IB related to five skill clusters: social, research, thinking, communication and self-management.

Concept-based teaching and learning
Social construction of meaning

Quality thinking is hard work. Concept-based models encourage collaborative group work to enhance thinking and problem-solving. Different minds working together scaffold each other and generate new ideas and solutions. The social construction of meaning and collaborative groups work is a significant aspect of all IB programmes. School days filled with teacher-dominated lectures to passive students, locked into parallel rows of desks, are hopefully a relic of past pedagogies.

Effective concept-based teachers in IB programmes understand the principles of synergistic thinking, the transfer of knowledge and socially constructed meaning-making. They have at some point made the following pedagogical shifts in their instruction if they began their teaching career in a traditional two-dimensional model. (Please also see the appendix for an example.)

<table>
<thead>
<tr>
<th>From two-dimensional instruction*</th>
<th>To three-dimensional instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>The goal is increased factual knowledge and skill development.</td>
<td>The goal is increased conceptual understanding supported by factual knowledge and skills, and the transfer of understanding across global contexts.</td>
</tr>
<tr>
<td>Teacher relies heavily on lecture to disseminate factual knowledge.</td>
<td>Teacher facilitates student inquiry into important interdisciplinary and disciplinary topics and issues using one or two key concepts as the conceptual draw.</td>
</tr>
<tr>
<td>Instruction and learning experiences focus on factual examples and definitions of concepts with assumed conceptual understanding.</td>
<td>Instruction and learning experiences utilize concepts along with factual content to ensure synergistic thinking. Teacher deliberately uses concepts to help students transcend the facts.</td>
</tr>
<tr>
<td>Teacher posts objectives for each lesson as required.</td>
<td>Teacher posts questions of different kinds (factual, conceptual, debatable) to engage interest and to facilitate synergistic thinking.</td>
</tr>
<tr>
<td>Students face the teacher in straight rows to ensure order and attention to the teacher’s instruction.</td>
<td>Students often work in groups to facilitate shared social inquiry, collaboration, synergistic thinking and problem-solving. Students may work independently, in pairs or groups, or across global contexts using the internet or other communication tools.</td>
</tr>
<tr>
<td>Teacher verbally summarizes the learning related to the objectives at the close of the lesson.</td>
<td>Teacher uses inductive teaching to draw the statement of conceptual understanding from students near the end of a lesson and posts the central or suggested supporting ideas for later connections to future topics in the curriculum. Students support their understanding with accurate facts as evidence of quality synergistic thinking.</td>
</tr>
<tr>
<td>Assessments measure factual knowledge and skills.</td>
<td>Assessments of conceptual understanding tie back to a central (or supporting idea) by incorporating specific language from the idea in the task expectations.</td>
</tr>
<tr>
<td>Teacher focuses on covering the required curriculum.</td>
<td>Teacher focuses on student thinking and understanding. He/she is cognizant of each student’s ability to think synergistically.</td>
</tr>
</tbody>
</table>

*The two-dimensional model is exaggerated in this paper to provide a clear contrast with the three-dimensional model.

Challenges and summary

There are challenges to the development and implementation of a concept-based, three-dimensional curriculum. But challenges indicate opportunities. The IB has the opportunity to meet the challenges to
refine the IB programmes so they are truly the most effective and engaging in the world. The greatest
chardles centre around curriculum development and programme articulation, teacher training and
assessment.

Fundamental to success in meeting each of these challenges is a solid understanding of concept-
base, three-dimensional curriculum and instruction.

1. Curriculum development. Quality teaching is supported by quality curriculums. The curriculum
must be concept-based to meet the goals of transfer of knowledge, deep conceptual
understanding, synergic thinking, intercultural understanding and personal intellectual
engagement. The IB continuum can be articulated and coordinated through the development of
common terminology and a common curricular framework of key concepts, related concepts, and
central and discipline-based supporting conceptual understandings. Unit planners can help ensure
that a concept-based teaching plan will be developed.

2. Teacher training. This challenge is critical to the success of a concept-based model. If teachers do
not understand the concept-based model and required shifts in pedagogy they will fall back on
traditional teaching methods and fail to effect transfer of knowledge and deep understanding. As
new schools and teachers are continually joining the IB family, creative ways of delivering the
teacher training need to be developed. Regional training centres around the world that certify IB
trainers after a rigorous training programme of one or two weeks with materials that teach the
concept-based model along with the other facets of the IB programmes could be part of the
solution. The critical point is that anyone training teachers on the concept-based model must
understand the model completely and be able to convey that understanding to others effectively.
(Administrators also need to be well trained on the meaning of a concept-based curriculum, what
to look for in instruction and how to support teachers in the implementation of the IB model.)

3. Assessment—The challenge here is to assess to the conceptual level of understanding, rather
than just to the factual level. The design of the classroom assessments must be part of the
teacher training programme. The IB external assessments also need to assess to the conceptual
level of understanding, while allowing schools to use local content to support the understandings.

Many nations lament the academic progress of their students year after year. Government officials
institute “solutions” and throw money at the problem but the “solutions” usually centre around increased
objectives to ensure topics are covered fully and more testing to make certain the curriculum is taught
as defined. However, the focus is on the content rather than on the development of the whole child—
social, emotional, intellectual and physical—a terrible mistake.

The IB is on the correct path to prepare citizens of the future for living, learning and working in global
environments, and for addressing the complex problems and issues that will undoubtedly arise. The IB
community is a family of passionate educators who will continue to develop common understandings in
curriculum and pedagogy to keep the ship on course. This journey could not be more important.

---

About the author

H Lynn Erickson is the author of three bestselling books: Stirring the Head, Heart and Soul: Redefining Curriculum and Instruction, 3rd edition January 2008; Concept-based Curriculum and Instruction: Teaching Beyond the Facts, 2002; and Concept-based Curriculum and Instruction for the Thinking Classroom, 2007, all published by Corwin Press. Erickson is a recognized presenter at national conferences in the areas of concept-based curriculum design, teaching for deep understanding and standards alignment.

Erickson was born and raised in Fairbanks, Alaska. She graduated from the University of Alaska in 1968 and taught at North Pole, Alaska before moving south. She taught various grade levels in California and moved to Missoula, Montana in 1976, earning master’s and doctorate degrees in curriculum and instruction and school administration. She has worked as a teacher, principal, curriculum director, adjunct professor and educational consultant over a 42-year career.
References


Appendix

Nevine Safire teaches the MYP. She has developed a concept-based unit around the central question “When does population growth become ‘overpopulation’?” To focus this unit she has chosen the key concepts of interactions and change from the MYP subject group guides. She is facilitating inquiry into the central idea “human/environmental interactions can become unbalanced, leading to changes with unintended consequences”.

Ms Safire has identified related concepts from science and the humanities to use in developing some supporting ideas to facilitate greater conceptual depth and understanding throughout the inquiry. Some of these concepts she found in the MYP subject group guides; others she extrapolated from the content she will be teaching:

<table>
<thead>
<tr>
<th>Migration</th>
<th>Environments</th>
<th>Overpopulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Scarcity</td>
<td>Climate change</td>
</tr>
</tbody>
</table>

Concept-based teaching and learning
<table>
<thead>
<tr>
<th>Population</th>
<th>Equilibrium</th>
<th>Conflict</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population density</td>
<td>Endangered</td>
<td>Adaptation</td>
</tr>
<tr>
<td>Natural resources</td>
<td>Trade</td>
<td>Sustainability</td>
</tr>
</tbody>
</table>

Ms Safire developed five or six supporting ideas to guide the formative unit work. One of her supporting ideas for this unit was “overpopulation can lead to conflict over scarce natural resources”. Within the lesson Ms Safire asks students to analyse both local and global contexts for population density. We will identify aspects of concept-based pedagogy as we follow some of Ms Safire’s thoughts during a concept-based lesson related to this supporting idea.

<table>
<thead>
<tr>
<th>Classroom actions</th>
<th>Ms Safire’s thoughts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ms Safire: We have been learning about population growth in our unit. In this lesson we are going to think about the effects of overpopulation. What does the term &quot;overpopulation&quot; mean to you?</td>
<td>I want to check students’ prior conceptions of the term and draw out a clear definition that we can build on.</td>
</tr>
<tr>
<td>Students view the video “World Population” by Population Connection.</td>
<td>This world map, adding the points of light representing increasing population growth over time, is an unforgettable visual of the growing problem.</td>
</tr>
<tr>
<td>Students work in table groups using a graphic organizer to predict the possible effects of rapidly increasing population density on the following: land, animal populations, plants, natural resources.</td>
<td>I wonder if students can foresee the potential problems that may be caused by rapidly increasing populations.</td>
</tr>
<tr>
<td>Analyse the graphic organizer and discuss the question “At what point might population growth be considered ‘overpopulation’?”</td>
<td>This provocative question requires students to articulate the destructive effects of an imbalance between the human and natural world.</td>
</tr>
<tr>
<td>View and discuss the video “Overpopulation and Its Effects on Our World”. <a href="http://www.youtube.com/watch?v=mWHo_eGaORU">http://www.youtube.com/watch?v=mWHo_eGaORU</a></td>
<td>This video supports, and further extends students’ understanding of the effects of overpopulation.</td>
</tr>
<tr>
<td>Teacher questions: Factual: Does our state (country, region) have areas of dense population? What issues in our region would suggest the danger of overpopulation? What areas of the world have very dense population centres? (Note: Ms Safire engages students with a research activity on overpopulation in different parts of the world and uses guiding questions to develop understanding.) Conceptual: Why do nations develop concentrated “population centres”? What would indicate that a population centre is “overpopulated”? How do increasing human/environmental interactions change the environment? How might these interactions lead to scarce natural resources? How do communities with scarce natural resources meet their survival needs? How can scarce natural resources lead to conflict between groups of people or nations?</td>
<td>I will ask students to bring their questions to the inquiry, but I also developed some factual, conceptual and provocative, debate questions for the lesson. I want to move students’ thinking beyond the local examples to the global perspective so that students can see the parallelism between local and global issues of concern. I want students to know which areas of the world have the greatest and least availability of natural resources; and to understand that scarcity of natural resources can lead to conflict.</td>
</tr>
</tbody>
</table>

- How can governments ensure that their citizens have the...
necessary natural resources for survival?

- Should nations with plentiful natural resources be required to share with nations who have scarce natural resources?

Be prepared to defend your position.

We have been inquiring into the idea of overpopulation. The following concepts have entered into our discussion. In your table groups see how many concept statements you can develop by showing relationships between the concepts below. Remember that these concept statements must transfer through time and across situations.

- Environment
- Interactions
- Change
- Overpopulation
- Natural resources
- Conflict
- Cooperation
- Scarcity

Be ready to cite specific factual examples to support your understandings.

I know that students will necessarily use synergistic thinking in this learning experience as they use the actual examples to support the construction of their conceptual understandings. I will be interested in seeing the relationships they find.

Assessment task:

You are the leader of a task force that has been commissioned by the city council to propose solutions to the problem of scarce water supplies caused by severe droughts in your area. A neighbouring area, which has also been affected by the drought, has blocked the downstream flow of water from a major source to ensure that all of their agricultural and industrial uses of water can be met. This blockage of water to your area has led to a heated conflict. You are to make a presentation to the city council that describes the water problem, and suggests creative and viable solutions—including a way to work with the neighbouring area to collaboratively develop a win-win solution. You can choose your mode of presentation.

This authentic performance task will help students learn to collaboratively problem solve using a local issue that has global implications. The task connects back to the supporting idea that scarce natural resources can lead to conflict and supports the IB aim of taking principled action to solve a community problem.

As we read Ms Safire’s thoughts during the lesson we can see her attention to concept-based pedagogy through the following:

- checking for prior understanding of concepts
- encouraging predictions
- using different kinds of questions (factual, conceptual, debatable) to move thinking beyond the facts
- engaging synergistic thinking
- assessing for conceptual understanding as well as facts and skills.
Section 2:

Agency
The Influence of Teaching
Beyond Standardized Test Scores: Engagement, Mindsets, and Agency
A Study of 16,000 Sixth through Ninth Grade Classrooms

Ronald F. Ferguson with
Sarah F. Phillips, Jacob F. S. Rowley, and Jocelyn W. Friedlander
October 2015

The Achievement Gap Initiative at Harvard University
Excellence with Equity
A Letter from the Raikes Foundation

The Raikes Foundation was pleased to commission this report from the Achievement Gap Initiative (AGI) at Harvard University. We welcome its release and anticipate a wide readership. The analysis by Ron Ferguson and his colleagues of over 300,000 Tripod surveys across 16,000 sixth and ninth grade classrooms offers an unparalleled opportunity to learn from what's working in teachers' classrooms. In their identification of agency dampers and agency boosters, the authors provide a nuanced picture of the balance of challenge and support behaviors teachers can employ to develop agency.

The report will inform our work at the Raikes Foundation where we are focused on empowering teachers to help their students develop learning mindsets and skills. This is a priority because research and experience in all kinds of classrooms have shown that empowering students with learning mindsets and skills can unlock their potential to grow in any subject, at any age. The Raikes Foundation is working with educators, parents and leading researchers to develop an evidence-based, teacher-tested toolkit of practical resources to help students cultivate important beliefs and abilities.

We are grateful for the Achievement Gap Initiative's leadership and so pleased that Ron Ferguson is among the 22 researchers participating in the Mindset Scholars Network, www.mindsetscholarsnetwork.org. The Network conducts original interdisciplinary research, builds capacity for high quality mindset scholarship, and disseminates the latest scientific knowledge through outreach to education stakeholders.

Thank you for your own commitment to this important work!

Best,

Zoe Stemm-Calderon
Director, Education Strategy
Raikes Foundation
www.raikesfoundation.org
CONTENTS

Foreword ii
Acknowledgments iv
Executive Summary 1
Chapter 1: Why Prioritize Agency? 13
Chapter 2: The Tripod 7Cs of Effective Teaching 18
Chapter 3: Agency and Agency-Related Factors 24
Chapter 4: Major Patterns in the Data 39
Chapter 5: Isolating the Influence of Teaching 50
Chapter 6: How Teaching Predicts Agency-Related Factors 56
Chapter 7: Ten Practical Implications for Teaching Agency 82
Appendix 94
References 103
About the Authors 110

This report is available for download at
http://www.agi.harvard.edu/publications.php
FOREWORD

Three years ago, the Consortium on Chicago School Research (CCSR) issued a report entitled, *Teaching Adolescents to Become Learners—the Role of Noncognitive factors in Shaping School Performance: A Critical Literature Review*. The CCSR report asserted a "need for more research on the role of school and classroom context in students' development and demonstration of noncognitive factors" (p. 75).

Officials at the Raikes Foundation sponsored the CCSR report. They commissioned this report as one response to the CCSR call for more research on how classroom contexts affect noncognitive factors.

In June 2015, as the present report was nearing completion, the CCSR issued a new report entitled, *Foundations for Young Adult Success: A Developmental Framework*. It resulted from an extensive literature review and a great deal of consultation with researchers and expert practitioners. In the new report, the authors describe features of an extensive research agenda focused on how institutions, including schools, foster development.

This report presents evidence concerning the influence of teachers on the development of what the first CCSR report called noncognitive factors and what their most recent report calls *foundational components and key factors*, prominent among which is *agency*.

About Tripod

Data from classroom-level Tripod student surveys have made this report possible. The Tripod Project for School Improvement (Tripod) emerged fifteen years ago from the first author's work with K-12 educators in Cleveland Ohio's racially diverse eastern suburbs. Since then, Tripod surveys have been continually refined and completed by students and teachers in thousands of schools across the United States. Tripod was the student survey used in the Measures of Effective Teaching (MET) project, sponsored by the Bill & Melinda Gates Foundation. Tripod data have also been used in U.S. Department of Education sponsored studies conducted by Mathematica, American Institutes for

---

Research, and Basis Policy Research. School systems commission the surveys and educators receive online reports to help with goal setting and school improvement. The work with school systems is based at Tripod Education Partners, Inc., a research and education firm located in Cambridge, Massachusetts. Rob Ramsdell and Ron Ferguson, the first author of this report, are co-founders and co-owners of Tripod. See www.tripoded.com for more information.

About The Achievement Gap Initiative

The Achievement Gap Initiative (AGI) at Harvard University is a university-wide effort based at the Harvard Graduate School of Education and the Malcom Wiener Center for Social Policy at the Harvard Kennedy School. Its mission is to bridge research and practice by framing important issues, producing and disseminating new research, and distilling implications for action by decision makers.

The AGI promotes excellence with equity as the defining goal. Not only should there be group proportional equality—where group-level characteristics such as race or socioeconomic status do not predict an individual’s achievement—there should also be excellence. Hence, the AGI is focused on scholarship, public education, and outreach activities to support governmental, civic, and private sector mechanisms aimed at raising achievement levels for all children while closing gaps between racial, ethnic, and income groups. Ron Ferguson, the first author of this report, is the AGI faculty co-chair and director. See www.agi.harvard.edu for more information.

This is an AGI report, commissioned by the Raikes Foundation.
ACKNOWLEDGMENTS

We express our sincere thanks to the Raikes Foundation and former program officer Craig Wacker for commissioning this report and for ongoing support as we have worked to complete it. In addition, we thank colleagues who invited us to present at seminars and the participants who provided helpful feedback. These include Gordon Berlin and researchers at MDRC in New York City, former Dean Deborah Vandell and faculty at the University of California at Irvine School of Education, and members of the National Mindset Collaborative, convened by Doug Yeager, Carol Dweck, and Greg Walton. Helpful feedback from public school teachers and administrators has come from participants in America Achieves, the Teacher Union Reform Network, and the American Association of School Administrators Collaborative. Educators in Bedford, New York, Ossining, New York, Tarrytown, New York, and Omaha, Nebraska also provided helpful feedback. Julie Hackett, superintendent of the Taunton, Massachusetts Public School District, was especially helpful. In addition, Ann Ballantine provided insightful advice and support on content, editing, and presentation.

Finally, we express our gratitude to the thousands of teachers and students whose perspectives and experiences are reflected in the data upon which this report relies.
EXECUTIVE SUMMARY

Today in the United States, producing higher scores on standardized tests of academic skills is the dominant goal of teacher professional development, the primary gauge of teacher productivity, and the almost single-minded focus of educator accountability. Certainly, reading, computing, and reasoning well are critically important to success as parents, citizens, and economic actors. Therefore, testing these skills in elementary and secondary schools to make sure that students learn them is warranted. At the same time, there is growing agreement that scores on standardized tests of academic skills are incomplete measures of the important things that students learn from their teachers. A major challenge facing educators, policy makers, and advocates is to achieve a better balance across the educational goals that we prioritize.

We present new evidence in this report that untested learning outcomes are measureable and that specific components of teaching influence them in nuanced and interesting ways. As targets for improved teaching and learning, these outcomes can supplement academic skills and knowledge as intentionally cultivated developmental foundations for school and life success.

The report relies upon data from over 300,000 Tripod student surveys administered in more than 16,000 sixth to ninth grade classrooms, 490 schools, 26 districts, 14 states, and in all major regions of the nation during the 2013-14 school year.

The Central Question

The report concerns the influence of teaching on emotions, motivations, mindsets, and behaviors that we associate with agency. We ask, "How do distinct components of teaching influence the development and expression of agency-related factors in sixth to ninth grade classrooms?" Agency is the capacity and propensity to take purposeful initiative—the opposite of helplessness. Young people with high levels of agency do not respond passively to their circumstances; they tend to seek meaning and act with purpose to achieve the conditions they desire in their own and others' lives. The development of agency may be as important an outcome of schooling as the skills we measure with standardized testing (Exhibit 1).
We examine how seven distinct components of teaching influence a number of agency-related factors, other things equal. Conventional wisdom would predict that the aspects of teaching we feature are positively associated with all types of learning outcomes. The empirical findings are mostly consistent with this prediction. However, there are caveats. For example, some of the ways that teachers strive to be helpful and caring can be agency dampers, serving to reduce rather than bolster agency, other things equal. In addition, some teaching components that tend to be agency boosters have the potential to be stressful in ways that can threaten an otherwise ambitious teacher's resolve. The report raises issues for teacher training and professional support as well as for individual educators who are striving to be reflective practitioners.

**Key Concepts**

The data come from student responses to Tripod surveys. When a student responds to a Tripod survey, his or her responses are
focused on one designated classroom. The items that measure teaching are conceptualized using the Tripod 7Cs™ framework as illustrated in Exhibit 2. In this framework, seven research-based components of teaching are represented using multi-item indices. The seven are:

1. Care—Teachers who care are emotionally supportive and interested in students.
2. Confer—Teachers who confer talk with students as well as welcome and respect student perspectives.
3. Captivate—Teachers who captivate make learning interesting and relevant.
4. Clarify—Teachers who clarify explain things clearly, provide informative feedback, and clear up confusion in order to make lessons understandable.
5. Consolidate—Teachers who consolidate summarize and integrate learning.
6. Challenge—Teachers who challenge students press them to think rigorously and to persist when experiencing difficulty.

7. Classroom Management—Effective classroom management entails developing a respectful, cooperative classroom climate with on-task behavior.

In addition, clarify has sub-components pertaining to lucid explanations, informative feedback, and clearing up confusion. Challenge has sub-components associated with requiring rigor and requiring persistence. Each 7Cs component and sub-component is represented by an index composed of multiple survey items.

The report examines how 7Cs components of teaching quality predict individual-level behaviors that express agency and a number of emotions, motivations, and mindsets that awaken and support the growth of agency. Most are associated with classroom engagement. Indices for the behavioral expression of agency, illustrated in Exhibit 3, include the following:

- **Punctuality**—The student tries hard to arrive to class on time.
- **Good Conduct**—The student is cooperative, respectful, and on task.
- **Effort**—The student pushes themself to do their best quality work.
- **Help Seeking**—The student is not shy about asking for help when needed.
- **Conscientiousness**—The student is developing a commitment to produce quality work.

- **Disengagement Behaviors** (that are the opposite of agency):
  - Faking Effort—The student pretends to be trying hard when they actually are not.
  - Generally Not Trying—The student is generally disengaged, exerting little effort.
  - Giving Up if Work is Hard—The student fails to persist in the face of difficulty.
  - Avoiding Help—The student does not ask for help even when they know they need it.
Exhibit 3
Student Expressions of Agency

- Punctuality
- Good Conduct
- Effort
- Help Seeking
- Conscientiousness

- Taking Effort
- Generally Not Trying
- Giving Up if Work is Hard
- Help Avoidance

All of the above pertain to the specific classrooms in which students complete surveys. It seems likely that each has implications as well for how students express agency in other settings. For conscientiousness, in particular, we address not only whether the student is learning to be more conscientious in the surveyed classroom, but also whether they consider themselves generally to be a conscientious person. We are interested in the degree to which teaching affects both learning of conscientiousness and changes in identity self-perceptions of conscientiousness.

Indices representing emotions, motivations, and mindsets that may awaken and support the growth of agency (Exhibit 4) include:

- Happiness—The student regards the classroom as a happy place to be.
- Anger—The student experiences feelings of anger in class (which may boost or dampen agency).
Exhibit 4
Emotions, Motivations, and Mindsets Associated with Agency

- **Mastery Orientation**—The student is committed to mastering lessons in the class.
- **Sense of Efficacy**—The student believes they can be successful in the class.
- **Satisfaction**—The student is satisfied with what they have achieved in the class.
- **Growth Mindset**—The student is learning to believe they can get smarter.
- **Future Orientation**—The student is becoming more focused on future aspirations (e.g., college).

The report grounds these concepts in the research literature and embeds them in organizing frameworks.

**Methodology for Teaching Quality**

An important methodological feature is in the way we use student responses to represent instructional quality. Each Tripod 7Cs teaching
quality component is measured using a classroom-level average of student responses. However, we make an important adjustment. Specifically, each classroom average for a teaching quality measure excludes the 7Cs teaching quality responses of the individual student whose emotion, motivation, mindset, or behavior is being predicted. This is important because it makes the teaching quality measure for each student more objective.

Findings and Interpretations

Empirical results show that each agency-related factor is predicted by a distinct pattern of 7Cs components. Furthermore, each 7Cs component and subcomponent plays an interesting role in the total story that emerges. It is a story in which students' emotions, motivations, mindsets, and agency-expressive behaviors are predicted by what teachers *ask* of them as well as by what teachers *give*. In the education research literature, what we call *asking* is associated with academic and behavioral *press*, while what we call *giving* is associated with social and academic *support*.

Agency Dampers and Agency Boosters

Our analysis identifies *agency dampers* as teaching practices that tend to reduce or dampen agency-related factors, and *agency boosters* as teaching practices that tend to increase them. *Agency dampers* appear mainly in relationship to *care, confer*, and the subcomponent of *clarify* called *clear up confusion*.

Other things equal, the effects of *care, confer*, and *clear up confusion* are statistically significant in the *undesirable* direction for some agency-related variables. The negative findings seem to indicate imbalances between what teachers *give* students (i.e., supports), on the one hand, versus what they *ask* of students (i.e., press), on the other hand. Specifically:

- *Care* may sometimes entail coddling (e.g., in an effort to be emotionally supportive, some teachers may be especially accommodating and this may depress student conduct as well as academic persistence);
*Conferring* may sometimes lack clear purposes (e.g., *confer* may operate primarily as a way of *caring*, *clarifying*, and *challenging*. When all of the other 7Cs components are being held constant, *confer* cannot be serving as a way of carrying them out and may therefore lack a clear purpose; conversations without clear purposes may undermine student effort and reduce time on task;  

*Clearing up confusion* may sometimes occur too automatically (e.g., too much teacher problem solving or clearing up confusion can deny students adequate incentive and opportunity to diagnose and correct their own misunderstandings, ultimately diminishing effort and conscientiousness).

Each of these *agency dampers* involves ways that support for students may inadvertently lower levels of press for performance.

An important category of *agency boosters* involves asking students to think more rigorously by striving to understand concepts, not simply memorize facts, or by being able to explain their reasoning. We find that when teachers challenge students to think more rigorously, they show:  

- Greater mastery orientation (i.e., personal commitment to learning);  
- Increased effort;  
- Increased growth mindset (i.e., belief that effort grows ability);  
- More conscientiousness; and  
- Higher future aspirations (e.g., interest in going to college).

In these ways, requiring rigor is an *agency booster*. However, we find that when teachers insist on rigorous thinking they risk at least slightly diminishing students’ happiness in class, feelings of efficacy, and satisfaction with what they have achieved.

These slightly dampened emotions in the short-term seem small prices to pay for the motivational, mindset, and behavioral payoffs we predict to result from requiring rigorous thinking. Combinations of teaching practices—for example, appropriately differentiated assignments, lucid explanations of new material, and curricular
supports to accompany demands for rigor—seem quite relevant in this context.

**Annual Gains Versus Aspirations**

An interesting and important contrast concerns the combination of Tripod 7Cs components that most strongly predicts annual learning gains versus an increased interest in going to college.

Prior research by the Bill & Melinda Gates Foundation Measures of Effective Teaching (MET) project has shown that annual achievement gains on standardized tests are predicted most strongly by aspects of academic press, specifically the Tripod 7Cs components for challenge and classroom management. The component for challenge, with its subcomponents for require rigor and require persistence, asks students to think hard and work hard. The component for classroom management asks students to behave themselves and stay on task. Thinking rigorously, sustaining effort, and staying on task may be sufficient to produce substantial learning gains, even if the teacher-student relationship is not what it should be and the interest and relevance of the material is relatively low.

However, being pressed to learn and experiencing learning growth may not produce a love of learning or a desire to continue learning over a lifetime. We find that the strongest predictors of increased college aspirations (future orientation) are the 7Cs components for care and captivate. When teachers connect with students in personal ways and make lessons fascinating, aspirations for future learning tend to rise.

The point is not that there is a trade-off between annual learning gains and higher aspirations. Instead, the point is that the most important agency boosters for each are different. A balanced approach to instructional improvement will prioritize care and captivate to bolster aspirations, and challenge and classroom management to strengthen the skills that standardized tests measure. Certainly, without the skills that tests measure, college aspirations might be futile. But in turn, without college aspirations, the payoffs to those skills may be limited.

---

3 Visit www.metproject.org to find MET project reports.
Ten Practical Implications for Teaching to Develop Agency

Based upon the findings above, we have distilled the following ten implications for teaching list in Exhibit 5 and described below:

1. **Care:** Be attentive and sensitive but avoid coddling students in ways that hold them to lower standards for effort and performance and may thereby undermine agency. At the same time, express interest in students’ lives, activities, and aspirations so they will feel known and inspired to follow your example.

2. **Confer:** Encourage and respect students’ perspectives and honor student voice but do so while remaining focused on instructional goals. Avoid extended discussions that have no apparent purpose and thereby fail to model self-discipline and effective agency.

3. **Captivate:** Strive to make lessons stimulating and relevant to the development of agency. If some students seem unresponsive, do not assume too quickly that they are disinterested. Some students—and especially those who struggle—purposefully hide their interest and their effort.

4. **Clarify with lucid explanations:** Strive to develop clearer explanations—especially for the material that students find most difficult. Also, related to both clarify and captivate and consistent with the themes in this report, develop lucid explanations of how the skills and knowledge you teach are useful in the exercise of effective agency in real life.

5. **Clarify by clearing up confusion:** Take regular steps to detect and respond to confusion in class but do so in ways that share responsibility with students. Strike a balance between keeping hope alive for struggling students, on the one hand, versus pressing them to take responsibility for their own learning, on the other hand.
Exhibit 5
Implications for Teaching Across the 7Cs to Develop Agency

| 1. CARE | Be attentive and sensitive, but don’t coddle. |
| 2. CONFER | Encourage and respect students’ perspectives, but don’t waste class time with idle chatter. |
| 3. CAPTIVATE | Make lessons stimulating and relevant while knowing that some students may hide their interest. |
| 4. CLARIFY: | Take regular steps to detect and respond to confusion, but don’t just tell students the answers. |
| - Clear up Confusion | |
| - Lucid Explanations | |
| - Instructive Feedback | |
| 5. CONSOLIDATE | Regularly summarize lessons to help consolidate learning. |
| 6. CHALLENGE: | Anticipate some resistance but persist. |
| - Require Rigor | |
| - Require Persistence | |
| 7. CLASSROOM MANAGEMENT | Achieve respectful, orderly, and on task student behavior by using clarity, captivation, and challenge instead of coercion. |
6. **Clarify with instructive feedback:** Give instructive feedback in ways that provide scaffolding for students to solve their own problems. Through instructive feedback, you provide the type of support that enables students to develop and express agency by correcting their own work, solving their own problems, and building their own understandings.

7. **Consolidate:** Regularly summarize lessons to remind students what they have learned and help them encode understanding in memory, even when they seem reticent or disinterested. Consolidation helps to solidify student learning.

8. **Challenge by requiring rigor:** Press students to think deeply instead of superficially about their lessons. Set and enforce learning goals that require students to use reasoning and exercise agency in solving problems. Expect some pushback from students who might prefer a less stressful approach. Try increasing *captivation* and *care* in combination with rigor in order to help mitigate the tension and make the experience more enjoyable.

9. **Challenge by requiring persistence:** Consistently require students to keep trying and searching for ways to succeed even when work is difficult. Emphasize the importance of giving their best effort to produce their best work as a matter of routine. Be confident that few things could be more important for helping your students to develop agency.

10. **Classroom Management:** Strive to achieve respectful, orderly, on-task student behavior in your class by teaching in ways that clarify, captivate, and challenge instead of merely controlling students through intimidation or coercion.
From Sage on the Stage to Guide on the Side

Alison King

In most college classrooms, the professor lectures and the students listen and take notes. The professor is the central figure, the “sage on the stage,” the one who has the knowledge and transmits that knowledge to the students, who simply memorize the information and later reproduce it on an exam—often without even thinking about it. This model of the teaching-learning process, called the transmittal model, assumes that the student’s brain is like an empty container into which the professor pours knowledge. In this view of teaching and learning, students are passive learners rather than active ones. Such a view is outdated and will not be effective for the twenty-first century, when individuals will be expected to think for themselves, pose and solve complex problems, and generally produce knowledge rather than reproduce it.

According to the current constructivist theory of learning, knowledge does not come packaged in books, or journals, or computer disks (or professors’ and students’ heads) to be transmitted intact from one to another. Those vessels contain information, not knowledge. Rather, knowledge is a state of understanding and can only exist in the mind of the individual knower; as such, knowledge must be constructed—or re-constructed—by each individual knower through the process of trying to make sense of new information in terms of what that individual already knows. In this constructivist view of learning, students use their own existing knowledge and prior experience to help them understand the new material; in particular, they generate relationships between and among the new ideas and between the new material and information already in memory (see also Brown, Bransford, Ferrara, and Campione 1983; Wittrock 1990).

When students are engaged in actively processing information by reconstructing that information in such new and personally meaningful ways, they are far more likely to remember it and apply it in new situations. This approach to learning is consistent with information-processing theories (e.g., Mayer 1984), which argue that reformulating given information or generating new information based on what is provided helps one build extensive cognitive structures that connect the new ideas and link them to what is already known. According to this view, creating such elaborated memory structures aids understanding of the new material and makes it easier to remember.

In contrast to the transmittal model illustrated by the classroom lecture—note-taking scenario, the constructivist model places students at the center of the process—actively participating in thinking and discussing ideas while making meaning for themselves. And the professor, instead of being the “sage on the stage,” functions as a “guide on the side,” facilitating learning in less directive ways. The professor is still responsible for presenting the course material, but he or she presents that material in ways that make the students do something with the information—interact with it—manipulate the ideas and relate them to what they already know. Essentially, the professor’s role is to facilitate students’ interaction with the material and with each other in their knowledge-producing endeavor. In the constructivist model the student is like a carpenter (or sculptor) who uses new information and prior knowledge and experience, along with previously learned cognitive tools (such as learning strategies, algorithms, and critical thinking skills) to build new knowledge structures and rearrange existing knowledge.

But how do we get from transmission of information to construction of meaning? Such a change can entail a considerable shift in roles for the professor, who must move away from being the one who has all the answers and does most of the talking toward being a facilitator who orchestrates the context, provides resources, and poses questions to stimulate students to think up their own answers.

Change is never easy; usually, however, changes are easier to bring about by modifying existing practices than by starting afresh. So, we will begin by looking at some practical active-learning activities that can be incorporated into a typical lecture; then we will move on to
the more formal approach of cooperative learning, an alternative to the lecture. This sequence will show how the professor can make a gradual transition from the role of sage to that of guide.

Promoting Active Learning

Active learning simply means getting involved with the information presented—really thinking about it (analyzing, synthesizing, evaluating) rather than just passively receiving it and memorizing it. Active learning usually results in the generation of something new, such as a cause-effect relationship between two ideas, an inference, or an elaboration, and it always leads to deeper understanding. However, students do not spontaneously engage in active learning; they must be prompted to do so. Therefore we need to provide opportunities for active learning to take place. A general rule of thumb might be as follows: for each major concept or principle that we present, or that our students read about in their text, we structure some activity that requires students to generate meaning about that concept or principle. For this approach to be effective, students must use their own words and experiences—not regurgitate the text or lecture.

An active-learning activity that can easily be incorporated into a lecture is “think-pair-share.” Let’s look at an example of how this works. Dr. Jones is lecturing to his Anthropology 101 class on the role of language in culture. After several minutes, he poses the question: “What do you think would happen if we had no spoken language? Think about that for a minute.” After a minute he continues, “Now pair up with the person beside you and share your ideas.”

Each of the examples of active learning listed in Table 1 can be similarly incorporated into a lecture and can be accomplished during a one- to four-minute pause in the presentation. When I use these tactics during a lecture, I simply stop talking for a few minutes and have students engage in one of the activities. Then I have selected students share the product of their activity before continuing with my presentation. Students either work alone or collaborate in pairs.

Guided Reciprocal Peer Questioning

Now let’s look at small group learning processes. These are methods that promote problem exploration and task completion by students working in small groups while also having individual students engage in interactive learning with their peers. In these small groups the student is simultaneously an active constructor of knowledge and a collaborator with peers in a shared construction of meaning; the role of the professor is to guide and facilitate this process.

Again, let’s begin with an instructional approach that is interactive, can be used in conjunction with the familiar lecture presentation format, and that gets students actively involved in constructing meaning. This is an approach that I have developed and that I call “Guided Reciprocal Peer Questioning” (King 1989, 1990, 1992).

Guided Reciprocal Peer Questioning is an interactive learning procedure that can be used by students in any area of the curriculum to help them actively process material presented in lectures or other classroom presentations. Students work in groups of three or four. They

<table>
<thead>
<tr>
<th>Table 1.—Learning Activities to Incorporate into a Lecture</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student activity</strong></td>
</tr>
<tr>
<td>Think-pair-share</td>
</tr>
<tr>
<td>Generating examples</td>
</tr>
<tr>
<td>Developing scenarios</td>
</tr>
<tr>
<td>Concept mapping</td>
</tr>
<tr>
<td>Flowcharting</td>
</tr>
<tr>
<td>Predicting</td>
</tr>
<tr>
<td>Developing rebuttals</td>
</tr>
<tr>
<td>Constructing tables/graphs</td>
</tr>
<tr>
<td>Analogical thinking</td>
</tr>
<tr>
<td>Problem posing</td>
</tr>
<tr>
<td>Developing critiques</td>
</tr>
<tr>
<td>Pair summarizing/checking</td>
</tr>
</tbody>
</table>
are provided with a set of generic questions to use as a guide for generating their own specific questions on the lecture content (see Figure 1). With the help of the question stems, each student individually writes two or three thought-provoking questions based on the lecture. Following this self-questioning step of the procedure, students engage in peer-questioning. They pose their questions to their group and then take turns answering each other’s questions in a group discussion format.

These generic questions are designed to induce higher-order thinking on the part of students. For example, simply formulating specific questions (based on the generic questions) forces students to identify the relevant ideas from the lecture, elaborate on them, and think about how those ideas relate to each other and to their own prior knowledge. Responding to others’ questions further extends such active learning.

A Classroom Example

Professor Tax-Fax lectures to his introductory principles of accounting class for twenty minutes on the topic of intangible assets. Then he pauses, signals to the class, and the students turn to their neighbors and form groups of three. Dr. Tax-Fax turns on the overhead projector to display a list of questions. All of the questions are content-free, such as, “What does . . . mean?” and “What conclusions can you draw about . . . ?” (see Figure 1). Dr. Tax-Fax expects his students to use these generic questions to guide them in formulating specific questions on the topic of intangible assets. Within a few minutes, each student in the class (working individually) has selected appropriate generic questions and has written down one or two specific questions. At another signal from Dr. Tax-Fax, the small groups begin their questioning and responding (see dialogue in Figure 2).

The students continue asking and answering each other’s questions for several more minutes until Dr. Tax-Fax indicates that their discussion time is over. He then brings the class together to share and discuss inferences, examples, and explanations generated by the different small groups and to clarify any misunderstandings that the students might have had regarding the topic of intangible assets.

An Analysis of the Example

Dr. Tax-Fax’s students were engaged in several forms of active learning during their guided peer-questioning and responding activity. First of all, they had to think critically about the lecture content just to be able to formulate their specific thought-provoking questions. To generate those questions, not only did the students have to identify the main ideas of the lecture, they also had to consider how those ideas relate to one another and to the students’ own existing knowledge. Second, in order to answer those questions, the students had to be able to analyze and evaluate ideas presented, apply the information in new situations, generate inferences from the lecture material, and identify relationships among the concepts covered.

More specifically, in order to respond to a student’s question, the other students in a group had to construct explanations and communicate them. Explaining something to someone else often requires the explainer to think about and present the material in new ways, such as relating it to the questioner’s prior knowledge or experience, translating it into familiar terms, or generating new examples. Such cognitive activities force the explainer to clarify concepts, elaborate on them, reorganize thinking, or in some manner reconceptualize the material.

Webb’s (1989) extensive research on interaction and learning in peer groups indicates that giving such explanations improves understanding for the individual doing the explaining. For example, in the sequence of dialogue shown in Figure 2, Maggie asked her group for the definition of the term intangible assets, and Fred, in the first part of his response, simply parroted Dr. Tax-Fax’s definition. However, Fred showed that he actually had made some meaning for the term when he later explained why the cookie recipe would be considered an intangible asset, thus suggesting that he had reorganized his thinking by incorporating that concept into his existing knowledge. Similarly, Sam’s inclusion of Mrs. Field’s cookie recipes as a new example of intangible assets was an indication of reconceptualization on his part. Furthermore, Sam’s explanation of how Mrs. Field’s recipe (an intangible asset) could lose value showed concept clarification—he really understood some of the nuances of the concept. Essentially, Sam was using his prior knowledge to make sense of the newly presented concept of intangible assets.

When students think about class material in these ways, they actively process the ideas and construct for themselves extensive cognitive networks that connect the new ideas and link them to what they already know (e.g., Mayer 1984). Developing such cognitive networks facilitates understanding and makes it easier to remember the new material. In the discussion precipitated by Sam’s “What if . . . ?” question, the three students together explored the relationship between maintaining the value of unique intangible assets and dissemination of the information that makes those assets unique. In speculating on the effects that the newspaper advertisement might have for Mrs. Field’s business, they undoubtedly forged new links among the ideas presented in the lecture and between those ideas and their own prior knowledge. For example, Fred integrated the new information about intangible assets with marketing/adverti-
He knew that the generic questions would control the quality of the specific questions students asked and that those questions in turn would influence the quality of student thinking and knowledge building during discussion. All he had to do was structure the situation to allow that to happen. Therefore, after providing the guiding questions, he arranged the class in groups of three with the requirement that they discuss the topic of intangible assets by taking turns asking and answering each others’ specific questions on the topic. Because this reciprocal questioning-answer procedure requires each individual to contribute questions and answers, all members of each group were obligated to participate, but no one individual dominated the discussion.

Professor Tax-Fax had his students work in small groups because he knew that learning through peer-group interaction results in cognitive benefits for each student far beyond those that an individual would experience working alone. He was aware that in small group learning contexts such as Guided Reciprocal Peer Questioning students are confronted with each others’ conflicting viewpoints on issues as well as differences in each other’s prior knowledge and current understanding of the topic, and, in attempting to understand each other’s views and come to agreement, individual students have to modify their own thinking. Each member of such a group makes important and necessary contributions to the construction of a shared understanding of the topic; however, each individual’s understanding and expression of it is idiosyncratic. Such learning exemplifies the social construction of knowledge—a model of the learning process that is constructivist in nature but that also emphasizes collaboration.

When Professor Tax-Fax ended the activity by calling on each group to share its ideas, he was extending the social construction of knowledge to a whole-class context. In doing so, he made sure that new inferences and understandings were disseminated across groups and that if groups arrived at conflicting meanings, those differing perspectives would be revealed and

-ing concepts (such as consumer characteristics and the deliberate comparison with specific competitors) that he had learned about in a different course.

Dr. Tax-Fax’s role in this activity was purely facilitative. As a guide on the side, he promoted knowledge building in unobtrusive but powerful ways. To begin with, he provided the students with question starters written at the higher levels of Bloom’s (1956) taxonomy of thinking. He was well aware of the importance of carefully selecting the generic question starters to be used.

Vol. 41/No. 1
could be reconciled through further whole-class discussion.

**Effects of Guided Reciprocal Peer Questioning**

In using Guided Reciprocal Peer Questioning with a number of college classes, I have found that teaching students to ask their own thought-provoking questions stimulates their critical thinking and promotes high-level discussion. Because of the reciprocal nature of this procedure, all students actively participate in the discussions. Even those students who are reluctant to participate in class for fear of asking the teacher “stupid” questions are less hesitant about posing such questions to their peers in a small group.

I have also found that students who are taught to ask and answer thoughtful questions perform better on subsequent tests of lecture comprehension than do students who use other comprehension strategies such as unguided group discussion or independent review (King 1989, 1990). Such an improvement in learning suggests that the students who engage in this questioning-answering process actually recontextualize the material. In fact, tape recordings of the students’ discussions have shown that students using Guided Reciprocal Peer Questioning give more explanations and highly elaborated responses to each other than do students who use either discussion or unguided reciprocal peer questioning strategies, thus indicating some degree of conceptual restructuring on the part of those particular students.

**Cooperative Learning**

The same sociocognitive benefits that derive from Guided Reciprocal Peer Questioning can be obtained from other instructional approaches that call for cooperative learning, such as Jigsaw, Constructive Controversy, and Co-op Co-op.

**Jigsaw**

Jigsaw is a cooperative learning procedure commonly used in classroom settings (Aronson et al. 1978). In jigsaw, as with all cooperative learning approaches, the professor says very little but unobtrusively arranges the context and facilitates the process. Jigsaw activities are designed so that each student in a group receives only part of the learning materials and must learn that part and then teach it to the others in the group. Thus, each student’s part is like one piece of a jigsaw puzzle; to understand the whole picture, students must have access to all parts of the learning materials. Because students must combine their pieces to complete the puzzle, each team member’s contribution is highly valued.

In implementing a jigsaw activity, the professor divides the material to be learned into several parts—usually no more than five or six. Each part must be a unique source of information that is comprehensible on its own without reference to any of the other parts. Students are assigned to “home teams” with as many members as there are parts to the learning materials, and each team member receives one part of the material. Students reassemble into “expert groups” by joining all of the other students who received that particular part. In their expert groups, students read and discuss their part of the material together to learn it thoroughly. Then they return to their home groups and teach the part they learned to the other members of their team.

In this way, each team member is an expert in one part of the material to be learned, and each team member learns material from the other experts on the team; thus, jigsaw emphasizes interdependence. Finally, each student is tested independently to assess individual understanding of the complete set of materials. In this way, jigsaw emphasizes individual accountability.

In a psychology course on theories of personality, for example, jigsaw might be used to present material such as differing theories of personality, alternative approaches to assessing personality, or specific examples of personality disorders. In some cases, a jigsaw teaching-learning approach might be used to provide an overview of a particular topic; in other cases, this approach might be used following the introduction of an area of study. (For developing specific uses for the jigsaw in various disciplines, consult Aronson et al. 1978).

**Constructive Controversy**

Another cooperative learning strategy for use with large classes is constructive controversy. In this procedure, students work in teams of four; pairs of students within teams are assigned to opposing sides of a controversial issue. Each pair researches its side of the issue and then the pairs discuss the issue as a team. The purpose of this discussion is to become more informed about the issue and engage in collaborative construction of meaning—not to win a debate about the issue. After some discussion, pairs switch sides and argue for the opposite side of the issue. Finally, each student takes a test on the material individually to determine that student’s understanding of the issue. Constructive controversy might be used in computer courses, for example, to encourage students to explore the ethical issues inherent in the use of computers, software, and telecommunications.

**Co-op Co-op**

Co-op co-op is a student-centered cooperative approach to learning and can be used for the study of any unit of course material or for any number of research or problem-solving projects. Students work together in small teams to investigate a topic and produce a group product that they then share with the whole class. Thus the name “co-op co-op”: students cooperate within their teams to produce something of benefit to the class; they are cooperating in order to cooperate. There are nine steps in implementing co-op co-op. Again, at each step the professor guides the process from the side, facilitating students’ interaction with learning materials and with each other.

**Step 1. Student-centered class discussion.** At the beginning of an instructional unit, the professor encourages the students to discuss their interests in the subject to be covered. This discussion should lead to an understanding among the professor and all the students about what the students want to learn and experience during the unit. The importance of this initial discussion cannot be
Step 2. Selection of student learning teams. Students self-select into four-to-five-member teams.

Step 3. Team topic selection. In their teams, students discuss their interests in the topics and then select a topic for their team. Each team should select a topic with which its members identify.

Step 4. Minitopic selection. Just as the class as a whole divides the unit into sections to create a division of labor among the teams in the class, each team divides its topic into minitopics to create a division of labor within the team. Each team member selects a minitopic.

Step 5. Minitopic preparation. After selecting their minitopics, students work independently to prepare their minitopics. Depending on the nature of the main topic being covered, the preparation of minitopics may involve library research, data gathering through surveys or experimentation, creation of an individual project, or some expressive activity such as writing a script or creating a video.

Step 6. Minitopic presentations. When students complete their minitopics, they present them to their teammates. These presentations should be formal. Presentations and follow-up discussion should allow all team members to gain the knowledge and experience acquired by each. Following the presentations, team members discuss the team topic like a panel of experts, critiquing the presentations and noting points of convergence and divergence. The professor should provide time for feedback and additional time for teams or team members to rework aspects of their reports in light of that feedback.

Step 7. Preparation of team presentations. Students integrate the minitopics for the team presentation. (Panel presentations in which each member reports on his or her minitopic are discouraged as they may represent a failure to reach high-level cooperative synthesis of the material.) The form of the presentation should be dictated by the content of the material. Non-lecture formats such as debates, displays, team-led class discussions, videotapes, simulations, role-playing episodes, or demonstrations are encouraged (as are the use of overheads and audiovisual materials).

Step 8. Team presentations. During its presentation, a team takes over the classroom and is responsible for how the class time, space, and resources are used.

Step 9. Evaluation. Being student-centered, co-op co-op calls for the class to have considerable say in how learning is evaluated as well as the criteria to be used in that evaluation. Therefore most evaluation will be self-evaluation or peer evaluation; however, the class may decide to include instructor evaluation also. Evaluation can take place on three levels: (1) team presentations (generally evaluated by the class or by the team itself), (2) individual contributions to the team effort (often evaluated by the team or the individual student), or (3) a write-up of the minitopic (often evaluated by the team).

Findings

Studies of group-based learning, conducted over the past twenty years, have shown that such approaches to learning can be effective in increasing student achievement (Slavin 1990). However, improved achievement seems to result primarily when the cooperative approach uses some sort of group goal and stresses individual accountability. Apparently, when students are individually accountable for their learning (e.g., when each member of the group must take a test) and a group goal is established (e.g., when every individual in the group must understand the material to pass the test), group members have incentive to help each other learn the material.

This sets up a condition of interdependence. Under these circumstances, group members tend to provide each other with elaborated explanations of concepts and processes so that everyone will understand the material and will excel on the test. As discussed earlier, explaining something to others improves one's own understanding (see Webb 1989). Cooperative and collaborative learning also have positive effects on self-concept, race relations, acceptance of handicapped students, and enjoyment of school (Slavin 1990).

Engaging our students in such active learning experiences helps them to think for themselves—to move away from the reproduction of knowledge toward the production of knowledge—and helps them become critical thinkers and creative problem solvers so that they can deal effectively with the challenges of the twenty-first century.

NOTE


REFERENCES


Section 3:

Developing a Growth Mindset
CHAPTER 3
CHARACTERISTICS OF THE INNOVATOR’S MINDSET

With continuous advances in technology and learning, a threeyear absence from teaching could easily seem like thirty to an educator who is returning to the classroom. It’s no surprise, then, that Lisa Jones, an accomplished high school teacher in Peel School district just outside of Toronto, Canada, felt as if she had fallen behind when she first came back to teaching after a three-year maternity leave. That feeling intensified when she pulled out the overhead projector.

Only a few years earlier, the overhead projector was considered to be a crucial technology in education. But as Lisa switched out the transparencies for her mitosis lesson, she suddenly realized how boring it was for her. The methods she had previously used to teach these concepts were not meeting the needs of her students today.

As Lisa shared her concern, it was obvious to me that she wanted to get better. She knew that, if she were a student, she would have hated the traditional mitosis lesson she had just delivered. She also knew she didn’t want to limit her students’ opportunities to the knowledge she had at the time—knowledge that hadn’t yet caught up with the current technology available to her and her students. As we talked about her options, I asked her to consider going onto Twitter and connecting with other science educators using the hashtag #scichat. Since she was new to Twitter, I showed a video called “Twitter in 60 Seconds” to give her a quick, introductory explanation of how teachers can use Twitter.
A week later, Lisa sent me the following tweet: "@gcouros you inspired me to do this lesson—here is one of the products!" She included a link to a student-produced YouTube video titled “Mitosis in 60 Seconds.”

I was blown away by the progress Lisa made in such a short time. With a desire to create new and better learning experiences, and a model of a new way of learning, she took a risk to make that happen. After watching the “Twitter in 60 Seconds” video, Lisa showed it to her students and asked, “Could you do this for mitosis?” Although she wasn’t confident she could create the video herself, Lisa knew her students would be able to figure out the technology—and maybe she could learn from them. Notice that Lisa still taught the mitosis content, but with an innovative and effective twist. In fact, she told me later that it was the first time every single student in her class passed the course.

Lisa’s experience is one example that proves we do not have to sacrifice innovation because of expectations or limitations. This goes back to the idea of being innovative inside of the box—in this case, creating new and better methods to teach the required curriculum. My belief is that great educators can work within the constraints of the system and still create innovative learning opportunities for their students. In fact, I think it is necessary to do so.

I believe that the majority of educators want to create engaging and effective learning experiences for their students to achieve the desired learning goals. No teacher wakes up in the morning and says, “I can’t wait to go to school and be mediocre!” I talk with teachers all the time and I know they desire to do great things for the students they serve. But at the same time, there are tremendous demands placed on teachers today and, too often, the narrative I hear from so many administrators is, “Teachers don’t want to change!” My belief is not that teachers don’t want to change, but they sometimes lack clear guidance and support to make the desired change. Effective leadership in education is not about moving everyone from one standardized point to the next but moving individuals from their point “A” to their point “B.”

**GREAT EDUCATORS CAN WORK WITHIN THE CONSTRAINTS OF THE SYSTEM AND STILL CREATE INNOVATIVE LEARNING OPPORTUNITIES FOR THEIR STUDENTS.**

When we approach leadership with an innovator’s mindset, we lead with empathy—meeting people where they are—to help them find or create solutions that work for them. If I, as an education and innovation coach, had told Lisa she’d better use technology or demanded to see specific grade outcomes for her course, she, like many teachers feel compelled to do, may have clung to what she knew and hoped for the best—even if that meant the lesson was boring and ineffective.
Empathizing with Lisa to understand her goals and providing both a model and encouragement empowered her to create a new learning experience—for her and her students.

**THE 8 CHARACTERISTICS OF THE INNOVATOR’S MINDSET**

I've identified eight crucial characteristics that are necessary for an innovator's mindset, not only for teachers but also for everyone involved in education. As you read through them, consider how Lisa demonstrated each of these characteristics in her teaching. (Many thanks to Sylvia Duckworth, @sylviaduckworth, for her illustration of these characteristics.)

1. **Empathetic**—In the previous chapter, I mentioned that educators have exceptionally high expectations for their own learning experiences because they focus on creating powerful learning opportunities for students every single day. High expectations are not bad, in fact, they can be used to our advantage. And it's why I regularly ask educators to consider the question: **Would you want to be a learner in your own classroom?**

   Empathetic teachers think about the classroom environment and learning opportunities from the point of view of the student, not the teacher. Lisa, for example, understood that the way she originally delivered the information did not allow for the students to interact with the content in a meaningful way. It was familiar and easy for her, but it didn't engage her learners' needs or appeal to their interests. To remedy that, Lisa flipped the classroom from being teacher-centric to learner-centric. She shifted her role and became the architect of the experience, while her students actually created the learning for themselves.

   Because of Lisa's willingness to develop relationships with her students, she understood them. She knew that at least some of them knew how to create media, and enjoyed doing so, but did not necessarily connect that process to what they did in school. By tapping into their interests, she crafted an innovative learning experience.

2. **Problem Finders/Solvers**—For years, schools have posed questions or problems to students, often in a linear model that forces students to follow certain steps to find the answer. The world is not step-by-step or linear; it's complex and often requires a messy solution. Sometimes it takes several attempts and iterations to solve real-life problems, and, sometimes, there are several correct answers. But solving a problem is only one part of learning. Educational thought leader Ewan McIntosh notes that finding the problem is an essential part of learning—one that students miss out on when we pose the problem to them first.
Currently the world’s educational systems are crazy about problem-based learning, but they’re obsessed with the wrong bit of it. While everyone looks at how we could help young people become better problem-solvers, we’re not thinking how we could create a generation of problem finders.4

We see elements of problem finding in Lisa’s approach. She felt her students were not experiencing deep learning, and she knew something was wrong, even though no one was challenging her to do anything different. Although she suggested the idea of creating a video, she also gave them the freedom to explore other options for sharing their learning in a compelling way. Additionally, she challenged her students to figure out how to create the presentation on their own.

Teaching students by example to be self-starters and to continuously evaluate how they might improve their education helps them learn how to effectively learn. When we stop simply telling students how to learn, and, rather, act as a “guide on the side,” we can support them in a way that encourages them to find their own solutions.

3. Risk Takers—Innovative teaching and learning involve taking risks. If we are to create new opportunities for the learners we serve, not everything we try will always work with every learner. If we’re honest, we’ll admit that’s always been the case. Many of our known “best practices” don’t serve a large number of our students. So why would we take risks in the classroom, especially when it may affect our students’ futures? The answer lies in the question. Risk is necessary to ensure that we are meeting the needs of each unique student. Some respond well to one way of learning, while others need a different method or format. Not taking the risk to find the best approach for each student might seem less daunting than trying new things, but maintaining the status quo may have dire consequences for our students.

When we think of risk, we shouldn’t just challenge what doesn’t work. It’s also necessary to question our “best practices.” Lisa could have continued using the overhead projector and transparencies, and some of her students would have done extremely well on any assessment given. But did they really gain deep knowledge about concepts and how to apply them? Or did they simply learn to regurgitate information? In many cases, we have conditioned kids to “schooling.” Accustomed to and successful in controlled learning environments, some students may fear being educated (and assessed) in any other way. But they don’t start out fearful. Think about how curious kindergarten students are and how many questions they ask when they first walk into school! I can guarantee that none of those little ones ask for a worksheet. We condition them to that. If we really want to serve our students and help them to develop into the leaders and learners of today and the future, taking risks in our practice is not only encouraged but necessary.

An educator with an innovative mindset will find the balance between drawing on experience while maintaining a willingness to try something new. Apple started by making great computers, and computers are still a major part of its business. But the company’s
leaders went out on a limb to create the iPhone—a risk that led to one of Apple’s biggest achievements. That “risk” also spawned other successful endeavors, including the App Store, iPad, and the Apple Watch. Likewise, for Lisa, the success her students experienced because of the risk she took has led to her implementing more innovative ideas. What is important is that she feels confident asking the question, “Is there a better way to teach this lesson to meet the needs of these learners?”

4. Networked—Networks are crucial to innovation. CIO.com writer Tom Kaneshige says, “Every idea is fundamentally a network of ideas... When you create an environment that allows the kinds of serendipitous connections to form, [innovative ideas] are more likely to happen.”

Being in spaces where people actively share ideas makes us smarter. Think about all of the actors who flock to Hollywood, or all of the country singers who end up in Nashville, or all of the “startups” that are situated in Silicon Valley. In these networks where so many people with similar interests in an area gather, innovation occurs almost spontaneously.

Face-to-face networking is still crucial and valuable in education, but, today, social media provides a place for ideas to spread. Steven Anderson, well-known educational speaker and writer says, “Alone we are smart, together we are brilliant.” The power of networking is sharing ideas, clarifying our thinking, and developing new and better ideas.

When students come to school, we continually tell them, “You need to share!” because we know the great benefit to their learning. Educators would all benefit if we decided to take our own advice. One way we can do that is through blogs. If you’re thinking, “I’m not a writer,” consider this: every opportunity to share with others on a global scale makes you think more deeply about what it is that you are sharing in the first place. In Clive Thompson’s Wired article “Why Even the Worst Bloggers Are Making Us Smarter,” he explains how having an audience can enhance learning:

Having an audience can clarify thinking. It’s easy to win an argument inside your head. But when you face a real audience, you have to be truly convincing.

Social scientists have identified something called the audience effect—the shift in our performance when we know people are watching. It isn’t always positive. In live, face-to-face situations, like sports or concerts, the audience effect can make athletes or musicians perform better—but it can sometimes psych them out and make them choke, too.

Yet studies have found that the effort of communicating to someone else forces you to pay more attention and learn more. Innovation (and enjoyment) flourishes when teachers collaborate to learn and practice new strategies. Isolation is often the enemy of innovation.

5. Observant—Let’s go back to the example of Lisa and her students. Lisa first connected with other educators through social media. Specifically, she searched the hashtag #scichat to find other teachers on Twitter who were interested in the subject of science. Online, she saw other teachers who pushed the expectations for what could be done by students. Yet her idea for having students create their own “Mitosis in 60 Seconds” videos didn’t come from that network but from watching the “Twitter in 60 Seconds” video. The lesson? Sometimes, the most valuable thing you get from the network isn’t an idea but the inspiration or courage to try something new.

Ideas for education today are not limited to education. Many educators have taken ideas from Google, FedEx, or from any number of YouTube videos, including TED Talks (TED is short
for technology, entertainment, and design). Chris Anderson, the entrepreneur, journalist, and publisher who curates TED Talks, calls the effect of web videos on learning, "crowd-accelerated innovation" (which will be discussed further in Chapter 11). Because of Chris's dedication to exposing others' brilliance to the entire world through online videos, millions of people have heard talks on edu-

SOMETIMES, THE MOST VALUABLE THING YOU GET FROM THE NETWORK ISN'T AN IDEA BUT THE INSPIRATION OR COURAGE TO TRY SOMETHING NEW.

cation done by the likes of Sir Ken Robinson and Rita Pierson. And you are not limited to the minds of those focused solely on the field education. People from a variety of fields, such as Daniel Pink (speaking about motivation), Barry Schwartz (sharing his thoughts on wisdom), or Susan Cain (focusing on the power of introverts), have had a major impact on educators worldwide.

Making connections between the powerful ideas and information that's being freely shared online allows educators to expand learning possibilities for their students. Josh Stumpenhorst is one educator who has done exactly that. After watching Daniel Pink's TED Talk and reading more from him about the topics of motivation and the power of autonomy, Josh created an "Innovation Day" in which his students proposed and worked on their own ideas and inventions. The day was so successful that Josh's students were willing to come to school on a Saturday to continue the learning. If Josh had not been willing to learn from ideas being shared outside of education and connect those ideas to the needs of his students, he might not have created such a powerful learning opportunity.

My dream is that, by creating a culture of innovation in education and sharing our ideas with the world, organizations will look to schools for ideas to become innovative, not the other way around.

Inspiration is everywhere and often in unexpected places; you just have to keep your eyes open.

—Unknown

6. Creators—Anyone can consume information, but that doesn't equate to learning. The Center for Accelerated Learning notes:

Learning is creation, not consumption. Knowledge is not something a learner absorbs, but something a learner creates. Learning happens when a learner integrates new knowledge and skill into his or her existing structure of self. Learning is literally a matter of creating new meanings, new neural networks, and new patterns of electro/chemical interactions within one's total brain/body system.  

For Lisa, the biggest shift in her classroom was from teacher-centric instruction to learner-centric creation. While many students in her class might have been able to regurgitate or re-share their understanding of mitosis, Lisa wanted them to really understand the concept and retain the knowledge. Creating something helped them make a personal connection to the information—an important key for deeper learning. Those connections need to happen every day in our schools. From a teaching standpoint, Lisa focused on the creation of new ideas, which led to the creation of new content and knowledge by the students. Creation is crucial.

With access to a plethora of digital resources and information, it's important to foster a culture of creation versus consumption. Instilling the idea of creation is especially important in light of the growing popularity of the "flipped classroom," in which students
watch some sort of video or connect with a resource at home and do their "homework" at school. The thing is that, whether delivered in person or through a video, the lecture focuses on the consumption of information. What if the "flip" was, instead of students watching a video, they created one in which they shared the objectives they needed to grasp? Consider how much deeper learning could be if "creation" was a non-negotiable in the learning for both us and our students.

7. **Resilient**—I love the Chinese proverb that says, "The person who says it cannot be done should not interrupt the person doing it." Oh, if only an innovative educator’s life were that easy!

For those with an innovator’s mindset, the reality is that their work will constantly be questioned simply because it is something new. Most people are far more comfortable with a known average than with dealing with the unknown—even if the unknown holds great possibility. Innovators will get pushback in the form of comments like, "Let's not throw out the baby with the bathwater." Objections like this are really meant to disguise people's fear of moving forward. With that in mind, innovators must be prepared to move forward, even when the risk of rejection is involved.

In Lisa's example, she ran the risk of her colleagues questioning the wisdom of going outside the lines. Or her administration might have expressed concern about the implications of students sharing their work on YouTube (even though many students already have a presence on there or other social media platforms).

Anything new and different can seem threatening. One thing I have learned is that, when thinking about moving forward, focusing on the question, "What is best for learners?" helps ensure you’re making the right decisions.

Aside from concerns expressed by colleagues, you may face pushback on new ideas from your students. As stated earlier, many students are so accustomed to school that projects that veer outside the traditional lines of education terrify them. School can easily become a checklist for our students (complete homework, tests, rubrics, graduation requirements, etc.). In contrast to multiple choice tests, learning that focuses on creation and powerful connections to concepts not only takes more effort but also more time. Yet, if we do not challenge our students in their learning experiences, we aren’t truly preparing them for the real world. Yes, it is important, if not crucial, to listen and respond to our students’ input, but it is equally important to help them become resilient and face adversity. The school environment is the perfect place to challenge and encourage them to stretch their thinking; and it's a safe place for them to try, fail, and try again.

Resilience is a necessary trait for innovators, but it's also a skill that all humans need to develop. Life is full of ups and downs. How you recover from failure and move forward is important to how you learn and how you live. As you push the edges of the norm with your innovative ideas, hold onto your conviction and passion. If you don’t believe in your idea, why would anyone else?

8. **Reflective**—Reflection, not only in innovation but also in education, is a practice to which we need to pay more attention. It is a process that is crucial to innovation, as it ensures we're asking valuable questions, such as "What worked? What didn’t? What would I/we change? What questions do I have moving forward?"

Questioning our efforts, progress, and processes is crucial to innovation. By constantly revisiting our learning in any space, we
find areas that can be tweaked, modified, reiterated, or even reinvented. Looking back is crucial to moving forward. Reflection also helps us, as educators, to make our own connections, and again, deepen our learning.

In education, how often do we actually build in reflection to our learning? We have a lot of schools that do variations of “Drop Everything and Read” (DEAR), a concept that encourages students to read and consume information. But few schools focus on encouraging students or educators to “Drop Everything and Reflect.” How might we all be impacted if we took time out of each day to the think about what we have learned and how it impacts our next steps?

HOW DOES AN INNOVATOR’S MINDSET AFFECT STUDENTS?

The hope that our students become innovators in schools that do not have educators who embody the same mindset is, at best, wishful thinking. If we do not model these characteristics and the willingness to innovate inside of the box, why would our students do anything different? They won’t.

We cannot limit innovation. Trying things that push us out of our comfort zone while determining every step our students take in our classrooms will not help them think differently. It is essential that we learn when to step in and when to step aside and allow—and even create—opportunities for our students to come up with their own types of learning opportunities.

If you go on to social media platforms, like YouTube or Vine, you will see students creating things we never could have imagined in our younger years. When I see these amazing creations, I always wonder whether these students are creating and developing these innovative ideas because of or in spite of schools? Our role is to empower students to see themselves as innovators who take responsibility for their own learning and leading.

What we model is what we get.

-Jimmy Cosec

As leaders, we cannot tell others they should be innovative while we continue to do the same thing. The characteristics we look for in our teachers and our students—empathizing, problem finding and solving, risk-taking, networking, observing, creating, bouncing back, and reflecting—should be embodied in our work as well. As a principal, I did not simply recreate what my former principals did. Instead, I constantly asked the question, “Would I want to be in a community where I was the principal?” Working from that empathetic mindset, allowed me to keep the teachers’ perspective in mind. There were lots of things that I took from my former administrators, but, I will have to admit, there are some things I chose not to replicate because I hated them as a teacher. If I hated them, why would my staff enjoy them?

Working with Lisa reminded me that educators are not scared of change, but they don’t always feel supported to take the risks in the first place. It’s crucial for administrators to understand that having support and feeling supported are two different things. Administrators often encourage risks while refusing to take chances themselves. Taking risks is also different from “openly taking risks,” where your team can see that you are willing to go out of your way to be an innovator as well. If an administrator takes a risk that no one can see, does anyone learn from it? Probably not. Leaders, whatever their role, will more easily affect change if they allow others to see them taking risks, failing, recovering, and risking all over again.

A NEW WAY OF THINKING

In preparation to write this chapter, I followed up with Lisa once more. She told me that her shift in thinking has changed everything for
her and her students. All it took for her to rethink her teaching was a single opportunity to do something new and experience some success. And that’s a key lesson: innovation is not about changing everything; sometimes you only need to change one thing. That experience can lead to new and better learning opportunities.

MOVING FORWARD

As we close out this chapter, I want to share what I call the mantra of an innovative educator.

I am an educator.

I am an innovator.

I am an innovative educator and I will continue to ask, “What is best for learners?” With this empathetic approach, I will create and design learning experiences.

I believe that my abilities, intelligence, and talents can be developed, leading to the creation of new and better ideas.

I recognize that there are obstacles in education, but, as an innovator, I will focus on what is possible today and where I can push to lead towards tomorrow.

I will utilize the tools that are available to me today, and I will continue to search for new and better ways to grow, develop, and share my thinking, while creating and connecting my learning.

I focus not only on where I can improve, but where I am already strong, and I look to develop those strengths in myself and in others.

I build upon what I already know, but I do not limit myself. I’m open to and willing to embrace new learning, while continuously asking questions that help me move forward.

I question thinking, challenge ideas, and do not accept, “This is the way we have always done it” as an acceptable answer for our students or myself.

I model the learning and leadership I seek in others. I take risks, try new things to develop, and explore new opportunities. I ask others to take risks in their learning, and I openly model that I’m willing to do the same.

I believe that isolation is the enemy of innovation, and I will learn from others to create better learning opportunities for others and myself.

I connect with others both locally and globally to tap into ideas from all people and spaces. I will use those ideas, along with my professional judgment, to adapt the ideas to meet the needs of the learners in my community.

I believe in my voice and experiences, as well as the voice and experiences of others, as they are important for moving education forward.

I share because the learning I create and the experiences I have help others. I share to push my own thinking and to make an impact on learners, both young and old, all over the world.

I listen and learn from different perspectives because I know we are much better together than we could ever be alone. I can learn from anyone and any situation.

I actively reflect on my learning because I know looking back is crucial to moving forward.

If we all embrace this mindset, imagine what education could become.
QUESTIONS FOR DISCUSSION

1. What risk might you take to change learning experiences?

2. How might you create an environment that fosters risk-taking?

3. How do you exhibit the innovator's mindset in the learning and work that you do currently?

NOTES


3. Image used with permission from Sylvia Duckworth, @sylviaduckworth, https://www.flickr.com/photos/15664662@N02/.


DNA is not destiny

Nick Miller

Published: February 19 2012 - 3:00AM

'THEY f--- you up, your mum and dad./They may not mean to, but they do./They fill you with the
faults they had/And add some extra, just for you.'

Philip Larkin's best-known poem sets the scene for one of the most hard-fought, even acrimonious
disputes in science. Ten years ago, the human genome crumbled into its component parts. Here,
we were told, was the blueprint for humanity. A child is a recipe written at conception, pre-
programmed with the genes for cancer, the genes for depression, obesity, politics and sexuality,
lives laid out in advance, assembling themselves in stop-motion as we helplessly watch.

But New York philosopher Jesse Prinz wants to call a halt to the "century of the gene". In a new
book, Beyond Human Nature, he gathers the arguments of a growing number of scientists who
take the side of nurture against nature, in a backlash against the tyranny of DNA.

As hard as it might be for parents to hear, he says, the evidence actually suggests that most of their
child's abilities, personality and future success depend on how and where they are raised.

Prinz is unmissable in the crowded Manhattan cafe where we meet. A mutual acquaintance
challenged me not to mention his hair in print but I have to; it's a faint pink (unlike the bright blue
it was a few months ago). He literally wears on his head a rejection of genetic determinism.

"There was a sense that we were being seduced," he says of the recent flood of publications from
human naturists (as opposed to nurturists). "But there was also a sense that we were missing out
on the most interesting aspect of the human story."

Prinz, a professor of philosophy at the City University of New York, speaks at a million miles an
hour with a persuasive power that surely harks back to his grandfather, a rabbi from Nazi Berlin
who spoke out against Hitler, moved to the States and became an activist and one of the organisers
behind Martin Luther King jnr's march on Washington.

Prinz, of course, dislikes such pat genetic explanations. He is proud of his parents but he prefers to
see his own personality and interests as being shaped by being raised by his intellectual, art-loving
family rather than by grand-parental DNA.

"I think a lot of what we call personality has to do with just the things that excite us in life," he
says. "You might appear to be totally unmotivated, sluggish, bored if you are surrounded by things
that don't interest you. You might appear as an extrovert who's open to new experiences if you are
surrounded by the things that you like."

To understand what Prinz is reacting against, it's worth rewinding a bit. Ten years ago, scientist
and author Steven Pinker wrote a bestseller called The Blank Slate, in which he set out the
arguments for a genetically programmed human nature.

"What is the best predictor that a person will become schizophrenic? Having an identical twin who is schizophrenic," Pinker wrote. "Not societal stress or schizophrenogenic mothers."

Pinker wielded the formidable weapon of the scientific study of identical twins: genetically identical beings.

"Identical twins, whether separated at birth or not, are eerily alike in just about any trait one can measure - verbal, mathematical and general intelligence, in their degree of life satisfaction, personality traits such as introversion, agreeableness, neuroticism, conscientiousness and openness to experience," he wrote. "They have similar attitudes towards controversial issues such as the death penalty, religion and modern music. They resemble each other in gambling, divorcing, committing crimes, getting into accidents and watching television. They even boast dozens of shared idiosyncrasies such as gigling incessantly, giving interminable answers to simple questions, dipping buttered toast in coffee and writing indistinguishable syndicated advice columns."

Twins studies are a powerful tool. By comparing twins with their siblings, scientists have created numbers that measure the "heritability" of a trait - the degree to which it seems to pass from parents to child. One of the "converts" to this field, known as behavioural genetics, is Geoffrey Miller, associate professor at the University of New Mexico. Miller genuinely believes behavioural genetics, along with evolutionary psychology (which explores ancient humanity for the evolutionary pressures that shaped modern human nature), will permanently rewrite the psychology textbooks.

"I felt like it was the most powerful way of looking at human nature and human behaviour that I had encountered and I still think that's the case," he says.

He says twins research, which has really taken off in the past 10 to 20 years, has established that virtually every aspect of personality is heritable.

For example, genetic differences account for about 50 per cent of the difference in general intelligence between children. This actually increases to 80 per cent by old age, Miller says, because higher intelligence tends to reinforce itself by driving inquiry but low intelligence will lead us to "watch reality TV, not read books and go to NASCAR races".

Then there is personality, often broken down by psychologists into separate traits such as extroversion, agreeableness or emotional stability. They turn out to be "moderately heritable", Miller says, somewhere between 30 and 60 per cent, depending on which trait.

The rest doesn't necessarily come from the family home, Miller says. "What you often find is family environment explains 10 to 20 per cent of the remainder [with the rest] basically noise, random effects."

Behavioural geneticists become ever bolder in what they believe genes can influence in our nature. Some studies show heritability in religiosity - how strongly you believe, how often you pray. Other show heritability in political attitudes, whether you incline to liberalism or conservatism. Even getting divorced seems determined by your DNA. As is being a psychopath.

"I think the important thing is for parents to recognise that they aren't at fault [if they happen to] raise a psychopath - though they might have been at fault in choosing the wrong mate," Miller says. But should this inspire despair? Doesn't it mean we can't improve our kids, that the die is set?
"If you're single, it's great news," he says. You just need to choose the right mate. It also means parents can relax: "You feed them and they will grow … you're not going to be able to change their basic intelligence or personality very much."

It even has implications for the education system. "I think the US is wasting hundreds of billions of dollars a year trying to give university educations to young people who can't actually handle university," Miller says.

This is where we need to turn back to Prinz. Because this kind of talk really makes him mad.

"I think the claims being made are often irresponsible and dangerously so," he says. "If you really thought there were biologically based differences in intelligence, the thing you should promote is IQ enhancement through education, because this would equalise the differences. The people who believe in genetic differences in intelligence often oppose affirmative action but it seems like they really should be the biggest supporters of affirmative action."

Prinz says people who argue for these biological stories often feel persecuted by the "political correctness police" and accused of supporting eugenics.

But they're not necessarily bad, he says. Just wrong.

For a start, in *Beyond Human Nature*, he mounts a rigorous attack on the twins studies that are the backbone of the naturist argument. "The heritability numbers are meaningless," he says.

The same mathematics would tell you that girls wearing pink or women wearing lipstick were heritable traits.

The deeper problem is that the separated twins used in the studies are often raised in very similar social and economic families - adoption agencies work hard to make this happen - so naturally the effects of environment will be hard to spot. And when scientists are comparing identical twins with their siblings, they fail to recognise that identical twins are treated far more similarly to non-identical siblings.

What about the amazing stories of coincidences between long-separated twins, the strange parallels in behaviour, of lives that seem to have been almost mystically connected? Surely that's genetic? Well, no, Prinz says. He likens the meeting of long-lost siblings to a first date, when we go through our interests and preferences and fixate on the similarities while brushing over the differences.

"It's like those moments when you hit on, 'Oh you like jazz, too' - suddenly this becomes a justification for your attraction and you can say, 'We have so much in common'," Prinz says. "But any two randomly selected individuals will have lots of very random things in common."

So what does make us who we are? Family, school, friends and most importantly the culture we are raised in, he believes.

A lot of Prinz's arguments come from the relatively unploughed field of cultural psychology. If genes are so powerful, he asks, how come different cultures seem to cause such strong differences in our behaviour?

There are some astonishing examples. People raised in Western countries tend to see the trees before the forest, while people from the Far East see the forest before the trees. In south-east Asia, there is a common form of mental illness, unheard of here, in which people go into a trance-like state after being startled.
One of Prinz's favourite examples involves Australian women. Researcher Stanley Milgram ran a famous series of tests in which he assessed people's level of obedience to authority by making them think they were electrocuting people by following an experimenter's instructions.

"He showed that 65 per cent of Americans were willing to administer what they believed to be a lethal dose of electricity on a stranger [and] the assumption was that he had discovered a human universal, that human beings are, in general, extremely obedient," Prinz says.

"Well, the same study, or a version of it, was run in Australia and the average level of obedience dropped down to 40 per cent and if you looked at Australian women, they were obeying authority to this extreme only about 10 per cent of the time. You can tell something about the spirit of independence and defiance and mistrust of authority [in Australia]; it's deeply entwined in the culture."

And perhaps most astonishingly, the language we speak seems to determine our very thoughts. This is the part of Prinz's book that is likely to offend the most scientists, who believe we have inherited a part of the brain primed to learn language - and they use that as a wedge to argue how other parts of our genetic inheritance can explain our behaviour.

Prinz believes the evidence points the other way - that babies are born "profoundly dumb ... less like their parents and more like the family pet". But they have big brains that can spot all sorts of patterns in the world and if one of those patterns consists of grown-ups constantly trying to communicate with them, they're going to pick it up. "There is not a single way of thinking shared by all human beings - a universal logic of thought. Rather, there are different thinking styles that emerge though enculturation," Prinz argues. "One thing really striking me about research on human behaviour is we have this very obvious variable, we have a very good understanding of how it works, it's very easy to measure and it's hugely influential, namely culture. But instead of going for it we tend to look for something ... biological.

"It's as if, when somebody has a gun to their head and they hand over their wallet to a mugger, we ignore the gun and say, 'Oh they must have been biologically disposed to be charitable', to give over their money to strangers.

"Well, no, they had a gun to their head, there's a glaring variable in their environment."

This has considerable consequences, Prinz says. It means we can, and must, do more to help the disadvantaged, especially when young enough to overcome their initial deprivation.

Anyone can turn out bad if their environment is bad. "Patterns of poverty, of alcoholism, of deprivation of opportunity are going to be much bigger players in governing whether somebody in a biologically isolated group ends up committing acts of abuse," Prinz says.

But also, it means the least promising student may, in the right environment, given the right help, outperform any son or daughter of privilege - studies show that IQ increases three points for every year of schooling, that a 15-point IQ difference can be cut in half by four years of university. It means our personality and intelligence grow as a result of how our parents treat us, how much they speak to us, how much they encourage us to embrace new experiences, how they deal with anxiety in front of us.

"Parents are, in the early years of life, the most important factor in determining intelligence outcomes, because they have so much control over a child's environment," Prinz says. "Parents [should] give their children challenging problems, convey an excitement about learning by modelling that enthusiasm, put a child in contact with materials of instruction that are stimulating."

The scientific debate goes on. But what does a parent make of all this? I thought I should go and ask one. So I spoke to my friend Hayley, who in 2010 had a beautiful daughter called Betty. And I filled Hayley in on some of the high points of Prinz's research.

The first reaction was comedy. After all, what do you expect a mother to do when told how vitally important talking to a young child is for their future academic achievement?

But after some introspection, Hayley emailed me this: "If science was to prove that pretty much personality was created from our environment I think I would feel surprised and annoyed. I'm constantly beating myself up about something ... so finding out that her personality is all 'up to me' is not comforting.

"If we were to find out that it was all genes then I think I would feel more comforted by that.

"I would still beat myself up about all those things but I wouldn't feel like I was failing her as much."

But, Hayley says, she doesn't think any of this will change the way she raises her daughter.

"I'm curious what creature she turns into but at the moment I'm enjoying the ride, watching her grow and taking the cues off her. And telling her that's she's OK - and loved."

Beyond Human Nature by Jesse Prinz is published in Australia next month by Penguin.

Hardback rrp $45

This story was found at: http://www.smh.com.au/entertainment/books/dna-is-not-destiny-20120218-1tyr.html
Section 4:

Assessment
ahead of the curve

The Power of Assessment to Transform Teaching and Learning

Solution Tree Press
Chapter 1

Using Assessments to Improve Teaching and Learning

Thomas R. Guskey

Large-scale assessment programs provide the foundation for nearly every modern education reform initiative. Policymakers and legislators at the state and national levels see assessments as essential for change. They believe that good data on student performance drawn from large-scale assessments will help focus educators' attention and guarantee success, especially if consequences are attached to the assessment results; however, large-scale assessments, like all assessments, are designed for a specific purpose—to rank-order schools and students for the purposes of accountability, and some do that fairly well. But assessments designed for ranking are generally not good instruments for helping teachers to improve their instruction or modify their approach to individual students. Students take these assessments at the end of the school year, when most instructional activities are near completion. Teachers do not receive the results until many months later, and by that time their students have usually moved on to other classrooms with different teachers. Finally, the results teachers receive usually lack the level of detail needed to target specific improvements (Barton, 2002; Hattie & Timperley, 2007; Kifer, 2001).
The assessments best suited to guide improvements in instruction and student learning are the quizzes, tests, writing assignments, and other assessments teachers administer on a regular basis in their classrooms. Teachers trust the results from these assessments because they relate directly to instructional goals in the classroom (see Guskey, 2007). Plus, results are immediate, relevant, and easy to analyze at the individual student level. However, to use classroom assessment to make improvements, teachers must change both the way they view assessment and the way they interpret results. Specifically, they need to see their assessments as an integral part of the instructional process and as an essential element in their efforts to help students learn.

Despite the importance of assessments in education today, few teachers receive much formal training in assessment design or analysis. A survey by Stiggins (1999) showed, for example, that less than half the states require competence in assessment for licensure as a teacher. Lacking specific training, teachers often do what they recall their own teachers doing: They rely heavily on the assessments offered by the publishers of their textbooks or instructional materials. When no suitable assessments are available, they construct their own in a haphazard way, with questions and essay prompts similar to those their teachers used. They treat assessments strictly as evaluation devices, administering them when instructional activities are completed and using them primarily to gather information for assigning students’ grades.

Making Assessments Useful

For assessments to become an integral part of the instructional process, teachers need to change their approach in three important ways: They must 1) use assessments as sources of information for both students and teachers, 2) follow assessments with high-quality corrective instruction, and 3) give students second chances to demonstrate success.
What makes these changes in approach so difficult, however, is that each change compels teachers to depart significantly from the practices they experienced as students. In other words, teachers must think about and use assessments differently than their teachers did.

**Use Assessments as Sources of Information for Both Students and Teachers**

Nearly every student has suffered the experience of spending hours preparing for a major assessment, only to discover that the material he or she studied was different from what the teacher chose to emphasize on the assessment (see Guskey, 2006). This experience teaches students two unfortunate lessons: First, they discover that hard work and effort often do not pay off in school because the time and energy they spent in preparation for the assessment had little or no influence on the results. And second, they learn that they cannot trust teachers (Guskey, 2000b). These are hardly the lessons responsible teachers want their students to learn.

Nevertheless, this experience is a common one for students because many teachers still mistakenly believe that they must keep their assessments secret. As a result, students come to regard assessments as guessing games, especially from the middle grades on. They come to believe that their success in school depends largely on how well they can guess what their teachers will ask on quizzes, tests, and other types of assessments. Some teachers even take pride in their ability to out-guess students. They include questions about isolated concepts or obscure facts just to see if students are reading carefully. Generally, teachers do not include such “gotcha” questions maliciously, but rather they do so often unconsciously because such questions were asked of them when they were students.

Classroom assessments that serve as meaningful sources of information do not surprise students. Instead, they are well-aligned extensions of the teacher’s instructional activities. Such assessments reflect
the concepts and skills the teacher emphasized in class, along with the criteria the teacher provided for how he or she would judge student performance. Ideally these concepts, skills, and criteria are also aligned with state, provincial, or district standards. Students see these types of assessments as fair measures of important learning goals. The results of the assessments facilitate learning by providing essential feedback on students' learning progress and by helping to identify learning problems (Bloom, Madaus, & Hastings, 1981; Stiggins, 2002).

Critics sometimes contend that this approach to assessment is "teaching to the test," but this is not necessarily the case. We have to ask, "What determines the content and methods of teaching?" If a test is the primary determinant of what teachers teach and how they teach it, then they are indeed teaching to the test. This occurs frequently today in schools and districts where student performance, as well as that of teachers, is judged by the results of a single large-scale assessment.

However, if desired learning goals or standards are the foundation of students' instructional experiences, then assessments of student learning are simply extensions of those same goals and standards. Instead of teaching to the test, teachers are more accurately "testing or assessing what they teach." They recognize that if a particular concept or skill is important enough to assess, then it should be important enough to teach. And if it is not important enough to teach, then there is little justification for including it in the assessment.

The best classroom assessments also serve as meaningful sources of information for teachers. Assessments provide teachers with specific guidance in their efforts to improve the quality of their teaching by helping identify what they taught well and what needs work. Gathering this vital information does not require sophisticated statistical analysis of assessment results. Teachers need only make a simple tally of how many students missed each item on the assessment or failed to meet a specific criterion. Figure 1 shows an example of this type of assessment analysis.
<table>
<thead>
<tr>
<th>Item</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem #1</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion 1</td>
<td></td>
</tr>
<tr>
<td>Criterion 2</td>
<td></td>
</tr>
<tr>
<td>Criterion 3</td>
<td></td>
</tr>
<tr>
<td>Criterion 4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem #2</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion 1</td>
<td></td>
</tr>
<tr>
<td>Criterion 2</td>
<td></td>
</tr>
<tr>
<td>Criterion 3</td>
<td></td>
</tr>
<tr>
<td>Criterion 4</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1: Sample Tally of Student Errors on a Classroom Assessment**

State assessment results sometimes provide similar item-by-item information, but concerns about content security and the cost of developing new questions each year often make assessment developers reluctant to offer such detailed information. Once teachers have tallied their students' results, they can pay special attention to trouble spots where large numbers of students made errors (in the example in Figure 1, items 7 and 8, and criterion 2 of problem #1).

When reviewing results, the teacher must first consider the quality of the item or criterion. In other words, is the problem with the instruction or is it with the assessment? Perhaps the question is ambiguously worded or the criterion is unclear. Perhaps students misinterpreted the question or the instructions. Whatever the case, teachers must determine whether or not these items and criteria
adequately address the knowledge, understanding, or skill they were intended to measure.

If teachers find no obvious problems with the item or criterion, then they must turn their attention to their teaching. When as many as half the students in a class answer a clear question incorrectly or fail to meet a particular criterion, it is not a student learning problem—it is a teaching problem. Whatever strategy the teacher used, whatever examples were employed, or whatever explanation was offered, it simply did not work.

Analyzing assessment results in this way means setting aside some very powerful ego issues. Many teachers may initially say, "I taught them. They just didn't learn it!" But with further reflection, most recognize that effectiveness in teaching is not defined on the basis of what they do as teachers. Rather, it is defined by what their students are able to do. If few students learned what was taught, can it be said that the teaching was effective? Can effective teaching take place in the absence of learning? Certainly not.

Some argue that such a perspective puts too much responsibility on teachers and not enough on the students. Occasionally teachers respond, "Don't students have significant responsibilities in this process? Shouldn't students be expected to display initiative and personal accountability? And besides, if they don't get it, that's their fault, not mine. I'm here to teach, and they're here to learn."

Indeed, teachers and students share responsibility for learning. Even with valiant teaching efforts, we cannot guarantee that all students will learn everything excellently. In fact, only rarely do teachers find items or assessment criteria that every student has answered correctly. There are always a few students who are unwilling to put forth the necessary effort, but these students tend to be the exception, not the rule. If a teacher is reaching less than half of the students in the class, the teacher's method of instruction needs to improve. And
teachers need this kind of evidence to help target their instructional improvement efforts.

**Follow Assessments With High-Quality Corrective Instruction**

If assessments provide vital information for both students and teachers, then it makes sense that they do not mark the end of learning. Assessments must be followed by high-quality corrective instruction designed to help students remedy whatever learning errors identified with the assessment (see Guskey, 1997). To charge ahead knowing that certain concepts or skills have not been learned well would be foolish. Teachers must therefore follow their assessments with instructional alternatives that present those concepts in new ways and engage students in different and more appropriate learning experiences.

Using high-quality corrective instruction is not the same as re-teaching, which often consists simply of restating the original explanations louder and more slowly. Instead, the teacher must use approaches that accommodate differences in students’ learning styles and intelligences (Sternberg, 1994). Although teachers generally try to incorporate different approaches when they plan their lessons, corrective instruction extends and strengthens that work. Students who have few or no learning errors to correct should also participate in enrichment or extension activities to help broaden and expand their learning. Materials designed for gifted and talented students are an excellent resource for such activities.

Developing ideas for high-quality corrective instruction and enrichment activities can be difficult, especially if teachers believe they must do it alone. Fortunately, they do not have to. Colleagues are some of the best resources for developing teaching strategies. Structured professional development opportunities can help teachers share strategies and collaborate on teaching techniques (Guskey, 1998, 2000a). Faculty meetings devoted to examining classroom
assessment results and developing alternative strategies can be highly effective. District-level personnel and collaborative partnerships with local colleges and universities are valuable resources for ideas and practical advice.

Occasionally, teachers express concern that if they take class time to offer corrective instruction, they will need to sacrifice curriculum coverage. But this need not be the case. Initially, corrective work must be done in class, under the teacher's direction. Efforts to involve students in corrective instruction once per week or during special study sessions conducted before or after school rarely succeed (see Guskey, 1997). In addition, teachers who ask students to complete corrective work independently, outside of class, generally find that those students who most need to spend time on corrective work are the least likely to do so. For these reasons, early instructional units will require more time, typically an extra class period or two.

However, as students become accustomed to this process and realize the personal benefits it offers, the teacher can drastically reduce the amount of class time allocated to corrective work and accomplish much of it during review sessions or with homework assignments. By not allowing minor errors to become major learning problems, teachers better prepare students for subsequent learning tasks, and thus less time is required for corrective work (Whiting, Van Burgh, & Render, 1995). For this reason, instruction in later learning units usually can proceed at a more rapid pace. By pacing their instructional units more flexibly, most teachers find that they need not sacrifice curriculum coverage to offer students the benefits of high-quality corrective instruction.

**Give Students Second Chances to Demonstrate Success**

To become an integral part of the instructional process, assessments cannot be a one-shot, "do-or-die" experience for students. Instead, assessments must be part of an ongoing effort to help
students learn. If teachers follow assessments with high-quality corrective instruction, then students should have a second chance to demonstrate their new level of competence and understanding. This second chance determines the effectiveness of the corrective process while also giving students another opportunity to experience success in learning, thus providing them with additional motivation.

Writing teachers have long recognized the many benefits of a second chance. They know that students rarely write well in their initial attempt. So these teachers build into the writing process several opportunities for students to gain feedback on early drafts and then to use that feedback to revise and improve their writing. Teachers of other subjects frequently balk at the idea—mostly because it differs from their personal learning experiences.

Some teachers express concern about the fairness of giving students a second chance and point out that, "Life isn't like that." They describe how a surgeon does not get a second chance to perform an operation successfully and a pilot does not get a second chance to land a jumbo jet safely. Because of the very high stakes involved, each must get it right the first time.

But how did these highly skilled professionals learn their craft? The first operation performed by that surgeon was on a cadaver, which clearly allows a lot of latitude for mistakes. Similarly, the pilot spent many hours in a flight simulator before ever attempting a landing from the cockpit. Such experiences allowed these professionals to learn from their mistakes and improve their performance. Similar instructional techniques are used in nearly every professional endeavor. Only in schools do students face the prospect of one-shot, do-or-die assessments, with no chance to demonstrate what they learned from previous mistakes.

All educators strive to help their students become lifelong learners, and to develop learning-to-learn skills. What better learning-to-learn skill is there than learning from one's mistakes? Mistakes
should not mark the end of learning; rather, they can be the beginning. Some assessment experts argue, in fact, that students learn nothing from a successful performance. Instead, they learn when their performance is less than successful, for then they can gain direction about how to improve (Wiggins, 1998).

Other teachers suggest that it is unfair to offer the same privileges and high grades to students who require a second chance as we offer to those students who demonstrate a high level of learning on the initial assessment. After all, these students may simply have failed to take responsibility or prepare appropriately for the assessment. Certainly we should recognize students who do well on the initial assessment and provide opportunities for them to extend their learning through enrichment activities. But those students who do well on a second assessment have also learned well. More important, their poor performance on the first assessment may not have been their fault. Maybe the teaching strategies used during the initial instruction were inappropriate for these students, but the corrective instruction proved more effective. If we determine grades and give special privileges (for example, honor roll membership) on the basis of performance and these students have performed at the same high level, then they certainly deserve the same grades and privileges as those students who scored well on their first try.

A comparable example is the driver's license examination. Many individuals do not pass this examination on the first attempt. On the second or third try, however, they reach the same high level of performance as others did on their first attempt. Should these drivers be penalized for not showing appropriate responsibility or being inadequately prepared? Should they, for example, be restricted to driving in fair weather only? Should they be required to pull their car over in inclement weather and park until the weather clears? Of course not, because they eventually met the same high performance standards as those who passed on their first attempt. The same
should hold true for students who show that they, too, have learned well.

The critical issue is this: What is the purpose of a grade? Is the purpose to punish students for not providing the teacher with precisely what was expected on the first try? If so, then a low grade due to inadequate performance in an initial attempt may be justifiable. If this is indeed the case, however, then the teacher must make this purpose clear and must be prepared to defend this purpose to everyone involved—students, parents, school officials, and others. On the other hand, if the purpose of the grade is to provide an accurate description of how well students have learned, then a different outlook is required. In this case, what students know and are able to do become the basis of the grade, rather than how or when they learned the information. From an educational perspective based on what is most helpful to students, this is clearly a more sound, defendable, and equitable position.

A Familiar Process

Using assessments as sources of information, following assessments with corrective instruction, and giving students second chances to demonstrate their learning may seem unfamiliar at first, but most teachers already do these things when they tutor individual students. If the student makes a mistake, the teacher stops and points out the mistake, then he or she provides the student with immediate feedback on the error. The teacher then re-explains the concept or understanding in a different way to help the student remedy the mistake. Finally, the teacher asks another question or poses a similar problem to ensure the student understands before moving on. The challenge for teachers is to use their classroom assessments in similar ways to provide all students with this sort of individualized assistance.
Successful coaches use exactly the same process. Immediately following a gymnast's performance on the balance beam, for example, the coach explains to her what she did correctly and what could be improved. The coach then offers specific strategies for improvement and encourages her to try again. As the athlete repeats her performance, the coach watches carefully to ensure that she has corrected the problem. Teachers who see their classroom assessments as the same type of demonstration of learning can help their students use the results to likewise improve their performance.

Successful students typically know how to take corrective action on their own. They save their assessments and review the items or criteria they missed. They rework problems, look up answers in their textbooks or other resource materials, and ask the teacher about ideas or concepts they do not understand. Less successful students rarely take such initiative. After looking at their grades, they typically crumple up their assessments and deposit them in the trash can as they leave the classroom. Teachers who use classroom assessments as part of the instructional process help all of their students do exactly what the most successful students have learned to do for themselves.

**A Vital Component**

Using classroom assessment to improve both teaching and student learning is not a new idea. Nearly 40 years ago, Benjamin Bloom showed how to conduct this process in practical and highly effective ways when he described the practice of mastery learning (Bloom, 1968, 1971). And despite the relatively modest change required to implement the process, extensive research demonstrates that it can have exceptionally positive effects on student learning. A study by Whiting, Van Burgh, and Render (1995) representing 18 years of data gathered from over 7,000 high-school students showed mastery learning to have a remarkably positive influence on students' test scores and grade-point averages, as well as on their attitudes toward school and learning. Another field experiment conducted in elementary-
middle-school classrooms showed that the implementation of mastery learning led to significantly positive increases in students' academic achievement and self-confidence (Anderson et al., 1992). Even more impressive, a comprehensive, meta-analysis of the research on mastery learning concluded:

Few educational treatments of any sort were consistently associated with achievement effects as large as those produced by mastery learning.... In evaluation after evaluation, mastery programs have produced impressive gains. (Kulik, Kulik, & Bangert-Drowns, 1990, p. 292)

Some researchers even suggest that the superiority of Japanese students in international comparisons of achievement in mathematics operations and problem-solving may be due largely to Japan's widespread use of instructional practices similar to mastery learning (Nakajima, 2006; Waddington, 1995). Research evidence also shows that the positive effects of mastery learning are not limited to cognitive or achievement outcomes. The process yields improvements in students' confidence in learning situations, school attendance rates, involvement in class sessions, attitudes toward learning, and a variety of other affective measures (Guskey & Pigott, 1988). This multidimensional impact has been referred to as mastery learning's "multiplier effect," which makes it an especially powerful tool in school-improvement efforts.

Assessments are a vital component in our efforts to reform and improve education. But as long as we use them only as a means to rank schools and students, we will miss out on their most powerful benefits. We must focus instead on viewing assessments in a different way, considering assessments for a broader array of purposes, and changing the way we use assessment results. As teachers, we must improve the quality of our classroom assessments to ensure that they are well-aligned with valued learning goals and state or district standards. When teachers' classroom assessments become an integral
part of the instructional process and a central ingredient in efforts to help students learn, the benefits of assessment for both teachers and students will be boundless.

References


Using Assessments to Improve Teaching and Learning


Anne Davies

Dr. Anne Davies is an author, consultant, and researcher who applies her expert knowledge of developing quality classroom assessments to her mission to increase the possibility of learning for all students. Educators at every grade level in Canada, the United States, and elsewhere have benefited from the high-touch support she provides during professional development events and multiyear projects. A world-renowned keynote presenter and professional development consultant, Dr. Davies’ genuine care and commitment to supporting educators and their important work make her an approachable and insightful specialist. Dr. Davies is the author and coauthor of more than 24 books and multimedia resources as well as numerous chapters and articles.

In this chapter, Dr. Davies explores the four research-based cornerstones for thoughtfully and deliberately involving students in the classroom assessment process to support learning:

- Formative classroom assessment
- Feedback
- Motivation
- Summative evaluation

Dr. Davies then examines four specific strategies in depth, using real school examples to illustrate how educators can:

- Define learning destinations so that students understand their achievement goals.
- Involve students as partners in co-constructing criteria.
- Multiply the amount of feedback students receive to “feed-forward” their learning.
- Engage students in collecting, selecting, reflecting on, and presenting evidence of their learning.

For more information about Dr. Anne Davies, visit www.annedavies.ca.
Assessing Student Learning: Why Reform is Overdue

Geoff Masters, ACER
Assessing student learning: Why reform is overdue

Geoff N Masters
Australian Council for Educational Research

Advances in our understanding of human learning require new approaches to assessing and monitoring student learning.

Much assessment thinking has changed little over the past fifty years. The field continues to be dominated by twentieth century introductory textbook concepts, including such dichotomies as formative versus summative assessment, criterion-referenced versus norm-referenced testing, quantitative versus qualitative assessment, informal versus formal assessment — distinctions that often hamper rather than promote clear thinking about assessment.

Assessment practice also has changed little over this period. Traditional, high-stakes examinations continue to dominate what is taught and learnt in many of our schools and universities. Greater use is now being made of promising new technologies, including banks of online assessment tasks, computer adaptive tests and technology-based assessments of ‘new’ life skills and attributes. However, while emerging technologies are capable of providing more innovative and informative explorations of student learning, much electronic assessment remains pedestrian and underpinned by traditional assessment thinking.

At the same time, progress in our understanding of learning itself is challenging long-held assumptions and pointing to the need for a paradigm shift in assessment theory and practice.

For example, substantial progress has been made in our understanding of human capacity for learning. It once was believed that individuals differed significantly in their capacity to learn. But research in neuroscience has shown how the plasticity of the brain enables almost all individuals to learn throughout the lifespan. This finding parallels the educational conclusion that, although students are at different points in their learning and are progressing at different rates, almost all students are capable of successful learning if motivated and provided with appropriate learning opportunities and support.

Research also is making clear the enormous variability in students’ levels of achievement and progress. Children begin school with very different social, cognitive, psychomotor and language development. Many of these differences do not disappear. In any given year of primary school, differences in reading and mathematics achievement are the equivalent of five or six years of school. And in some areas of learning and development, variability appears to increase across the school grades.

We also know that, in mixed-ability classrooms, students learn best when provided with learning opportunities matched to their varying interests and progress. Learning is maximised when tasks are targeted just beyond individuals’ current levels of attainment — in the region where success is possible, but often only with scaffolding and support.

An implication of these observations is that educational assessment is best conceptualised as a process of discovering where learners are in their learning and development. Although it is common to refer to the ‘multiple purposes’ of assessment, assessment has only one fundamental purpose: to establish where learners are in their progress at the time of the assessment. This information can then be interpreted and used in a variety of ways. For example, students’ achievements can be interpreted by reference to the performances of other students nationally or internationally, by reference to achievement expectations or standards, or by reference to past performances to study trends or growth over time. The results of assessments can be used to inform starting points for teaching, to evaluate the effectiveness of educational programs and interventions or to award qualifications. For teaching purposes, it sometimes is desirable to obtain more detailed information to diagnose specific student misunderstandings or errors, but once again, the single underlying purpose is to discover where learners are in their learning. Much unnecessary complexity has been introduced into the assessment literature through failure to recognise and begin with this simple truth.
The process of establishing where students are in their learning depends on a thorough understanding of the learning terrain through which they are progressing: typical paths of development; sequences in which understandings normally are established; and side-tracks in the form of common errors, learning difficulties and misunderstandings. Assessment as the discovery of where students are in their learning requires much more than familiarity with the intended curriculum. It depends on expert understanding of how learning occurs in a domain – a reference ‘map’ that is built from research and knowledge about learning itself.ii

Essential to this approach to assessment is an appreciation of learning as ongoing progress. At the heart of all educational effort is the intention of student growth, development or improvement. Rather than being limited to specific courses, semesters or years of school, the progress that students make usually occurs incrementally over extended periods of time. For example, in areas such as reading, mathematics and science, progress typically occurs across the entire period of schooling. The role of assessment should be to establish where students are on these long-term continua of learning and what progress they are making over time.

To establish where students are in their learning, evidence is required, usually in the form of observed performances on classroom activities or assigned assessment tasks. However, individual tasks are rarely, if ever, of intrinsic importance. Students may never again have to read and answer questions about the particular piece of text or solve the particular mathematics problems used in an assessment. Specific tasks are merely convenient but interchangeable vehicles for collecting evidence about what is really of interest – a student’s underlying reading ability, for example, or level of achievement in an area of mathematics. And establishing where students are in their learning always involves an on-balance inference with an accompanying degree of uncertainty.

This conceptualisation of assessment stands in stark contrast to the traditional use of assessment to determine how much of what a teacher has taught each student has successfully learnt. Traditional assessments are made not in relation to an understanding of long-term learning progress, but in relation to a specific corpus of taught content. The onus is on students to learn this content and the role of assessment – whether during or upon completion of a course – is to judge how well they have done this. Conclusions about ‘how much’ students have learnt commonly are expressed as percentages, which may then be converted to grades to convey the extent of each student’s success (or failure).

Under traditional approaches, it is common to treat ‘curriculum, teaching and assessment’ as separate activities. The role of teachers is to teach the curriculum, the role of students is to learn, and the role of assessment is to judge how much of the taught content students have learnt. By contrast, a view of assessment as professional investigation sees assessment as an integral part of good pedagogy. This view is consistent with the role of assessment in other professional work – for example in medicine and psychology – where the purpose is not so much to judge as to understand for the purpose of making informed decisions.

Research into learning highlights the need for investigative approaches to assessment. Learning is rarely, if ever, a process of passively taking in and storing new information. Even from a very young age, learning is a process of actively trying to make sense of the world. Learners interpret what they see and hear in terms of what they already know. They construct their own mental models and understandings which are sometimes inaccurate or only partially correct. And it is clear that misconceptions, if not identified and addressed, can be significant obstacles to further learning.iii

Research also shows that students sometimes can succeed on traditional forms of assessment while holding fundamental misconceptions. For example, physics students can sometimes recall formulae and substitute numerical values correctly to answer examination questions while holding fundamental misunderstandings about relationships between force and motion.

Studies comparing experts and novices in various fields show that what distinguishes experts from novices is not only extensive knowledge of a field, but also the frames of reference that experts have for organising and making sense of that knowledge. Experts have deep understandings of concepts, principles and big ideas in a field which allow them to see patterns in information...
and to transfer their knowledge to new and unseen contexts.

The implications of these research findings are that educational assessments must do more than establish whether students can reproduce what they have been taught, and teachers must be more than deliverers of curriculum content and judges of student success. The investigative process of establishing where students are in their learning must include an exploration of students’ understandings of important concepts and principles. An appreciation of learners’ own mental models and misunderstandings can provide important starting points for teaching (ie, assessments for learning). Assessments of factual and procedural knowledge will continue to be important, but perhaps more important in the future will be the assessment of students’ abilities to organise and use this knowledge and to apply their understandings to the solution of complex, real-world problems.

In the past, assessment methods often have been more concerned with judging success and making reliable and fair comparisons of student performances than with investigating and understanding student learning. And the desire for large-scale implementation under standardised conditions, with a quick turnaround of results, often has resulted in assessments requiring only that students reproduce what they have been taught through the provision of ‘correct’ answers.

Some educators have reacted against assessments of this kind by arguing that ‘authentic’, in situ assessments are always preferable to assessments based on specially-designed assessment tasks, or that ‘school-based’ assessments made by classroom teachers are always preferable to externally-developed assessments. But these are over-reactions. When the purpose of assessment is to explore and understand where students are in their learning, there must be a willingness to use the methods best able to provide this information, whatever form they take.

Day-to-day observations made by classroom teachers generally provide the richest information for establishing where students are in their learning. Ideally, teachers would have intimate and precise knowledge of each student’s progress and learning needs and would use that knowledge to personalise and focus their teaching efforts, often by grouping students with similar needs. As noted already, assessments of this kind depend on expert understandings of the relevant learning domain as well as professional skill in exploring learning progress.

Advances in technology are making it possible to incorporate professional knowledge of this kind into more sophisticated tools for investigating learning. Rather than testing only factual and procedural knowledge, these tools explore student thinking, including by testing hypotheses about misunderstandings and gaps in an individual’s learning. Intelligent forms of assessment in the future will be less concerned with judging how much a student has learnt and more concerned with diagnosing and understanding the details of an individual’s learning.

Research in neuroscience and cognitive psychology also is revealing the important role of emotions in learning. People are more likely to learn and to remember if intrinsically motivated and emotionally engaged. In classroom settings, learning is promoted by ‘learning cultures’ in which all students are expected to learn successfully, are highly engaged and feel safe and supported in their learning. Conversely, negative emotions such as stress and fear of failure have been shown to impede learning and memory. In classroom settings, these emotions can be the result of ‘performance cultures’ in which learning is extrinsically motivated and students compete with each other for success.

Other research has shown the importance of positive attitudes and beliefs about learning. Learners are more likely to learn successfully if they believe that they are capable of learning – in other words, if they have positive views of themselves as learners. They also must believe that effort will result in success. Effective learners are more likely to monitor their own learning, to recognise what they do not know and to be proactive in seeking out what they need to make further progress. Learners are assisted in these processes by relevant and timely feedback that guides action and enables them to see the progress they are making over time.

These research findings relating to emotions, attitudes and beliefs have implications for how
assessments of learning are conducted and how the results of assessments are reported and used.

Some forms of assessment promote ‘performance’ rather than ‘learning’ cultures. For example, one-off, end-of-course examinations usually are designed to judge and compare students on the amount of course content they have learnt – often for the purposes of ranking and selecting students for the next phase of education – rather than to monitor and understand learning progress. In such assessments, learning can be driven more by external pressure for results than by curiosity and intrinsic motivation. And this pressure often distorts teaching and learning by encouraging cramming and creating unacceptable levels of stress for students and their families.

_The paradigm shift now required in assessment is from judging how much of a body of taught content students have successfully learnt to establishing where students are in their long-term learning and what progress they are making over time._

For this reason, one-off, high-stakes assessment events probably have a limited future in the assessment of student learning. In some contexts, there will continue to be a need to ensure that minimum performance standards have been met, but such assessments could be undertaken when learners feel ready to be assessed rather than in a single assessment event.

There are significant implications, too, for methods of reporting and monitoring student learning. Traditional reporting methods, such as percentages and grades, are more consistent with ‘performance’ than ‘learning’ cultures. Percentages and grades are used to describe how much of a body of taught content students have learnt. But these reporting methods are incapable of showing learning progress, and indeed usually mask progress. A student who receives a ‘D’ year after year is given no sense of the progress they are actually making. And worse, they are likely to infer from this outmoded method of reporting that there is something stable about their capacity to learn: they are a ‘D’ student.

It sometimes is argued that students and parents ‘understand’ A to E grades; but they do not because course grades usually do not represent consistent, interpretable levels of achievement. Grading is more appropriate for describing the quality of agricultural produce or the products of industrial manufacturing than for describing learning. The educational challenge is to develop ways of reporting that show where students are in their long-term learning, what progress they are making (ie, assessments of learning) and what might be done to support further learning.

Finally, the uses to which assessments are put also can encourage ‘performance’ rather than ‘learning’ cultures. Assessments conducted to understand and promote student learning can be undermined and distorted when the results of those assessments are then used for other, unintended purposes. For example, external attempts to use test results to drive performance inevitably change classroom teachers’ attitudes and behaviours. There is growing evidence that the linking of rewards and sanctions to test results not only fails to produce the desired improvements, but also results in a range of responses that are inconsistent with what we now know about effective teaching and learning._

---


Towards a growth mindset in assessment

Follow this and additional works at: http://research.acer.edu.au/ar_misc

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
http://research.acer.edu.au/ar_misc/17
Towards a growth mindset in assessment

Geoff N Masters
Australian Council for Educational Research

The approaches we take to assessing learning, the kinds of tasks we assign and the way we report success or failure at school send powerful messages to students not only about their own learning, but also about the nature of learning itself. Assessment and reporting processes shape student, parent and community beliefs about learning—sometimes in unintended ways.

This essay describes three general approaches to evaluating and providing feedback on the outcomes of learning. Each approach is based on a particular way of thinking about what it means to learn successfully, and each has implications for how students view themselves as learners and how they understand the relationship between effort and success. It is argued that commonly used approaches frequently send unhelpful messages.

1. Providing ‘success’ experiences

The first approach is based on tasks chosen because they are within students’ capabilities and are likely to be completed successfully. Underpinning this approach is a belief that, if students are given tasks on which they are likely to succeed, then the resulting success experiences will make learning more pleasurable, increase engagement, build self-confidence and lead to further learning success. In contrast, the experience of failure is assumed to make learning less pleasurable, lower self-confidence and lead to disengagement and thus poorer learning outcomes.

Because, under this first approach, students are assessed on tasks chosen to ensure a high probability of success, most students perform well and so receive praise for their performance. By praising success, teachers endeavour to promote positive attitudes, build self-esteem and encourage all students in their learning.

There are several unintended consequences of this approach. First, when teachers assign tasks only within students’ current capabilities, they risk not challenging and stretching students and minimising learning by keeping students within their comfort zones. There is considerable research evidence that learning is most likely when students are given challenging tasks just beyond their comfort zone, in what Vygotsky (1978) called the ‘zone of proximal development’, where success is possible, but often only with assistance.

Second, when teachers praise students for success on easy tasks, they risk sending the message that success at school can be achieved with minimal effort. Rewarding success on unchallenging tasks does little to develop students’ understandings of the relationship between effort and success.

Third, by providing success experiences for almost everybody, this approach can encourage the view that success is an entitlement—that every student is a good learner and is entitled to good results and positive feedback. By protecting students from failure, this first approach does little to develop healthy attitudes to risks, challenges, mistakes and failure.

Psychologist Carol Dweck argues that, rather than giving students easy tasks within their comfort zones and providing praise for succeeding on these tasks, teachers should be communicating to students that unchallenging tasks are a waste of time:

Many educators think that lowering their standards will give students success experiences, boost their self-esteem, and raise their achievement... Well, it doesn’t work. Lowering standards just leads to poorly educated students who feel entitled to easy work and lavish praise.

(Dweck, 2006, 193)

2. Judging performances against ‘standards’

The second approach has been developed as a response to the first. Underpinning this second approach is a belief that, by specifying ‘standards’ to be achieved by all students in each year of school, and by judging and reporting performances against these standards, learning expectations and thus achievement levels will be raised.
The appeal of this approach is that it sets clear expectations for student performance. Grounded in the well-established industrial processes of specifying quality standards, judging performances against standards and grading products for their quality, this approach has particular appeal to politicians because it can be represented as rigorous (setting explicit standards against which performances are to be judged) but also fair (equitable in the sense that it holds all students to the same expectations).

This approach has the added advantage of being consistent with the way society generally thinks about schooling and what it means to succeed or fail at school: the role of teachers is to teach the curriculum specified for the year level, the role of students is to learn what teachers teach, and the role of assessment is to establish how much of what they have been taught students have successfully learnt. Students who demonstrate most of the expectations for their year level are rewarded with high grades; students who demonstrate few of those expectations receive low grades and may be judged to have ‘failed’.

The problem with this second approach is that it suffers from many of the same disadvantages as the first. It often is no better at helping students understand the relationship between effort and success. It often does not provide students with stretch challenges. And it often encourages fixed mindsets about learning ability.

How is this possible? The answer lies in the variability of students’ achievement levels within each year of school. In any given year of school, the most advanced 10 per cent of students typically are between five and six years ahead of the least advanced 10 per cent of students (Harlen, 1997; Masters & Forster, 1997; Wiliam, 2007). Children begin school at very different points in their social, cognitive, emotional and psychomotor development. Many of these differences persist throughout the years of school. As a consequence, rather than being at a similar stage in their learning, students in any given year of school are in reality spread over a wide range of achievement levels.

This is not to say that students who are at different stages in their learning are not making good personal progress. They often are. It is simply that less advanced students are tracking five to six years behind the most advanced students. And these relativities tend to be maintained across the years of school. One of the best predictors of student achievement in the later years of school is achievement in the earlier years.

We may wish that this were not the case. It may be our intention that all students of the same age should be at very similar points in their learning and development. However, the reality in our schools is that this is not the situation, and almost certainly never has been. The problems with the second approach arise from the attempt to ignore this fact.

In reality, students commence each school year with very different levels of readiness for the year-level curriculum that teachers are about to teach. Some are still several years behind. Inevitably, these students struggle, master less of the year-level curriculum than other students, and are judged and graded accordingly. Often these students perform below the year-level standard year after year. In fact, there is some evidence that, in mathematics, less advanced students, on average, fall further behind each year (Wiliam, 2007; Masters, 2013).

When students’ performances are graded against year-level expectations, some less advanced students can receive the same low grade year after year. The feedback these students receive is that they are consistently performing below standard and below other students. A to E grades provide little or no sense of the learning progress that individuals actually make over time. A student who receives a ‘D’ year after year could be excused for concluding that they are making no progress at all when, in reality, they may be making as much annual improvement as a student who consistently receives an ‘A’. And worse, they may conclude that there is something stable about their capacity to learn – that is, they are a ‘D-student’. Such demotivating messages undermine students’ beliefs in the relationship between effort and success and frequently lead to disengagement. As Grenny et al., (2013) observe, for many less advanced students, ‘dropping out [of school] is a sane response to persistent disappointment and repeated reminders that they’re performing below average’.

© 2013 Australian Council for Educational Research
19 Prospect Hill Road, Camberwell, VIC 3124
AUSTRALIA

www.acer.edu.au
However, the problems with this approach are not limited to less advanced students. They apply equally to more advanced students. When learning expectations are couched only in terms of year-level standards, these common expectations can fail to challenge and extend more advanced students. For example, in some secondary schools it is common for all entering students to be taught the same mathematics curriculum and to be assigned the same mathematics tasks during their entire first year. (Some schools justify this on the grounds that it gives them a year to ‘sort students out’.) This practice inevitably disadvantages more advanced students who are ready for more challenging work.

And, in some classrooms, it is common for students to be given ‘free time’ when they complete set class work. Rather than extending more advanced students with challenging, more difficult material, this practice makes the completion of assigned class work the common goal for all students. (In fact, there is anecdotal evidence of reluctance on the part of some teachers to give additional work to more advanced students because this could be interpreted as a form of ‘punishment’ for finishing set work early.)

Adding to this concern is a finding by Patrick Griffin and his colleagues at the University of Melbourne that teachers are less able to identify intervention strategies to assist more advanced students. These observations may explain why more advanced students, despite receiving higher grades, do not always make as much progress in their learning as less advanced students. In their study of progress in reading and mathematics, Griffin and colleagues concluded:

- **Students at the bottom levels of the proficiency scale are improving rapidly.**
- **Students at the top end of the scale are hardly improving at all.**

(Griffin et al., 2013, 5)

Observations of this kind also may help to explain why the decline in achievement levels at 15 years of age over the past decade has been greatest among more advanced students (Thomson et al., 2011).

And there is a risk of these students, too, developing unhelpful beliefs about the relationship between effort and success. Because they begin each school year five to six years ahead of some other students, more advanced students sometimes achieve high grades with limited effort. These students can develop a belief that, because they are ‘smart’ — that is, ‘A-students’ — they do not have to make an effort in the way that other students do. And, as Carol Dweck observes, there is no research evidence that more advanced students are more inclined than less advanced students to enjoy challenges or to extend themselves.

This second approach – assessing, judging and grading student performances against year-level ‘standards’ – was intended to challenge and motivate students, encourage effort and raise achievement levels. In practice, it often has the opposite effect on student attitudes and behaviours. The costs to learning and achievement in our schools are potentially significant and certainly justify the search for an alternative.

### 3. Assessing ‘growth’ over time

The third approach is focused on establishing the points that individuals have reached in their learning, setting personal stretch targets for further learning, and monitoring the progress that individuals make over time. Underpinning this approach is a belief that, at any given time, every student is at some point in his or her learning and is capable of further progress if they can be engaged, motivated and provided with relevant learning opportunities. Rather than expecting all students of the same age to be at the same point in their learning at the same time, this approach expects every student to make excellent learning progress over the course of a school year, regardless of their starting point. In other words, this third approach sets high expectations for every student’s ‘growth’.

Carol Dweck refers to this way of thinking as a growth mindset:

*When [teachers and students] change to a growth mindset, they change from a judge-and-be-judged framework to a learn-and-help-learn framework. Their commitment is to growth, and growth takes plenty of time, effort and mutual support.*

(Dweck, 2006, 244)
Towards a growth mindset in assessment

When students’ performances are assessed from the perspective of a growth mindset, the focus is not so much on ‘judging’ as on understanding where individuals are in their learning at the time of assessment. What knowledge, skills and understandings do they currently demonstrate, regardless of how other students are performing or what the intentions may be for students of this age or year level? To answer this question it may be necessary to investigate and diagnose in some detail the difficulties that individuals are experiencing or the misunderstandings that they have developed.

Assessment information of this kind provides starting points for teaching and learning. It enables learning activities to be selected and designed to maximise the likelihood of successful further learning. It also assists teachers and students to set targets for learning. Rather than being based on common year-level expectations, these learning targets are personalised; they set realistic stretch challenges for individual learners.

When assessments provide information about where students are in their learning at the time of assessment, they also provide a basis for monitoring individual progress over time. Assessments of progress are an alternative to judging success only in terms of year-level standards. Under a growth mindset, success is defined in terms of the progress each student makes, or the ‘distance travelled’.

Importantly, the adoption of a growth mindset does not represent a lowering of expectations. On the contrary, it sets high expectations of every learner, including more advanced students who sometimes are not challenged or stretched and hardly improve at all. Under a growth mindset, ‘failure’ is defined not in terms of year-level expectations, but as inadequate learning progress.

The adoption of a growth mindset also invites a change in thinking from a belief that there are ‘good learners’ who meet year-level expectations year after year, and ‘poor learners’ who perform below standard year after year, to a belief that, although students may be at different points in their learning and may be progressing at different rates, all are capable of good learning progress.

And, when learning is evaluated in terms of the progress that individuals make, the relationship between effort and success is clarified. Students’ self-confidence is built, not through success on easy tasks, but when they are able to see the progress they are making, when they appreciate how the quality of their work has improved, and when they succeed on challenging tasks that once were beyond them.

Many existing learning frameworks provide a basis for assessing student growth. School curricula that define clear progressions of learning across the years of school make explicit what long-term growth in a domain looks like, and so provide a basis for establishing individuals’ current levels of attainment and for monitoring growth over time. Such a range of empirically-based ‘proficiency scales’ and ‘developmental continua’ (Masters, 2013).

No small challenge

This essay has argued for defining, assessing and reporting school learning in terms of the progress that individuals make. However, this is no small challenge. Success at school usually is assessed not in terms of the progress that individuals make (for example, over the course of a school year), but by judging and grading performances against age/ year group expectations. Although letter grades are a relatively recent phenomenon — they appeared for the first time in some North American higher education institutions in the late 19th century and were widely used in schools only in the 20th century — they have come to define what it means to learn successfully at school. Reform depends first on a change in mindset.

Added to this is the challenge of developing credible and easily understood alternatives to current reporting practices. The kinds of reports called for in this essay would provide information about: (1) where students are in their learning at the time of assessment (eg, what they currently know, understand and can do); and (2) how much progress they have made over some specified time (eg, a school year, a semester). Good reporting alternatives of this kind generally do not exist. In their absence, the practice of reporting success in terms of year-level expectations is often justified on the grounds that parents wish to know how students are performing in relation to others of the same age. However, this may be less true if parents also had good information about where exactly students are in their learning and what progress they are making over time.
Changing mindsets and developing assessment and reporting tools to support such change are long-term educational agenda. They almost certainly require a transition phase in which processes based on differing mindsets operate in tandem. A starting point is a wider appreciation of the ways in which efforts to provide ‘success’ experiences and to evaluate learning in terms of common year-level ‘standards’ fail to engage and challenge some students and encourage fixed rather than ‘growth’ mindsets in our schools.

References


Section 5:

Creativity
Artful Thinking

"If a picture is worth a thousand words, a painting must be worth two thousand."—Arianna Bonnes, 9th grade

In New Bedford, Massachusetts, Chris Jones' history class has been studying the Renaissance. Arianna's comment follows a discussion about a painting that was guided by three simple questions: What do you see in this painting? What do you think about that? What does it make you wonder? Arianna's wonderful insight—that works of art are dense with meaning and rich with communicative power—is evidence that the class discussion was deep and far-ranging, filled with questions, ideas, and meaningful connections. This shouldn't be surprising. When young people engage in extended explorations of works of art, they find it quite natural to do such things as ask provocative questions, make careful observations, explore multiple viewpoints, and uncover multiple meanings. Not only are these powerful forms of thinking in the arts, they are also powerful forms of thinking in other areas of learning. This is the basic idea that underlies the Artful Thinking program—that exploring works of art can help students learn how to think in ways that empower learning across the curriculum.

Artful Thinking is a program that helps students learn how to think by looking at art.

Developed by Project Zero at the Harvard Graduate School of Education, Artful Thinking is designed to be used by teachers in any grade and in any subject. The purpose of the program is twofold: To help teachers create rich connections between works of art and their curriculum; and to help teachers use arts experiences as a touchstone for developing students' thinking dispositions. The program focuses on looking at art rather than making art, and it is part of the Visible Thinking Network at Project Zero—a research-based approach to teaching thinking that links several Project Zero initiatives and a growing international network of schools and other learning organizations.

From good thinking to good thinkers: A dispositional approach

Most educators believe that it's important to teach students to think. Traditionally, efforts to teach thinking foreground the teaching of thinking skills—reasoning skills, problem solving skills, and the like. Thinking skills are certainly important. But if we want students to use their skills frequently, in diverse and novel contexts, then simply teaching skills isn't enough. Research at Project Zero and elsewhere has shown that motivation, values, cultural context, and alertness to opportunity are also important factors in developing the intellectual behaviors—the thinking dispositions—that are characteristic of good thinkers.

The Artful Thinking Palette

There are many thinking dispositions worth cultivating—curiosity, open-mindedness, reasonableness, to name just a few. Artful Thinking focuses on a set of six thinking dispositions that have special power for exploring works of art and other complex topics in the curriculum. They are questioning & investigating, observing & describing, reasoning, exploring viewpoints, comparing & connecting, and finding complexity. Each of these dispositions has specific intellectual behaviors associated with it. As a set, the six dispositions are synergistic: Observing naturally leads to reasoning, which connects to questioning, which in turn links to connection-making, and so on. Artful Thinking uses the image of an artist's palette to express this synergy.

Thinking dispositions and thinking routines

Dispositions are formed when people routinely engage in specific patterns of behavior. Accordingly, in the Artful Thinking program, thinking dispositions are developed through the use of "thinking routines"—short, easy-to-learn procedures that help students enact thinking-dispositional behavior in and across the six areas of the palette. For example, recall the three discussion questions that prompted Arianna's insight about a painting being worth two thousand words: What do you see? What do you think about that? What does it make you wonder? These questions comprise a thinking routine that connects to two dispositions on the palette—observing & describing and questioning & investigating. Other thinking routines encourage exploring multiple viewpoints, forming careful interpretations, finding complexity, and so on. Thinking routines are designed to be used flexibly and frequently. Students can use them solo or in small or large group settings, they can be used across subject areas.

THINK / PUZZLE / EXPLORE

A routine that encourages questioning and inquiry

Consider an artwork/topic:
- What do you think you know about this artwork or topic?
- What questions or puzzles do you have?
- What does the artwork or topic make you want to explore?
CREATIVE QUESTIONS
A routine for creating thought-provoking questions

Brainstorm a list of at least 12 questions about the artwork or topic.
Use these questions to help you think of interesting questions.

Why...
What are the reasons...
What is the significance of...
What if...
How would it be different if...
Suppose that...
What if we knew...
What would change if...

matters, and they can be used with a wide range of topics and works of art. Above all, they are designed to deepen students’ thinking about the topic at hand, whether it is a painting, an historical event, or a mathematical operation.

Artful Thinking and Visible Thinking

Too often, students are exposed only to the final, finished products of thought—the finished novel or painting, the established scientific theory, the official historical account. They rarely see the patterns of thinking that lead to these finished products, yet it is precisely these habits of mind that students need to develop. A key part of Artful Thinking involves making students’ thinking visible by documenting their unfolding thought processes as they use thinking routines. Making thinking visible in the classroom provides students with vivid models of what the process of good thinking looks like and shows them how their participation matters.

CLAIM / SUPPORT / QUESTION
A Reasoning Routine

- Make a claim about the artwork or topic.
- Identify support for your claim.
- Ask a question related to your claim.

Artful Thinking and visual art

Thinking dispositions, thinking routines, and visible thinking are emphasized in all Visible Thinking initiatives at Project Zero. Though all Visible Thinking initiatives are art-friendly, Artful Thinking is distinctive in that it was developed to explicitly bring out the connection between art and thinking. There are two reasons for this. The first has to do with how works of art make us think, and the second has to do with what they make us think about.

In terms of how art makes us think, consider the kinds of things we have in mind when we talk about teaching thinking. We want students to learn to ask thoughtful questions, to construct careful explanations, to explore new viewpoints, to see the complexity and dimensionality of the topics they study, to find puzzles worth pursuing, and so on.

These forms of thinking come naturally when looking at art, because art naturally invites them. When Arianna tells us that a painting is worth two thousand words, she’s telling us that works of art are packed with meaning. And she’s right: Works of art are metaphorical, often multi-layered and ambiguous, often full of detail. They express artists’ intentions and their un-intentions and they condense many meanings and purposes. Moreover, works of art are made with the purpose of engaging our attention. Artists generally want us to look and ponder and explore. So one deep connection between looking at art and learning to think is this: By both design and default, art naturally invites deep and extended thought.

Of course works of art are more than simply a powerful vehicle for teaching thinking; they are also important things to think about. The second reason to connect looking at art and learning to think has to do with the meanings of artworks themselves and the multiple ways they connect to the curriculum. Works of art provoke rich, multilayered meaning-making in ways unlike other disciplines. They raise questions, evoke connection-making, and in many ways transform the shape of inquiry. In doing so, they have the power to transform students’ historical inquiry into a personal and contemporary one.

There are many ways to connect art to the curriculum, from targeted connections between the content of artworks and specific topic or themes, to more open-ended approaches that leave loose the directions in which a work of art will lead. Artful Thinking is in favor of any and all curricular connections, so long as students are invited to think directly and deeply about an artwork itself. Art gets shortchanged when it is used superficially merely as illustrative aid to a set of facts, such as when a painting is used simply to illustrate the costumes of a particular era or the geography of a particular region. Artful Thinking avoids this shortfall because thinking routines—the mainstay practice of Artful thinking—are designed to engage students in thinking deeply about the artwork or topic at hand. They allow for the “superficial read,” which after all is part but not all of an artwork’s meaning, but they also push students to unpack the depth and complexity of works of art by inviting them to ask creative questions, make diverse observations, explore multiple viewpoints, and seek personal connections.

Artful Thinking research

For the most part, Artful Thinking is a development and dissemination project. Its purpose has been to create a program for use in schools and other educational contexts. The program has its roots in Project Zero’s long history of research on thinking dispositions, as well as in previous Project Zero projects on teaching thinking, such as Art Works for Schools and Innovating with Intelligence. However, there is also an ongoing research component to Artful Thinking. Our research agenda currently has two goals: To help us understand the effects of the program on student thinking and learning and to help us better understand the nature of thinking and learning more generally. For example, through the use of student-generated concept maps, we are currently investigating how Artful Thinking affects students’ ideas...
about good thinking: What it looks like, what it's for, and the kinds of things students believe "count" as good thinking. Additionally, we are investigating how the program affects students' ability to explore and interpret art and whether these abilities transfer to non-art contexts. We are also investigating how the program affects both students' and teachers' concepts of art—for example, their ideas about art's purposes and meanings, and their ideas about art's connection to learning and to their own lives.

**Artful Thinking past and present**

Artful Thinking was developed in collaboration with Traverse City Area Public Schools (TCAPS) in Traverse City, Michigan, as part of an Arts in Education Model Development and Dissemination grant from the US Department of Education. It is also being used in New Bedford, Massachusetts, in Montgomery County, Maryland, and many of its practices are being used by Visible Thinking sites worldwide.

**Current Project Staff**

Shari Tishman, Patricia Palmer

**Funding for this project has been provided by**

 Traverse City, Michigan Area Public Schools through an Arts in Education Model Development and Dissemination grant from the U.S. Department of Education.
ARTFUL THINKING PALETTE

**REASONING**
What makes you say that?
Claim / Support / Question

**QUESTIONING & INVESTIGATING**
Think / Puzzle / Explore
Creative Questions
See / Think / Wonder

**EXPLORING VIEWPOINTS**
Step inside
Circle of Viewpoints

**OBSERVING & DESCRIBING**
Beginning / Middle / End
Looking: Ten Times Two
Listening: Ten Times Two
Colors / Shapes / Lines
The Elaboration Game

**FINDING COMPLEXITY**
Parts / Purposes / Complexities
Complexity Scale

**COMPARING & CONNECTING**
I used to think...now I think
Connect / Extend / Challenge
Creative Comparisons
Headlines

Artful Thinking | Project Zero, Harvard Graduate School of Education
Thinking Palette

Comparing & Connecting

- Headlines
- Creative Comparisons
- Connect / Extend / Challenge
- I used to think, ...now I think

Exploring Viewpoints

- Circle of Viewpoints
- Step Inside

Finding Complexity

- The Complexity Scale
- Parts / Purposes / Complexities

Observing & Describing

- The Elaboration Game
- Colors / Shapes / Lines
- Looking: Ten Times Two
- Listening: Ten Times Two
- Beginning / Middle / End

Questioning & Investigating

- See / Think / Wonder
- Creative Questions
- Think / Puzzle / Explore

Reasoning

- What Makes You Say That?
- Claim / Support / Question

The framework takes the image of an artist's palette as its central metaphor. Typically, a palette is made up of a relatively small number of basic colors which can be used and blended in a great variety of ways. The artful thinking palette is comprised of 6 thinking dispositions — 6 basic colors, or forms, of intellectual behavior — that have dual power: They are powerful ways of exploring works of art, and powerful ways of exploring subjects across the school curriculum.
The Artful Thinking palette comes alive through the use of “thinking routines.” Each thinking disposition has several thinking routines connected to it. Thinking routines are short, easy-to-learn mini-strategies that extend and deepen students’ thinking and become part of the fabric of everyday classroom life. They are used flexibly and repeatedly — with art, and with a wide variety of topics in the curriculum.
1. **Observe and describe**
   Look at the painting for at least one minute. Let your eyes wander. What do you see? List five words or phrases that describe any aspect of the work.

   Look again
   Move in closer. List five more words or phrases that describe any aspect of the work.

2. **Brainstorm** a list of 3-5 questions about the artwork or topic. Use these question-starters to help you think of interesting questions:
   
   - I wonder...
   - Why...?
   - What are the reasons...?
   - What if...?
   - I am puzzled by...
   - How would it be different if...?
   - Where...?
   - What if we knew...?
   - If I could interview the artist, I'd ask...

   Then, select one question to discuss for a few minutes.

   **Reflect:** What new ideas do you have about the artwork or topic that you didn’t have before?

3. **Connect** Compare the painting with another work in the room. How are the two works similar? Brainstorm a list of 3-5 things that the two works have in common.

   What kinds of connections did you find?

   Why do you think the curator chose to hang these works in the same room?
The Story Routine: Main, Side, and Hidden

A routine for exploring the complexity and depth of an incident, document, painting or photograph

After closely looking at or reading the source document, identify and begin to elaborate on:

1. What is the main or central story being depicted or documented?
2. What is the side story (or stories) happening on the sidelines or around the edges that may not necessarily involve the main characters?
3. What is the hidden story—that other story that may be obscured, neglected, or happening below the surface that we aren’t readily aware of initially?

Purpose: What kind of thinking does this routine encourage?
This routine helps students to dig into events and explore documents in more depth by constructing different narratives related to those events. This may raise new questions of inquiry and wonderings to be explored. The routine also encourages students to identify additional points of view, beyond those of the central characters, which traditionally might not be presented when examining events.

Application: When and where can it be used?
This routine asks students to identify and begin to tell multiple narratives stemming from a photograph they are looking at, a story they are reading, a work of art they are examining, an historical event being discussed, or a newspaper account under review. Use this routine when you want to go beyond the main story itself or want to explore new perspectives or alternative accounts of events, perhaps of marginalized actors. The main story is that which is being centrally presented and often the purpose or main intent of the account or image. Often instruction stops with the main story, but that needn’t be the case. The side story encourages learners to look at what may be happening on the sidelines and to explore the perspective of those supporting the story and impacted by it but perhaps not playing one of the major parts. Alternatively, a side story might be the action happening right alongside the main event but is being overshadowed by it. The hidden story or the other story usually doesn’t present itself directly. One needs to go beneath the surface to identify what isn’t being stated, what is obscured, or what perspective isn’t being heard or seen.

Launch: What are some tips for starting and using the routine?
To uncover these narratives requires both looking closely and looking beyond. As such, it may be useful to combine this routine with See-Think-Wonder, focusing on the See and Think steps to identify the main story as well as what might be happening on the sides. Wonderings could them explore the possible hidden stories. Alternatively, three columns (Main, Side, Hidden) could be drawn on the whiteboard and the class could identify the characters that exist for each section. Once the characters are identified, events and actions can be charted for each column as well. Using this information, students can then craft their own versions of the main, side, and hidden story either orally or in writing. Such accounts would be based on the information identified but enhanced by details that are creatively imagined to produce a work of historical fiction. Alternatively, students might select an emerging side or hidden story to research and present this factual information to the rest of the class to help in understanding the complexity of the event being explored.

Developed for Denso by Ron Ritchhart (2012)
Colors, Shapes, Lines
What are they like? What do they do?
A Routine for Exploring the Formal Qualities of Art

1. Take a minute to look at the artwork. Let your eyes wander over it freely. What do you see? Take a few observations from students and then move on to the next step.

2. Observe and describe the colors, shapes, and lines in detail. Make 3 columns.

<table>
<thead>
<tr>
<th>COLORS</th>
<th>SHAPES</th>
<th>LINES</th>
</tr>
</thead>
<tbody>
<tr>
<td>What colors do you see? Describe them.</td>
<td>What kinds of shapes do you see? Describe them.</td>
<td>What kinds of lines do you see? Describe them.</td>
</tr>
</tbody>
</table>

3. Choose a kind of color, shape, or line that you listed.* How does it contribute to the artwork overall? (How does it help the artwork "work?") Consider:

- How does it contribute to how the artwork feels?
- How does it contribute to the mood of the artwork?
- How does it contribute to how the artwork looks?
- How does it contribute to the story the artwork tells?
- How does it contribute to the ideas in the artwork?

* Do this with at least two elements. They can be chosen from any column.

4. What new ideas do you have about the artwork? What do you see now that you didn’t see before?
# Layers

A routine for structuring analysis of creative works

<table>
<thead>
<tr>
<th>NARRATIVE</th>
<th>The story, the back or pre story, the other or hidden story, the message</th>
</tr>
</thead>
<tbody>
<tr>
<td>AESTHETIC</td>
<td>The appeal (what pulls you in?), the reward or take away, the skill/mastery of the artist on display, the new/different/unusual</td>
</tr>
<tr>
<td>MECHANICAL</td>
<td>Technique, Form/structure, Methods, Symbolism</td>
</tr>
<tr>
<td>DYNAMIC</td>
<td>Surprise, Tension, Emotion and Movement</td>
</tr>
<tr>
<td>CONNECTIONS</td>
<td>To other works (in and out of the medium/genre), to history, to oneself, to the artist’s other works or personal life.</td>
</tr>
</tbody>
</table>

**Purpose: What kind of thinking does this routine encourage?**

The routine provides learners with a structure for looking analytically at creative works through a variety of different frameworks.

**Application: When and Where can it be used?**

There are many layers through which one can approach or look at any creative work (literature, dance, painting, etc.). Some layers may be more appropriate than others given the work being examined. Part of analysis involves selecting appropriate frames or layers to use in one’s analysis. Selecting interesting and unexpected layers can help one understand the work better. Sometimes this means rejecting the obvious layer and starting with one of the other layers.

**Launch: What are some tips for starting and using this routine?**

After looking closely at a creative work to fully notice what is there, students select a “layer” from the list to use in their analysis. This analysis can be done individually, with a partner, or whole group. Initially you may want to introduce one layer at a time with the whole class so that students have some collective experience using the layers. Initially the analysis should be done verbally so that students hear and can build on other’s ideas and contributions. Other possible ways of using the layers are:

1) To identify prominent and hidden qualities: In this work, what layer immediately speaks to you? What makes you say that? Which layers seem more distant? What makes you say that?

2) To compare and contrast: Use the layers to contrast 2 works to see how they relate to one another. Looking at 2 works, where do you see connections as well as differences in terms of the layers?

3) As a sieve: Pick one element from each of the layers through which to explore the work.

4) As a source of questions: Use the layers and their elements to identify questions you want to ask an expert about the work of art.
SIMPLEXITY

A routine for exploring the connection between the simplicity and complexity

Identify instances of complexity and simplicity and then explore their connection

SIMPLICITY
Where does it appear in the object/event/plan/design? Where does it hide and lay unseen? How does each of these complex elements contribute to the whole?

COMPLEXITY
Where does it appear in the object/event/plan/design? Where does it hide and lay unseen? How does each of these complex elements contribute to the whole?

SIMPLICITY
- How do the simple and complex elements relate to each other?
- How does the simple give rise to the complex?
- How is the complex made simple?
- What dominates your experience of this object/event/plan/design?
  The complex or the simple?

Purpose: What kind of thinking does this routine encourage?

The routine provides learners with a structure for uncovering complexity and looking closely. Through analytic looking, learners explore how elements of simplicity and complexity are not opposites but often work together in the creation and realization of an object/event/plan/design.

Application: When and Where can it be used?

This is a framework that may be used from anything from analyzing a story to a work of art to an everyday object. It is particularly well suited to examining issues around design. Designers (of apps, hardware, buildings, etc.) often encounter complexities that they must find elegant, simple solutions for. Likewise, the execution of simplicity is sometimes complex. This routine helps students develop greater appreciation for these two qualities and how they function and may be helpful before their own engagement in the design process.

Launch: What are some tips for starting and using this routine?

You may want to begin by exploring quick associations with words “complexity” and “simplicity” to help students see the many different ways each may appear. Explain that “simplicity” is looking at the connection and relationship between these two ideas; how one relates to the other. (Note: there are other definitions of simplicity as well, such as how simple things become more complex over time). Pick an object or image that has both complex and simple elements operating together. Have students working in pairs or as a whole group identify those elements. You can either consider simplicity or complexity separately or at the same time. Push students by asking where simplicity or complexity might be hidden from view. If students have missed an important element, draw their attention to it and ask how they would classify it and why.

Using some or all of the simplicity questions above, explore how the simple and complex elements identified relate to each other.
# Five Continua for Assessing Thinking

<table>
<thead>
<tr>
<th>Continuum One</th>
<th>BEYOND THE GIVEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBVIOUS</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>States the obvious, just scratches the surface, doesn’t stretch to go beyond the given information or surface story.</td>
<td>Probes beneath the surface, reaches beyond the obvious, stretches for new observations, questions, connections.</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Continuum Two</th>
<th>ELABORATED</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIMPLISTIC</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Under-described, no detail or nuance, overly broad, overly simplified, overly generalized.</td>
<td>Rich in detail, evocative, imaginative, nuanced, descriptive</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Continuum Three</th>
<th>MULTI-DIMENSIONAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONE-DIMENSIONAL</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Only explores one layer or dimension. Doesn’t see complexity, layers, or multiple viewpoints.</td>
<td>Explores several layers or dimensions Recognizes complexity, recognizes that there are different levels, layers, or perspectives.</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Continuum Four</th>
<th>GENERATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESTRICTED</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Closure-oriented. Doesn’t seem to want to explore big ideas or ask hard questions. Resists exploring possibility of bias, resists open-ended inquiry.</td>
<td>Expands or extends thinking in new directions. Broadens understanding, opens up new lines of inquiry. Often reflects curiosity and openness.</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Continuum Five</th>
<th>ESSENCE-CAPTURING</th>
</tr>
</thead>
<tbody>
<tr>
<td>TANGENTIAL</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Doesn’t capture or recognize important themes, characteristics or elements. Strays from the topic. Hovers over unimportant details or ideas.</td>
<td>Insightful, captures the heart of things, identifies key themes, characteristics or elements. Sees deep structure. Shows an appreciation for the relative importance of things.</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
ARTFUL THINKING: BEGINNING/MIDDLE/END AND PERCEIVE/KNOW/CARE ABOUT

The Artful Thinking approach, designed by Project Zero at Harvard University, teaches students how to think critically through the use of simple routines that guide them to develop processes for observing and analyzing art. The art is a force for developing thinking skills and thinking dispositions that students can then apply across the curriculum. Artful thinking involves several thinking strategies. In this lesson, students will practice two critical thinking approaches: Beginning/Middle/End and Perceive/Know/Care About.

Grade Level
Adaptable for all grades

Common Core Academic Standards
- CCSS.ELA-LITERACY.CCRA.R.1
- CCSS.ELA-LITERACY.CCRA.SL.1
- CCSS.ELA-LITERACY.CCRA.W.1
- CCSS.ELA-LITERACY.CCRA.W.9

PA Academic Standards for Art
- Anchor Standard #7: Perceive and analyze artistic work
- Anchor Standard #8: Interpret intent and meaning in artistic work

Art Images Required

Click on the titles below to view high-resolution photographs on the Philadelphia Museum of Art website. Images that are also available in the Artstor Digital Library are indicated by an ID number or search phrase. Entering that number or phrase into the Artstor search bar will direct you to the corresponding image in that database.

- My Friends, 1887, by Viggo Johansen
  - Artstor search: Not available
- The World Upside Down print series, 1952, by Antonio Frasconi
  - Artstor search: Not available
- Figures in a Landscape, 1972-73, by Sidney Goodman
  - Artstor search: 1974-112-1

For more information, please contact Division of Education and Public Programs: School and Teacher Programs by phone at 215-684-7580, by fax at 215-236-4063, or by e-mail at educate@philmuseum.org.
Lesson Process

BEGINNING/MIDDLE/END

1. Direct students to spend a minute observing My Friends by Viggo Johansen and to consider, on their own, what might be going on in the picture.

2. Have students share their ideas with a partner and then open the discussion to the entire group. Each time a student shares an interpretation, ask that student what he/she saw in the picture that supports the idea.

3. Once key details and a few different interpretations have been shared, distribute the Beginning, Middle, or End handout (attached at the end of this lesson). Instruct students to consider whether this scene belongs at the beginning, middle, or end of a story and follow the instructions on the handout to expand upon what they see. Depending on your time limit, students could use the three boxes for quick notes, a detailed story, or even a cartoon.

PERCEIVE/KNOW/CARE ABOUT

1. Direct students to spend a minute observing selections from the World Upside Down print series. Have each student select an image that attracts his/her attention, and ask them what in that image is unexpected or surprising.

2. Select one work from the series to model the thinking routine. Ask a volunteer to select a person, animal, or thing in the image and ask the group to imagine themselves inside that point of view.

   a. Start by asking, "What can the person or thing perceive?" You may want to define 'perceive' by telling students that it covers everything the person or thing might be aware of through their senses, including sight, sound, smell, etc. For example, students could note what direction the character is facing and what that character might be focusing upon.

   b. As a group, consider, "What might the chosen person or thing know or believe?" (This asks students to interpret context clues and develop a more complete story. They should use visual clues such as expression, body language, props, and setting to support their ideas.)

   c. Finally, ask the group, "What might the chosen person or thing care about?" (This asks them to empathize with the character and consider the emotional impact of the scene and the relationships between different people, animals, or objects.)

3. Once you have modeled the routine, divide students into pairs and distribute the Perceive/Know/Care About handout. Each pair of students should try to select characters with opposing viewpoints from the World Upside Down print series. The pairs should work on their own to fill in the handout, but will collaborate later in the lesson. (NOTE: Alternate version: Ask students to write monologues from the point of view of their chosen character and have the other students guess who they are writing about from the context clues provided in the monologue.)

For more information, please contact Division of Education and Public Programs: School and Teacher Programs by phone at 215-684-7580, by fax at 215-236-4063, or by e-mail at educate@philamuseum.org.
Assessment

1. Through individual journals, encourage students to consider what they have learned about art and about critical thinking by responding to the following prompts:
   “I used to think . . . now I think . . . about art.”
   “I used to think . . . now I think . . . about how art connects to things we study in school.”
   Discuss journal responses.

2. Present to the class Sidney Goodman's 1972–73 painting Figures in a Landscape. Ask them to respond in writing to the prompt, “What's going on in this picture?” Have students choose one of the two thinking routines described above and respond to the painting using that routine.

For more information, please contact Division of Education and Public Programs: School and Teacher Programs by phone at 215-684-7580, by fax at 215-236-4063, or by e-mail at educate@philamuseum.org.
BEGINNING, MIDDLE, or END

A ROUTINE FOR OBSERVING AND IMAGINING

Choose one of these questions:

- If this artwork is the beginning of a story, what might happen next?
- If it is this artwork is the middle of a story, what might have happened before? What might be about to happen?
- If this artwork is the end of a story, what might the story be?

Use your imagination to tell a story in words or a cartoon in the boxes below. One box should cover what you think is going on in the artwork itself and the other two help expand on the story.
PERCEIVE, KNOW, CARE ABOUT
A ROUTINE FOR GETTING INSIDE VIEWPOINTS
Choose a person, animal, or thing from a work of art.
What can the person or thing perceive?

What might the person or thing know about or believe?

What might the person or thing care about?

For more information, please contact Division of Education and Public Programs: School and Teacher Programs by phone at 215-684-7580, by fax at 215-236-4063, or by e-mail at educate@philamuseum.org.
What would your character be thinking or saying? Choose one bubble and write from their point of view. Use your notes about what they perceive, know, or care about to enrich your writing.

For more information, please contact Division of Education and Public Programs: School and Teacher Programs by phone at 215-684-7580, by fax at 215-236-4063, or by e-mail at educate@philmuseum.org.
An Artful Thinking Classroom

Jessica Ross
Harvard Graduate School of Education and
The United Learning Charter Public School

I projected the image onto the screen and gave my eighth-grade students several quiet minutes to take notes in their journals before I began asking questions.

"What do you see?" I asked. They answered in rapid fire: "Dark colors, tombstones, guitar, casket, horse skeleton, dusty red skies, corner of holy people, skeleton pit, naked guy, angry mob, brights, mountain of fire..."

"So, what do you think this might be about?" This time, they took a minute to look back at their journals before sharing responses.


After I captured some of their responses on chart paper, I concluded with, "And what does this image make you wonder?" These answers came quickly and didn't require aid of their journals. "I wonder why this was created, what was the motivation? What does triumph mean? I wonder if they are in war. I wonder if this ties into what we are studying next. I wonder what the time period is."

My students and I spent over thirty minutes that day analyzing The Triumph of Death by Pieter Bruegel the Elder. First they observed and described what they saw, and then they compared and connected their observations to previous knowledge. And finally, they questioned and explored viewpoints through their wonderings. We all looked at that image many times that day, going back to it whenever someone in the class noticed something that someone hadn't seen before or made a new connection.

The Triumph of Death
The students' thoughts were captured and documented both individually (in their journals) and collectively (on chart paper). Once they started to share their thoughts out loud, I scribbled new connections and extensions that were made as the discussion moved beyond the initial observations that were noted in journals.

The work of art, along with the See/Think/Wonder routine, and the documentation of their thinking was used to launch a unit about medieval history. For my eighth-grade humanities class, I designed the unit based on both World History I and English Language Arts curriculum frameworks and then used the Artful Thinking approach to shape an arc of instruction with specific thinking goals in mind.

Shari Fishman, the Artful Thinking project's Principal Investigator, has described the goals of the program as follows:
Artful Thinking was developed to explicitly bring out the connection between art and thinking. There are two reasons for this. The first has to do with how works of art make...
us think, and the second has to do with what works of art make us think about. By both
design and default, art invites deep and extended thought.

The five components of the Artful Thinking program include:
Thinking Routines
Thinking Dispositions
Art & Curricular Connections
Visible Thinking/Documentation
Study Groups: reflective professional practice

What Happens in an Artful Thinking Classroom?
Thinking Routines
It all starts with thinking routines. There is a wealth of these available, and many
teachers already employ them in their classrooms. I decided to introduce the unit with
the See/Think/Wonder routine, which first asks students to make careful observations
and then interpretations based on those observations. These are accessible entry points
even for a student who doesn’t have specific background knowledge. Once observations
are reported, students then begin to consider what the information reveals; finally, they
are asked what they wonder—which opens the door to further inquiry.

Thinking Dispositions
As the students progressed through the unit and gained knowledge about Europe during the
medieval period, I made choices about subsequent thinking routines based on the
thinking dispositions I was aiming for. Thinking dispositions are an inclination towards
certain modes of intellectual behavior. The six thinking dispositions that are
instrumental in considering art and subject area content include:
- Reasoning
- Finding Complexity
- Exploring Viewpoints
- Questioning & Investigating
- Observing & Determining
- Comparing & Connecting

The introductory routine I used, See/Think/Wonder, is a means to get at the disposition
of questioning and investigating. As the students began to learn about the Middle Ages, I
used another routine, Connect/Extend/Challenge to have them synthesize information
from their readings by connecting the new information to what they’ve learned before, to
notice what information extended their thinking, and to make note of ideas that were
puzzling or that challenged their thinking.

The Connect/Extend/Challenge routine does reinforce the disposition of questioning and
investigating, but is more heavily grounded in the disposition of comparing and
connecting, which is why I chose to use that routine once the students had a stronger
foundation of content knowledge.

Art (among other things) and Curricular Connections
Throughout the unit we examined text, images, and primary source documents using a
variety of thinking routines in order to both learn about the time period and develop
thinking dispositions. For example, as we added to our knowledge base, I pushed the
notion of exploring viewpoints by having the students compare two paintings depicting
women from different social classes.

Visible Thinking/Documentation
http://www.highschoolhigh.org/unesco/issue10/artful_thinking_classroom/
The next step was to decide how to document the process of our learning by making our thinking visible at various stages. We did this through journals, wall charts, discussions, and presentations of learning. This documentation served as both formative and summative assessment during the unit. I had a wide range of data gathered from each learning experience to determine the levels of individual and group understanding about the content and about dispositional skill building. The student work also highlighted misconceptions, lack of historical perspective, and confusion on the part of students. For example, the students were more familiar with the class and social distinctions of early American history than with medieval European history, and while they could use their knowledge to make some connections to the plight of peasants and serfs during the Middle Ages, it was necessary to use a variety of sources for them to appreciate the women’s sphere or the role of the church during that time period.

Study Groups: Reflective Professional Practice

So that I could appropriately challenge all of my students, address their misconceptions, and get feedback on the next steps of my plan, I brought examples of student work to my study group. Reflective teacher practice is a vital component of the Artful Thinking approach. Study group members use the “Looking at Student Thinking,” (L.A.S.T.) protocol devised to carefully examine student thinking in order to support teachers as they consider the next steps of the learning process (see Appendix). According to Tishman (2007), “This jump—from making student thinking visible, through thinking centered activities, to shaping instruction so that it further enhances student performance—is assessment in the most authentic sense.”

When my colleagues and I looked at documentation, we noticed that some students struggled to consider the content using a historical perspective. With this insight, I returned to primary sources that the students had examined already, like the Hereford Map and the map, the largest known medieval map. We took another look, using a different thinking routine and revisited the concept of worldview from a medieval European perspective.

Finally, toward the end of the unit, we used the Headlines routine to capture some of the big ideas we had learned that could have been newspaper headlines from the time period. Students suggested: “Foreign Invaders Attack. While One Rules, the Rest Suffer. The Epidemic Continues. Death Fights Back. Serfs and Slaves Suffer...” The Headlines routine also helps to develop the disposition of comparing and connecting. The students are asked to capture core ideas and basic things up. We used these headlines as a transition point to “fellow the stories” as we left the Middle Ages and moved towards the Renaissance.

Reflection

Adopting Artful Thinking involves weaving together many of the successful practices that teachers already use: thinking routines, assessing with documentation, setting dispositional goals, making learning visible and engaging in reflective teacher practice through study groups. The most daunting and rewarding challenge for the teacher is the new viewpoint on the role of art in the classroom. It becomes a matter of trust, not in picking the perfect art piece to illustrate a content standard, but rather trusting the ability of art to surface thought-provoking questions, connections, and areas of inquiry. It also requires trust to learn from the process with your colleagues and from your students.

I have been using thinking routines with art, objects, poems, nonfiction text, charts, graphs, cartoons, film, you name it, for the past seven years. Sometimes when I look at the student thinking that I collect, I realize I could have picked a different thinking routine to get at the disposition I was trying to develop and so I adjust the next time. I encourage teachers to just dive in and play, don’t fret over picking the perfect art piece, or text example, or thinking routine. Have fun, invite the students to share their thinking about a piece of work and you will be on your way towards creating an Artful Thinking classroom.

References


Tishman, S., & Palmer, P. (2007), Works of art are good to think about: A study of the impact of the Artful Thinking program on students’ and teachers’ concepts of art, and students’ concepts of thinking. In Evaluating the Impact of Arts and Cultural Education. Parks Centre Portsmouth, 89-101 (in French and English)

The Artful Thinking program was developed at Project Zero (1972), a research group at http://www.highechohigh.org/issue10/artful_thinking_classroom/
the Harvard Graduate School of Education (HGSE). For more information, visit:

Project Zero Website: http://www.pz.harvard.edu

or

Artful Thinking Website: http://www.old-pz.gse.harvard.edu/tc
Thinking About Thinking

Preservice Teachers Strengthen Their Thinking Artfully

By Susan L. Barahal

Educators often take thinking for granted. After all, thinking is something that we do naturally, without "thinking" much about it. Teachers expect their students to think about what they learn and to learn to be good thinkers.

But what does it mean to be a good thinker? How can teachers know when students are thinking deeply? Can teachers help students become better thinkers? Indeed, how many teachers even discuss the topic of thinking with students? And what do we want students to learn and think deeply about?

For two years, I’ve explored the topic of thinking with small groups of preservice teachers to learn how looking at and exploring art images could deepen and strengthen their own thinking and that of their students. Over time, these preservice teachers shifted in their thinking about thinking and became more aware of the importance of deep thinking. They showed greater appreciation for the value of identifying “big ideas” in their respective subject disciplines and of rethinking their curriculum and classroom lessons around these ideas.

What exactly does it mean to think critically? Since thinking is “invisible,” how can teachers identify it and recognize when it’s taking place?

During these group discussions, I often asked my preservice teachers if they wanted their students to be good thinkers, and they seemed surprised and a bit perplexed by my question. Of course they wanted their students to be good thinkers. In fact, they said they wanted their students to be “critical thinkers.”

But what exactly does it mean to think critically? Since thinking is “invisible,” how can teachers identify it and recognize when it’s taking place? Finally, how can teachers model good thinking to their students and create a classroom environment that en-
courages thinking and provides the time that thinking takes?

In our group meetings, we discussed the challenges our preservice teachers faced in their new roles as teachers and explored possible strategies that could inform their teaching and help them and their students become good thinkers. I also introduced my preservice teachers to the concept of Artful Thinking in order to learn whether it could strengthen their thinking and help put them in touch with the “big ideas” in their subjects that would lead to more meaningful classroom lessons.

**ARTFUL THINKING**

Project Zero at the Harvard University Graduate School of Education developed Artful Thinking to help students learn to think by looking at and exploring works of art. At the core of Artful Thinking is the Artful Thinking Palette (see Figure 1), which provides a meaningful visual reference for the kinds of thinking that can take place in the classroom.

Each of the six colors on the palette represents thinking dispositions. In practice, different types of thinking are often combined or overlap with one another. So, just as an artist working with paint can create limitless hues and shades from just a few different colors, students move naturally among various thinking dispositions, blending or overlapping the various kinds of thinking.

In addition to the metaphor of the artist's palette, Artful Thinking uses simple thinking routines to help guide student thinking. These routines are flexible and can easily be used to strengthen student thinking about virtually any topic or subject, from a math problem to an historical document, from a poem to a work of art. (See Figure 2 for one example of a thinking routine, the “Think/Puzzle/Explore” routine.)

Thinking routines are like many other classroom routines with one significant difference. Instead of helping students know what to do, thinking routines are designed to provide students with a framework for thinking and to help them to develop stronger thinking skills. When used regularly, thinking routines help students master and internalize new thinking processes until they become second nature. Using works of art is a good place to begin practicing these thinking routines. Because most works of art are inherently provocative, they encourage viewers to make observations and ask questions, thereby engaging them in different types of thinking. For these reasons, exploring works of art can help broaden the “thinking repertoire” of students and alert them to situations in which using different thinking modes can expand their understanding of a subject or issue.

**PRESERVICE TEACHERS MEET ARTFUL THINKING**

In our first group meeting, I asked my preservice teachers to create a “mind map” of the kinds of thinking in which they engage when learning something new. This exercise not only helped jump-start their efforts to think about thinking, but also made their thinking processes visible for them, documenting the extent, and the limits, of their thinking.

I also asked my preservice teachers to bring in samples of student work so we could examine them closely for evidence of thinking. For example, in examining a student writing sample, one preservice teacher described it as “complex” because the sentences were in order and built upon each other. Another noted that the writing sample reflected a “love of writing” because the student referenced his next piece of writing (alluding to a sequel).

These reviews and discussions around student work represented a powerful teaching tool for my preservice teachers, because they could clearly see evidence of thinking in
the student work and see that their students were capable of deep thinking. The discussions also made my preservice teachers think about how best to further develop this thinking capacity in students.

We then looked at an art image using the Artful Thinking routine called “Think/Puzzle/Explore” (see Figure 2). Using this routine to explore the art image resulted in a lively discussion, with each observation generating more comments and questions. My preservice teachers were also struck by how many of their questions were open-ended and not easily answered. We talked about the importance that asking good questions can play in the thinking process.

FIG. 2
An Example of the Think/Puzzle/Explore Routine

THINK:
What do you think you know about this topic?

PUZZLE:
What questions or puzzles do you have?

EXPLORE:
What does the topic make you want to explore?

Finally, we discussed the kind of thinking that they wanted students to engage in and what they could do to strengthen student thinking. Here, the Artful Thinking Palette gave them a vocabulary when talking about thinking and helped them visually categorize the kinds of thinking they wanted in their classrooms. My preservice teachers could also clearly see how they could use the Artful Thinking Palette with their own students to facilitate conversations about thinking.

IMPACT OF ARTFUL THINKING

The preservice teachers met regularly with me for nine months and also were required to e-mail me one journal reflection each week that focused on a particular issue regarding their student teaching experiences. They also were required to write a final reflection about trends, themes, and growth during their teaching internship.

At the end of the year, I also asked my preservice teachers to create a new “mind map” of their thinking, to compare their two maps, and to write a reflection about the differences they observed.

My notes, the preservice teachers’ journals, and their before-and-after “mind maps” were rich with information on how my preservice teachers’ thinking changed over the course of the year. Here are some of the findings that emerged from this year-long research.

Few teachers think about thinking. The most striking finding of this research was that my preservice teachers agreed that, before we began meeting together, they had never thought about thinking or how they might create opportunities for their students to think. Further, while a few had incorporated art into lessons to reinforce a point or illustrate an event, none had thought about using art images to deepen student thinking and make connections to the “big ideas” in the curriculum.

Early in the year, all of my student teachers said they wanted students to be critical thinkers. But, they had not really thought about what that meant, nor had they thought about what the evidence of good thinking would look like. Writing in his journal toward the end of the year, one preservice teacher observed:

I think thinking and utilizing activities that promote thinking were very useful. Personally, I never thought about thinking. Forcing myself to do this has allowed me to think deeper and broaden my lessons to accommodate critical thinking rather than only content. These models also showed me how to tackle these topics in an organized fashion rather than jump into topics without directions.

One preservice teacher became conscious of how important the thinking process is when making art. She wrote:

I used to think about the role of thinking, but did not think about making this thinking visible. Now, I think it is important to provide students with vocabulary to talk about thinking. It is important to make thinking visible (literally) in the art room. I used to think constantly while working, but did not take specific notice about the types of thinking I was using.

Teaching to think requires a “big picture” perspective. Looking at art images using the Artful Thinking routines forced my preservice teachers to move beyond the details in an image and focus on the larger issues in the composition, such as patterns, relationships, and multiple perspectives. This approach helped my student teachers synthesize information, make sense of the details, and discover that art images have the potential to elicit multiple interpretations, to
resonate with personal associations, and to activate prior knowledge.

As such, my preservice teachers were better able to appreciate the importance of identifying big ideas in their respective disciplines and to better appreciate the unifying themes and connections that exist among all subjects.

This was made evident by the experience of one of my preservice teachers, who was creating a final exam on the Shakespeare play, Macbeth. Initially, she planned to ask students a number of true and false questions that focused on factual information from the play. Most of her students would never again encounter Shakespeare, so I asked her what she wanted her students to remember about the play. That led to a discussion about the “big ideas” in Macbeth and to a change in the approach that she took regarding the exam.

As she wrote in her reflection journal:

This Friday, I gave my students their Macbeth unit exam . . . . After speaking with you and reviewing my objectives and goals, I decided that the objective of the test would be for the students to show their ability to manipulate, control, and extrapolate on their already learned knowledge of Macbeth. The test results I have received are very promising.

Another preservice teacher reflected on the Artful Thinking approach and how it helped inform her art making. She wrote:

Thinking about these strategies in studio courses helps us develop a philosophical view about the world which can feed our imaginations to enhance our art-making process.

Teaching to think requires a focus on the student and looking for clues as to the current state of their thinking. My preservice teachers had never looked closely at student work for evidence of thinking or for clues as to whether their teaching approach was achieving the desired results. However, through our work together, my preservice teachers began to dissect student work and to glean important insights into student thinking.

For most, it was the first time they understood that student work could provide a window into their thinking and into their understanding of the subject matter. In fact, most said their primary focus had been on the content of what they wanted to teach and their agenda for each class. But our discussions shifted their focus to their students and what their students were thinking and understanding.

Additional Resources

Artful Thinking web site:
www.pz.harvard.edu/at/cc_intro_new.cfm


Visible Thinking web site:
www.pz.harvard.edu/vt/VisibleThinking_html_files/03_ThinkingRoutines/03a_ThinkingRoutines.html
One preservice history teacher wrote:

I used to think that I needed to give lectures to students, information to stimulate their thinking. Now I think that there are many opportunities to think before lectures, such as assessing prior knowledge and thought that can help them think about and apply what they already know to different situations.

Thinking about thinking provoked my preservice teachers to think about teaching differently and to be more student centered.

**Teachers must be both facilitators and mentors.** Teaching to think is different than merely teaching content and requires an inquiry approach. My preservice teachers quickly found that the Artful Thinking routines required them to be facilitators and mentors, rather than conveyors of information. The routines moved my preservice teachers toward more of an inquiry-based approach and helped them develop and practice their facilitation skills.

Furthermore, the inquiry-based approach led to even more questions and to uncovering additional perspectives that otherwise would have gone unnoticed. My preservice teachers commented on how the thinking routines encouraged these additional discussions and generated additional ideas.

A preservice biology teacher wrote:

I would like to see students explore opposite viewpoints more in the classroom when they are presented with information in the book. I want the students to question who came up with the theories that are presented and how they went about creating these theories. In science class, I can go about getting the students to think harder by questioning how students came to their conclusions. Forcing students to explain their reasoning will help me understand their thought process.

A preservice history teacher wrote:

I used to think as a teacher I should only be a facilitator to higher-level thinking. Now I think that students look for affirmation when they take a chance in higher-order thinking. I need to be a facilitator and a mentor.

**Good thinking can be cultivated.** Finally, good thinking can be cultivated and guided, but teachers need time to foster thinking in their students, and students need time and opportunities to practice their thinking. Thinking routines can play an essential role in helping students deepen and strengthen their thinking and can gradually lead to the internalization of effective thinking dispositions.

For example, one preservice teacher wrote:

It is difficult to get students to automatically put on a "thinking cap," so to speak, and to start thinking deeply. While there are many routines, it's hard to get them to actively do it themselves. When they come across other situations, will they be able to sit down and think of a thinking routine? So the challenge is to get them to do it automatically.

Another preservice teacher observed:

The focus on deep thinking helped me to think more about what deep thinking is (its definition) and how it can be encouraged and brought out in my students in the classroom.

The preservice biology teacher wrote:

I used to think that students were automatically critical of information that they received and that I would try to convey science in an interesting manner and that it would be relatively simple. Now, I think that students are critical thinkers, but they have to be taught certain critical thinking techniques that will enhance their learning experience.

**MEANINGFUL CONNECTIONS**

I was anxious to see how preservice non-art teachers and art educators would react to the Artful Thinking Program and discussions around the topic of thinking. My preservice teachers seemed energized by the discussions and eager to have opportunities to think about thinking and the big ideas facing them as new teachers. So much of their mental energy is focused on organizational and housekeeping tasks, which are essential issues to new teachers and not to be minimized. Yet, our discussions were welcomed and sparked the preservice teachers to think in new ways.

Indeed, by focusing on the big ideas in their teaching, the preservice teachers were able to make meaningful connections with their students. They focused on student thinking and designed lessons that resonated with student interests and prior knowledge.

Thinking about thinking provoked my preservice teachers to think about teaching differently and to be more student centered.

How to promote creativity in the classroom

December 17, 2015 9:08pm AEDT

Author

Dan Davies
Dean of the Cardiff School of Education,
Cardiff Metropolitan University

Teach children to think creatively to solve tomorrow’s problems. Air Images/www.shutterstock.com

Creativity is a big deal in the 21st century classroom. Many countries include it as a core aim for their students in national curricula and even countries such as Singapore that come top of world education league tables are recognising the need for more of it in their schools.

This surge of interest in creativity among teachers, school leaders, academics and governments is partly driven by a growing belief that a fast-paced global economy requires workers with the flexibility of mind to adapt to constant change rather than follow a traditional career path.
How to promote creativity in the classroom

We live in a world where increasingly complex problems require creative solutions and where individuals’ lives can be enhanced by the greater sense of agency that comes with having opportunities to explore their own creativity.

Yet, surprisingly few teachers describe themselves as creative. This is perhaps because they have a performance-related, arts-based model of creativity in their minds, such as playing a musical instrument, painting a picture, acting a part in a play, writing a unique song, poem or story. This is in contrast to a broader definition of creativity as the ability to make connections between two previously unrelated ideas or contexts – what has been called “bisociation” by the Hungarian-British writer Arthur Koestler.

In 2013, I led a team undertaking a systematic review of Creative Learning Environments in Education for the Scottish Government. Looking at a number of studies, we found that in order to promote creativity among their pupils, teachers need to unpick their preconceptions about what it means to be creative as part of the professional learning process.

Let teachers be creative

They need to be given permission to innovate and improvise by school leaders, which is risky in a school culture structured around high-stakes testing. Once given this permission and support, teachers can develop creative learning environments for their students. This comprises both the physical environment of the classroom and a teaching environment with the following characteristics:

- students are given some control over their learning
- there is a balance between structure and freedom
- teachers are “playful”
- time is used flexibly
- relationships between teachers and learners include high expectations, mutual respect, modelling of creative attitudes, flexibility and dialogue
- students work collaboratively and assess each other

While each of these characteristics on its own might seem like a description of good teaching, it is their combination which creates the environment to promote creativity.

Two examples I uncovered during my research can help illustrate this. One teacher I observed in Somerset surprised his class by setting up a series of activities on their tables while they were out at break to introduce the topic of “gases”. These consisted of a candle burning, a series of plastic cups containing different numbers of marbles, and pairs of inflated and deflated balls.

The teacher gave no vocal instruction, but there were question cards with the activities, for example:

Watch the candle as it burns, what do you notice? Look at how the marbles are arranged, shake them, what is happening? Squeeze the two rugby balls, what can you say?

https://theconversation.com/how-to-promote-creativity-in-the-classroom-51838

11/12/2017
Initially bemused, groups of pupils soon began interacting with the exhibits and discussing their ideas. This unexpected start to the lesson — out of the normal routine — together with an invitation to look at everyday phenomena differently, provided the “hook” needed to engage children’s enthusiasm in a new scientific topic.

**Abstract concepts, made fun**

Another science co-ordinator at a South Gloucestershire primary school used stop-frame animation with plasticine models (like the Wallace and Gromit films) to help children understand forces in real-life situations. Working in groups of two or three, the children were asked to tell a story with their short animations that would involve everyday examples of forces in use.

One group of three girls shot a simple story of two boys having a fight “pushing each other over” and a dog jumping on top of them. They then annotated the resulting short movie on the computer with labels such as “push”, “pull”, “gravity” or “air resistance”. One child commented:

> You can be more creative when you do animation, because you can design what you’re going to do, and you get to think things through, like what forces you’re going to use and how the forces work.

Not only did this experience help reinforce children’s understanding of the tricky and abstract conceptual area of forces, it also enabled them to exercise choice, make links with other areas of the curriculum and engage in critical reflection as they viewed the results of their work.
Examples such as these demonstrate how teachers' own creativity and willingness to take risks can promote creativity in the way their students are learning. Such teaching for creativity is no laissez-faire, easy option – it requires careful preparation. As Thomas Edison said of genius, it’s “1% inspiration, 99% percent perspiration”.

Facts matter. Your tax-deductible donation helps deliver fact-based journalism.

Make a donation
Section 6:

Teaching and Pedagogy
The Role of the Teacher in the Thinking Curriculum

Introduction

*What has been the traditional role of the teacher?*

Traditionally, there has been a disconnection between education and thinking (Golding, 2006b). That is, people used to think that a student’s intelligence was a constraint on the level of education they received, but that relationship was not reciprocated, meaning education didn’t have an influence over a student’s intelligence (Golding, 2006b). This line of thinking resulted in a passive role for the teacher, “… all a teacher could do was make sure they set work that students had the ability to complete, and perhaps stream the more intelligent students into difficult, academic subjects” (Golding, 2006b, p.3)

*What is the new role of a teacher with relevance to the thinking curriculum?*

The core belief of a thinking curriculum is to develop good thinkers (Golding, 2006b). Good thinkers can be defined as not only having good thinking skills, but also having the disposition to use those skills appropriately (Golding, 2006b). Pietzner (2004) suggests that the role of the teacher in a thinking curriculum is to educate students to become more effectual thinkers and use their intrinsic cognitive resources, to better effect. He goes on to say that the teacher needs to encourage creativity, problem solving strategies and a philosophical approach to questioning to help students achieve a depth of inquiry. But how does a teacher go about achieving such noble goals?

The concept of a thinking curriculum is heavily influenced by Vygotsky. Vygotsky (1978) believed that the most important influence on a child’s cognitive development was the people within the child’s world, especially adults. The special role of the teacher was to scaffold students’ thinking. That is, to push students to the limit of their knowledge, to rectify the imbalance between new information and existing information, which he referred to as the zone of proximal development (as
cited in Adey & Shayer, 2002). Vygotsky suggested this be achieved through the interaction of three key components. Firstly, adults mediate the way a child interprets questions or problems, reframing them in a way that the child is capable of understanding. Secondly, they model effective problem solving strategies, leading by example. Thirdly, by utilising a language which focuses on the learning process (as cited in Alty & Pout, 2006). Thus, it can be seen that the role of teacher in a thinking curriculum is truly multifaceted. The best way to summarise the teachers’ involvement in a thinking curriculum is to categorise their contributions into several roles. These roles include; Teacher as model, teacher as facilitator, teacher as mediator, teacher as motivator and teacher as commentator.

Just as important as the teachers’ new roles, are difficulties teachers may have with taking on these roles. The traditional notion of a teacher is often an integral part of identity for educators. With that identify come certain ideologies. The thinking curriculum questions and often disregards such ideologies, which teachers may find quite confronting (Danielewicz, 2001).

**Teacher as a Model**

Sometimes, as teachers, we can forget how much we influence our students (Hughes, 2004). Students often model their behaviour on what they observe their teacher doing, thus it is our responsibility to model good behaviours. The idea of students learning through imitations has been further reinforced by the discovery of mirror neurons (Thompson & Collingwood, 2006). Mirror neurons enable a person to learn a skill just by observing another person do it. This is achieved by the learner’s neurons firing in the exact same areas as the ‘doer’. Empirical evidence suggests that mirror neurons function just as effectively for cognitive skills as they do for motor skills (Iacoboni, 2006).

One way to use this fact to enhance the thinking curriculum is to have the teacher model the practices of an effective learner (Claxton & Carr, 2004). Teachers can exhibit modelling by asking
questions and answering questions asked of them, admitting when they make mistakes and needing time to think. Additionally, it is beneficial to make their thinking observable to students by thinking aloud or explaining the process that led them to make particular decisions (Golding, 2004). To make this modelling explicit the teacher might use phrases like” perhaps”, “if we look it another way” “maybe I’m wrong but I think…” (Alty & Pout, 2006)

Another way to model is for the teacher to offer up a piece of their own work for analysis with the students. Sharing work brings the class to another level of understanding as they work together as a learning community (Costa & Kallick, 2000).

More importantly, through modelling the teacher gives the student a language that they can discuss their learning with (Claxton & Carr, 2004).

*Why might a teacher find this difficult?*

Often when we are modelling thinking skills, we must first start with an item that is causing cognitive conflict. This translates as something that doesn’t necessarily make sense or fit in with our current understanding of that phenomena (Adey & Shayer, 2002). By acknowledging that something is causing us cognitive conflict we must admit that we don’t know something. Despite teachers being regular human beings, they find it very difficult to admit that they don’t know everything, especially to students. Even the teacher who is most committed to a thinking curriculum, myself included, must fight with their inner id to overcome this need to be right in order to model how intelligent thinkers go about curing their cognitive conflict. Not surprisingly, students actually respond better to a teacher who can admit their fallibility, as it provides a platform for critical and constructive dialogue (Jones, 2001). Even though it is considered human to defend our biases, beliefs, actions and knowledge, it is also human to overcome the protective instinct by tapping into the courage to learn to change (Costa & Kallick, 2001).

*Teacher as a Commentator*
The purpose of a teacher acting as a commentator is two fold. Firstly, it gives teachers a chance to make explicit that they and the students’ are working within a thinking curriculum. They stipulate the importance of thinking skills, in the context of education and the ‘real-world’. By doing this, they set the scene; define the purpose of being there to students in terms that draw their attention to the development of learning as a valued goal (Claxton & Carr, 2004)

Secondly, commentary allows the teacher to highlight when thinking skills are being used, both correctly and incorrectly, and the consequences of such behaviour. This results in powerfully scaffolding students’ attention towards characteristics of their own performance. Teacher feedback can serve to identify and construct both achievement and improvement, but it is through the mutual construction of attainment and improvement that students can move from passive recipient to dynamic participant in the course of discussing and making choices about their learning (Alty & Pout, 2006).

*Why might a teacher find this difficult?*

If the teacher is not aware of thinking skills or incompetent in their correct uses, how will they be able to identify them in their students? The only answer here is professional development and training. The idea being that if teachers can be instructed to be effective thinkers then that style of thinking will be shared with their students (Golding, 2006b).

**Teacher as a Facilitator “the art of teaching is the art of assisting discovery” – Mark Van Dotten**

Atkin (1993) recommended that the role of the teacher requires a shift from trainer to facilitator. As a facilitator, “…the teacher acts as an intermediary between the learner and the learning” (Costa & Kallick, 2000). This provides opportunities for children to take responsibility for their own learning (Knight & Storey, 2004). Students need help in moving into a consistent informal research
mentality, where the solution to tasks lies not in knowing the precise answer, but in knowing how to find out the answer (Saddington, 2004).

A teacher could achieve this through the use of coherent neutral questions to get students to chase a line of thought, rather than the more traditional approach of evaluating answers that will shut down their thinking or worse still provide answers (Golding, 2006a). Alternatively, they may provide scaffolds for reflection, as this is vital to the process of meaning making for students (Costa & Kallick, 2000). Furthermore, it is important that teachers organise instruction in such a way as to facilitate the production of knowledge rather than consumption of knowledge (Costa & Kallick, 2000). The ideal being the teacher providing feedback that focuses on ways of thinking, rather than the content (Alty & Pout, 2006).

_Why might a teacher find this difficult?_

Not so long ago teachers were deemed the fountain of knowledge, the academic expert, the sage on the stage (McWilliams, 2009). What a grandiose image? It’s not difficult to see why a teacher wouldn’t want to give that up. Yet, this is exactly what a thinking curriculum aims to do. It is much more advantageous to give a student the skills and cultivate the disposition in them that will enable them to create their own knowledge (Golding, 2006b) or at the very least learn how to source knowledge for themselves.

_Teacher as a mediator_

The teacher acts as a mediator for several factors:

1. _Climate_

   It is vital that the thinking curriculum teacher mediates a classroom culture that is safe, supportive and provides the opportunity to engage in thoughtful discussion and reflection (Golding, 2006b). A safe climate enhances complex thinking by offering challenge, without threat. Threat is a well known inhibitor of the development of flexible thinking (Alty &
Pout, 2006). It is thus imperative that the teacher create and maintain a milieu of relaxed alertness, involving low threat and high cognitive challenge. That being said, it is of equal importance that the teacher not overstimulate or confuse the students (Alty & Pout, 2006). Additionally, classroom management strategies should foster reflective thinking, allowing the student to predict the consequences of their actions (Golding, 2006b).

2. **Group Learning**

According to Vygotsky social instruction is an integral part of learning (1978, as cited in Adey & Shayer, 2002). As such, it is important that teachers mediate collaborative situations so that students effectively share in and shape each others ideas. This can be achieved through the careful engineering of students’ construction of experience through well managed group work. The teacher’s role as mediator of the students’ interaction with each other and with the material is best achieved through the careful and strategic use of questions. The teacher uses questions to focus the students’ attention (Alty & Pout, 2006).

Furthermore, in a class of 25 or more students there are bound to be conflict of ideas. It is necessary that the teacher mediate such conflicts. In doing so, s/he should exploit this learning opportunity by demonstrating how to respectfully disagree with a peer, use evidence to support your argument, consider things from an alternative point of view and to use persuasion. It is through such events that students see the benefit of collaboration in their learning (Alty & Pout, 2006)

3. **Individual Learning**

The teacher must mediate the learning of individuals as well as those of the group. To have all students in flow means catering to different levels of challenge. All students should be working ahead of ability, in a cognitively demanding fashion, yet they should not feel that
success is beyond their reach (Alty & Pout, 2006). This can be achieved by the teacher ‘giving back’ problems to their students rather than giving them easy steps and recipes to solving the problem.

Why might a teacher find this difficult?

People like to be needed, and teachers are no different. Yet, a thinking curriculum promotes students independence. If a true thinking climate has been attained, the teacher’s voice in class discussions is now minimal (Alty & Pout, 2006). The teacher’s role has changed from central to hardly noticeable. This can lead to feelings of insignificance, in that the teacher feels they are not contributing to their students’ learning. However, it is important that teachers understand that if their role has been reduced to a minimum in a discussion, they have helped their students achieve what many adults cannot do, to communicate effectively and in a respectful manner. Furthermore, you will know your students well enough to know when to push them that bit further or even pull them back.

Additionally, an area that may require refinement for the teacher is their questioning technique, as questions are vital to the mediator role of the teacher. Areas of focus should include developing a range of sophisticated questioning techniques including open and closed questions, reflective questions, questions that provoke or clarify and a sense of timing to know when to use which type of question (Alty & Pout, 2006).

Teacher as a Motivator

To Learn

To achieve a deeper understanding of knowledge, students must be motivated to do so. However, there is often a fine line between motivating and demotivating a student (Harris & Bell, 1990).
One way to motivate student is by having high expectations of them. Once again we refer to the work of Vygotsky (1978 as cited in Alty & Pout, 2006) and his theory of having students working just above their cognitive ability (zone of proximal development).

How does a teacher maintain high expectations without making them too high? The easiest way to achieve motivation, as well as challenge, is to engage the students. Engagement involves authentic learning in which the students can identify with the content (Mimbs, 2005). Gabler and Schroader (2003) wrote “authentic learning can motivate those students who are bored, disinterested, or lacking necessary skills- the very students you might think could never do this” (p.202).

Furthermore, they suggest the use of student directed projects that are “driven by challenging, intriguing guiding questions… allow for a degree of student choice…build on the opportunities for students to share what they have learnt with classmates.” (p.418).

To Take Risks

As students begin to construct new knowledge, opinions or insights there is an element of risk for them. One way a teacher can motivate students to take risks with their learning is to encourage students to accommodate failure and to see it as a positive learning strategy so that “knowing what you don’t yet know” is appreciated (Alty & Pout, 2006).

Why might a teacher find this difficult?

Some teachers confuse engaging students as being entertainers. For some teachers, this may indeed be the case. However, entertainment is not a vital component of engagement. Engaging students means discussing components of the curriculum that are relevant to them or making the content relevant to the student. In fact, by engaging the students in content you achieve the opposite effect. Instead of the teacher having the “starring-role” the students become the central figures of the class. Teachers may also be burdened by guilt about letting a student take a risk that results in “failure”. It is important for both the student and the teacher that failure is just another form of learning and
risks need to be taken if any gains are going to be had. Learning how to handle failure is a great
skill to have as we hardly ever go through life getting everything right all the time.
Additionally, It is not easy for a teacher to make the transition from a position of making sure
students are ‘comfortable’ and motivated to one where s/he deliberately pushes students into their
discomfort zone while still maintaining motivation. (Alty & Pout, 2006).

Additional Challenges for the Thinking Curriculum Teacher

The Attitude of the School

The role of the teacher in a thinking curriculum can be made even more difficult when there is not a
culture of thinking in the school. This results in mixed teaching methods. Thus some teachers are
still teaching “traditionally”. If you’ve done your job well as a thinking curriculum teacher the
students will make comparisons between these two styles of teaching. You may even be labelled the
“bad” guy. “Why can’t you just tell us the answer like Mr. Smith does?” or “We’d never be able to
talk this much in Mrs. Brown’s class.” or “You don’t teach us anything?”. This kind of comparison
and questioning should be encouraged, as it will help the students to identify the benefits of why
you teach the way you do.

The Overcrowded Curriculum

Many teachers believe that while thinking skills are important, they are not as important as other
educational goals. Facts and figures, skills and knowledge are what students really needed to pass
exams and pursue their careers. In an overcrowded curriculum, where time is limited, teachers’ tend
to abandon thinking, and merely pay it lip service, in favouring of covering content (Golding
2006b). After all, it is the content that student and teacher performance is based on.
Conclusion

None of these role are mutually exclusive, there is often overlap between them. For example motivating the students to learn also involves a certain degree of modelling, that is the teacher must model valuing thinking to motivate the students to display good thinking. Even if a teacher does take on all these roles they will not achieve success unless part of their core value system is that students can learn to think more effectively. The important thing for teachers is to understand what they’re doing and why they’re doing it, because unless they are truly onboard thinking skills will continue to be marginalised within the curriculum. Teachers need to create new goals for themselves in how they value themselves and how they identify a job well done in a thinking curriculum. Teaching is all the more rewarding for a thinking curriculum; we just need to realign our values (Harris & Bell, 1990).

References


UNIVERSITY OF MELBOURNE
MELBOURNE GRADUATE SCHOOL OF EDUCATION
DEAN’S LECTURE
27TH SEPTEMBER 2011

LET’S GET SERIOUS ABOUT TEACHER QUALITY:
THE NEED FOR A NEW CAREER ARCHITECTURE FOR AUSTRALIA’S TEACHERS

Professor Stephen Dinham OAM

‘The quality of teaching is the main driver of successful student learning outcomes.
Australia’s teaching profession and its schools constitute an infrastructure that is critical to its survival in an increasingly global economy.
Every student deserves teachers who are suited to teaching, well trained and qualified, highly skilled, caring and committed to moving forward the learning of their students.’

(Dinham, Ingvarson & Kleinhenz, 2008)

The importance of the teacher and the school

Until the mid-1960s the conventional wisdom was that schools made almost no difference to student achievement, which was largely pre-determined by socio-economic status, family circumstances and innate ability (‘Coleman Report’, 1966). However, research has powerfully refuted that view. We now know that teachers, teaching and schools can make a significant difference to student learning and development. As a result, there has been a major international emphasis on improving the quality of teachers and teaching, particularly since the 1980s. ii

‘... the most important factor affecting student learning is the teacher. ... The immediate and clear implication of this finding is that seemingly more can be done to improve education by improving the effectiveness of teachers than by any other single factor’. iii

We now know how teacher expertise develops, what good teaching looks like and what good teachers know and do. However we also know that teacher quality varies widely within schools and across the nation.

‘Thus, the major challenge in improving teaching lies not so much in identifying and describing quality teaching, but in developing structures and
approaches that ensure widespread use of successful teaching practices: to make best practice, common practice.\textsuperscript{iv}

Whilst socio-economic status (SES) and family background can each exert moderate to large influences on student learning and development\textsuperscript{v}, these are not ‘life sentences’, as thought previously.

‘... school improvement by itself has potential to make an enormous difference in the lives of children even if broader social change is slow in coming. The children who depend most on good schooling for academic growth are the least likely to receive it. If school improvement begins early in life and if sustained, the most disadvantaged children stand to benefit most. This reasoning suggests that increasing the amount and the quality of schooling to which these children have access would reduce inequality in academic achievement.’\textsuperscript{vi}

What’s wrong with the current system for attracting, preparing, developing, recognising and rewarding teachers?

The first point to note is that there is no ‘current system’, but rather a series of largely disconnected, inconsistent, ineffective, incompatible teacher salary-career structures across the Australian educational landscape.\textsuperscript{1}

Previous attempts to drive improvement in teacher quality and to attract, retain, recognise and reward accomplished teachers have largely failed. One of the key reasons is that such schemes – of which there have been many in Australia - have never been mainstreamed to form a common, effective salary and career structure.\textsuperscript{vii}

‘Present arrangements in teaching do not encourage, reward or indeed require advanced professional learning. ... It is clear that there is a broad consensus that action is needed to radically strengthen procedures for recognising and rewarding teachers who reach high teaching standards.’\textsuperscript{viii}

Extensive international and national experience has shown that seemingly simplistic measures such as paying teachers on ‘merit’ or by ‘results’ are doomed to fail. Bonus schemes for teachers don't work in driving improvement in the quality of teaching. Rewarding and punishing schools on the basis of ‘results’ makes little sense when we know that teacher quality varies more within than between schools.

Whilst commencing salaries for teachers are comparable to those in similar professions, present lock-step annual incremental salary and career structures for teachers are 19th-century industrial artefacts that see teachers’ salaries peak too soon and at too low a level.

\textsuperscript{1} In saying this, I am not advocating a single national industrial award for teachers - but a common, recognisable and accepted framework for career long development, recognition and reward.
At present, more than three-quarters of Australia’s teachers are at the top of such salary scales where they earn less than 1.5 times the salary of a beginning teacher. This difference is too small and smaller than that in comparable countries where the differentiation is typically of the order of 1.75 to 2.25 or even higher. ix

Whilst too high a proportion of beginning teachers resign in their first three years – up to 25% - there is also a hidden resignation spike associated with teachers reaching the top of such salary scales after 8-10 years of teaching, a time at which salaries are rising steeply for the most able practitioners in other professions. x

David Berliner has suggested that moving from novice status to achieving competence as a teacher takes around two to three years. The development of a high level of skill, however, takes five to seven years and a great deal of work. xi The professional standards frameworks developed and being introduced across Australia by AITSL reflect this reality. xii

Other ‘one shot’ strategies to address teacher quality (e.g., higher entry scores, pre-service exit testing) won’t work when used in isolation. We need to address teacher quality at every step of the teacher career progression from entry onwards, i.e., at every key point of leverage. xiii

Caution is needed however. Rather than simply assessing teachers and making judgements on their capabilities at various career points in the hope that this will somehow lead to improvement, we need to use the concept of assessment for learning and development, something now widely advocated for student learning and equally applicable in this context.

The role of professional teaching standards in a new teacher career-salary structure

The solution to the complex situation outlined to this point lies in combining two present agendas; the introduction of national teaching standards, and the call for ‘bonus’, ‘merit’ or ‘performance’ pay to reward the most able teachers. Together these can systematically inform, drive and reward improvement in teacher effectiveness.

Professional teaching standards are now seen as crucial in promoting quality teaching and much work has been done in Australia over the past 20 years by professional associations, employers and jurisdictions, work that has produced a vast array of standards, frameworks and approaches designed to articulate, engender and in some cases, recognise and reward quality teaching. Although many of these standards have never actually been operationalised, this has been important work which has not been wasted. The involvement of the profession in thinking about and articulating what teachers need to know, do and understand has been an important precursor to what we have now and what we have to do now.

For the first time, Australia has a set of national teaching standards for teachers at four levels – Graduate, Proficient (both mandatory) and Highly Accomplished and Lead (both
optional). The introduction and adoption of these standards - and accompanying standards for teacher education course development and accreditation - provides an ideal opportunity and vehicle to move from the present ramshackle approach to teachers’ career structures to a nationally consistent model of professional learning, recognition and reward fit for a profession.²

In a report in 2008 for the Business Council of Australia (BCA)⁶ prior to the release of the current standards, Dinham, Ingvarson and Kleinhennz suggested that equilibrium in a national certification system for Australia’s teachers would see around 30% of teachers at the Highly Accomplished level and 10% at the Lead[ing] (Teacher) level. The remaining 60% of teachers would be seeking or have gained certification at the mandatory Proficient [registered] level (see Figure 1 below).³ In our report we suggested that the top of the Proficient salary scale should be around twice the salary of a beginning teacher. Those who achieved Highly Accomplished teacher status would have access to a salary scale that would enable them to learn up to 2.5 times the salary of a beginning teacher whilst those who achieved Lead teacher status could earn more. Teachers would not need to reach the top of salary scales in their particular level before seeking higher level certification.

We recognised in our report for the BCA that the issue of requiring a certain level of certification for a particular position was a decision best left to employers, given the diversity of Australian schools. It was our hope that eventually all teacher industrial agreements would incorporate a salary-career structure which was consistent with the proposed framework outlined below in Figure 1.

We now have national standards, but as yet we don’t have the framework. Whilst the development, endorsement and acceptance of the National Professional Standards for Teachers is a great achievement, the standards represent a beginning and a means, rather than an end. It is what happens next that is important.

² Universities have had such a broadly consistent and (inter)nationally recognised model for decades, with five levels (Tutor A-Professor E) and around 25 salary steps, along with systems of allowances and rewards for higher duties and positions of responsibility. A Professor typically earns 2.75 times the salary of a first step Tutor, a ratio consistent with the differentiation between a first year teacher and Lead(ing) Teacher in Figure 1.

³ This is indicative. We did not suggest the use of quotas for the higher levels.
Some key issues and areas for caution

However, there are a number of key issues that need to be addressed to achieve a true standards-based career structure for Australia’s teachers. Firstly, there is a need for effective performance standards that have authenticity, utility and add value.\textsuperscript{xvi} Performance indicators have yet to be fully developed. Secondly, how teachers are assessed and who conducts these processes are crucial questions. The profession must be meaningfully involved and feel ownership of the processes.

Professional learning opportunities, programs and resources must be developed and provided to accompany and support the standards and these need to be consistent with broader issues such as the emerging national curriculum being developed by the Australian Curriculum, Assessment and Reporting Authority (ACARA).\textsuperscript{xvii} Employers will need to ensure their processes and programs for appointment, induction, supervision, professional development, appraisal and promotion are consistent and supportive of assessment and accreditation of teachers at the various levels of the AITSL standards.
Portability, currency and maintenance of certification are also significant issues. Most important, however, is the question of whether certification of teachers at the voluntary *Highly Accomplished* and *Lead* levels will be rewarded financially through integration into existing and future salary and career structures. If this fails to occur, then the AITSL standards will gain little traction in driving improvement in the quality teaching. They will be decorative rather than generative.

It is essential that the assessment processes for certification are valid, reliable and credible within and across the profession. The assessment processes should be based upon the national standards and include observation of teaching and school visitations as well as ‘off-site’ assessment and validation across schools and systems. Assessors/observers must be fully trained and over time, should have been certified as *Highly Accomplished / Lead* teachers themselves. This involvement is essential in both recognising and utilising the expertise of the profession and involving and empowering it in the process.

The goal of advancing quality teaching is best achieved by having all Australian teachers certified at some point of the teacher career continuum from *Graduate*, to *Proficient*, to *Highly Accomplished* and to *Lead* teacher level (and possibly beyond in the case of aspiring and practising principals although the new National Standard for Principals is not intended to be used, at least directly, as a performance standard).

Whilst the lower *Graduate* and *Proficient* levels could be assessed by the existing teacher registration authorities within states and territories on a nationally consistent basis; because of the prestige and importance of the *Lead* teacher level, this should be truly national and conducted through or on behalf of AITSL. The *Highly Accomplished* level, being potentially far larger, could follow piloting of the *Lead* level procedures. Ideally, certification at the *Highly Accomplished* level would also be conducted nationally for reasons of consistency and credibility rather than being left to states and territories or individual schools and employers.

However, there are a number of additional issues. A poor ‘rubber stamp’ process with too many unsuitable teachers gaining certification at the higher levels could lead to a salary blow-out with little credibility or gain from the exercise.

Worthy teachers failing to be recognised will also undermine credibility, take-up and possibly morale. Experience tells us quotas won’t work either. The process must be fair and criterion-referenced.

The standards are intended to guide and drive professional learning and not just provide a frame to judge teachers. It will be necessary to measure teacher *performance* and *growth* against the standards at each level. Assessment for each level must be linked to evidence of student learning, the intended impact of teaching. Multiple measures of student learning will be required, including, class, year and school-based assessments, teacher portfolios.
(these need careful consideration\textsuperscript{xxi}), NAPLAN results (only where relevant as NAPLAN is a poor proxy for individual teacher quality) and use of other standardised measures and data sources. Key documents such as the \textit{Melbourne Declaration}\textsuperscript{xxii} and the \textit{Charter for the Australian Teaching Profession}\textsuperscript{xxiii} should be used as frames for the type of evidence of student learning and development that will be required. This goes far beyond a ‘scrapbook’ approach.

Joan Baratz-Snowden, a key figure in the field, has noted that:

\begin{quote}
‘The most valid assessment processes engage teachers in the activities of teaching - activities that require the display and use of teaching knowledge and skill and allow teachers the opportunity to explain and justify their actions.’\textsuperscript{\textit{xxiv}}
\end{quote}

As well as the standards for the various levels of certification, protocols and frameworks will need to be developed to guide teachers, their mentors, supervisors and assessors in the gathering and presentation of suitable evidence of performance and achievement across such areas such as student learning, contributions to the development of peers and to the school community, involvement in extra-curricular activities, student support, and leadership and service to the profession.

Research and evaluation need to be built into the processes of teacher development, assessment and certification from the start. As noted, professional learning will be essential to the application and outcomes of this overall process.

The scale of the undertaking is daunting given the number of teachers currently practising in Australia and the level of rigour required, yet it can and must be done.

\textbf{Time for action}

After many years of false starts the key pieces of the quality teaching and learning puzzle – including associated areas such as national testing, national accreditation of teacher education programs and the national curriculum - are coming together quite quickly. The new national standards for teachers (and the standard for principals\textsuperscript{\textit{xxv}) have the potential to drive, recognise and reward teachers’ professional learning and development from the graduate through to the ‘expert’ or \textit{Lead} level of teaching.

We are on the cusp of a new era of teacher professionalism and the AITSL standards and the application of these are integral to that development.

With the increased expectations placed upon schools, teachers and principals it is vital we have a clear understanding of the sorts of qualities and capabilities needed to meet these and to determine whether they are in fact being demonstrated and achieved.
However to reiterate, it is vital that the new national teaching standards and the associated measures for teacher professional learning, assessment and certification are integrated into salary and career structures across the profession.

This can’t be mandated centrally and will be a matter for negotiation within each industrial award. This will not be easy and it won’t be cost-neutral but if it doesn’t occur, we will be left with the current loosely connected, ramshackle, outdated, increasingly unattractive, and weak system for attracting, developing, retaining, recognising and rewarding teachers and we can expect to continue to slide down the ranks on international measures such as PISA. xxvi

Applications for teacher education courses in Victoria fell by 9% in 2010/2011. Whether this represents a temporary aberration or whether it is indicative of a longer-term trend is impossible to determine at the moment. However there is a danger that we have reached a tipping point in the attraction of teaching as a career. We will all be watching the level of demand for teacher education places later this year with great interest.

I have written previously that:

‘The biggest equity issue in Australian education today isn’t computers, new buildings or equipment. It’s each student having quality teachers and quality teaching in schools supported by effective leadership and professional learning in mutually respectful local community contexts.

Life isn’t fair, but good teaching and good schools are the best means we have of overcoming disadvantage and opening the doors of opportunity for young people.’ xxvii

If we cannot attract, prepare, support, develop, recognise, reward and retain quality teachers, we won’t be fulfilling our commitment to the young people of Australia as expressed in the Charter for the Australian Teaching Profession and the Melbourne Declaration and we will all be the poorer for it.

References


Dinham, S. (2008). 'Driving Improvement in the Quality of Australian Education: Points of Leverage', Australian College of Educators, Victorian Branch Oration, University of Melbourne, 15th August. [http://works.bepress.com/cgi/viewcontent.cgi?article=1079&context=stephen_dinham&sei-redir=1&search%2DDriving%2DImprovement%2Din%2Dthe%2DQuality%2Dof%2DAustralian%2DEducation%2DPoints%2Dof%2DLeverage%2D2](http://works.bepress.com/cgi/viewcontent.cgi?article=1079&context=stephen_dinham&sei-redir=1&search%2DDriving%2DImprovement%2Din%2Dthe%2DQuality%2Dof%2DAustralian%2DEducation%2DPoints%2Dof%2DLeverage%2D2)


[http://www.pisa.oecc.org/pages/0.2987_en_32252351_32235731_1_1_1_1.html](http://www.pisa.oecc.org/pages/0.2987_en_32252351_32235731_1_1_1_1.html)


sdinham@unimelb.edu.au