3 Constructivism in an Age of Non-Constructivist Assessments*

Daniel L. Schwartz *Stanford University*
Robb Lindgren *Stanford University*
Sarah Lewis *Stanford University*

Constructivism is a theory of knowledge growth and life-long development built on a philosophy of pragmatism (Dewey, 1916). In the context of formal education, it is frequently used as pedagogical label for sense-making activities including discovery, inquiry, exploration, and hands-on learning (Duffy & Jonassen, 1992). It is often set in opposition to shaping behavior, where external reinforcements regulate learning, as well as direct instruction, where students are told or shown what to do.

Constructivism *writ large* has fared relatively well in education. Lillard and Else-Quest (2006), for example, found that Montessori education leads to better academic and social skills. Russell, Hancock, and McCullough (2007) found that participation in undergraduate research experiences, regardless of specific mentoring styles, increased student interest in pursuing advanced degrees in science. However, constructivism *writ small*—constructivism applied to single lessons or instructional units—has not fared as well. Klahr and Nigam (2004), for example, demonstrated that explicitly telling young children the control of variables strategy led to improved learning compared to having the children simply conduct experiments without guidance. Similar findings have led some scholars to the conclusion that constructivist pedagogies are inconsistent with cognitive architecture because they withhold information that can be readily told or demonstrated (e.g., Kirschner, Sweller, & Clark, 2006). This conclusion cannot be completely warranted, given what we know, for example, about the generation effect (Slamecka & Graf, 1978). Given pairs of synonyms, people will remember a word better if they explicitly have to generate missing information, as in the case of FAST: R_P_D versus reading FAST: RAPID. Nevertheless, their analysis does lend itself to the question, "Wouldn't it be more efficient to simply tell students what they are supposed to do and know?"

Some of the discrepancy between the outcomes of constructivism writ large and constructivism writ small has to do with the nature of the assessments that are used to evaluate pedagogical effectiveness. Constructivist pedagogies writ

* The writing of this chapter and the research by Sears (2006) were supported by the National Science Foundation under Grant No. SLC-0354453. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

small are often evaluated through non-constructivist means. Students receive tests that measure how well they developed their efficiency at remembering facts, executing skills, and solving similar problems. These assessments present something of a mismatch to larger constructivist goals. Dewey (1916) stated, "the aim of education is to enable individuals to continue their education … the object and reward of learning is continued capacity for growth" (p. 117). Given this, a constructivist-tailored assessment should examine students' abilities and dispositions to construct new knowledge, not just execute old knowledge. This approach would be consistent with the assessments of constructivism writ large, which often indirectly measures students' ongoing abilities to construct knowledge, for example, by examining cumulative effects or by seeing whether students are more inclined to engage new content and new situations both in and out of school (Boaler, 2002).

This chapter shifts the application of constructivism away from instruction and places it instead in the realm of assessment. We begin by noting that constructivism is a sweeping theory so broad that it is difficult for it to dictate specific instructional design decisions. We are not alone in this observation. Mayer (2004) expresses faith in constructivism as a theory of knowledge attainment, but states that "a doctrine based approach to constructivism does not lead to fruitful educational practice" (p. 17). While this may be accurate, we argue that constructivism's breadth is quite valuable when used to assess educational outcomes.

Constructivism is a broad vision of learning; it is not just an approach to instruction. It enables us to consider students' abilities to create new knowledge when they are outside of instruction and we no longer have control over precise instructional variables. By shifting the focus to assessment, we can ask the question, "What experiences prepare students to construct knowledge in the future and in the wild?". This question is important because learning should not end once students leave the classroom and lose a teacher's direct guidance. By creating constructivist assessments, it will be possible to identify the elements of instruction—constructivist or otherwise—that facilitate the development of continued learning.

We justify this shift in focus by presenting three related lines of research. The first demonstrates the significance of constructivist assessments by showing that constructivist-inspired activities prepare students to construct knowledge from direct instruction later. The second line of work shows the value of constructivist assessments for detecting the special benefits of constructivist-inspired instruction. The third line of work demonstrates that targeting constructivist outcomes is compatible with promoting the efficiency outcomes favored by direct instruction, but that direct instruction may not always be compatible with constructivist outcomes. We conclude by tentatively working backward—given constructivist outcome measures, it is possible to start determining which elements of instruction lead to those outcomes.

Specificity in Instructional Theory

Different instructional techniques are suited to different instructional outcomes. If one's instructional goal is the development of sustained interest, then solo drill and practice will probably not fare very well. Instead, instruction that creates a social matrix of participation is likely to be more useful (Barron, 2004). In contrast, if the goal is to create efficiency in highly stable and repetitive contexts, then individual drill and practice may be very effective. One style of instruction does not fit all outcomes.

The relation between instruction and outcomes is mediated by the specific learning processes that are engaged. For instance, simply changing people's beliefs about an interaction can modify the learning processes. Okita, Bailenson and Schwartz (2007) had people discuss the mechanisms of fever with a graphical human character in virtual reality. By using the virtual-reality set up, the researchers were able to hold all information and interaction constant across participants and conditions. There were two conditions: participants were told the character was either controlled by a person or controlled by a computer. Even though there were no differences in the available information or interaction, people learned more about fever if they thought the character was controlled by a person. People had higher arousal when they thought they were in a social interaction, and arousal levels correlated positively with conceptual learning.

From a psychological perspective, there are many different internal mechanisms that regulate learning, and different instructional conditions can engage learning mechanisms differentially. People can learn by being told; they can learn by observing social models; they can learn through spatial navigation; they can learn through reinforcement; they can learn through deliberate practice; they can learn by exploration; and they can even learn implicitly without any intent or awareness they are learning at all. These pathways of learning engage different brain circuitry (e.g., Seger, Prabhakaran, Poldrack, & Gabrieli, 2000). They are not simply multiple instances of a single "learning module." Each learning process has benefits for particular types of content and outcomes. For example, implicit learning is thought to be important for language acquisition, and direct instruction can interfere with implicit processes (Reber, 1976). The instructional challenge involves deciding which combination of learning processes and environmental supports will yield which desired learning outcomes (Hmelo-Silver, Duncan, & Chinn, 2007). The goal of fitting instruction to outcomes was a major constituent of Gagne's seminal work on the conditions of learning and helped to create the field of instructional design (Gagne, 1985).

In our experience, constructivism tends to be too large and general a philosophy to be useful for the precise handling of the many specific ways and reasons that people learn. Constructivism is not at the right level for deriving specific instructional decisions. Sometimes hands-on learning is valuable and sometimes it is not—knowing the microscopic details of when it is valuable is difficult to derive from constructivism alone.

This is not to say that constructivism does not have an important role to play in the design of instruction. Instructional theories that are more specific tend to focus on one class of learning process and assume "all other things being equal." For example, cognitive theories of learning tend to be silent about design decisions involving motivation, but any instructor can testify to the importance of making instructional choices involving student engagement and interest. One benefit of constructivism for instruction is that it orients educators toward important values including overall student growth, interest, and agency. This way, when educators consider specific learning processes, they do not lose the bigger picture of education.

Constructivism as a Guide to Assessment

Although we believe that the broad concept of constructivism invites the wrong level of analysis for designing specific instructional moments, we do see constructivism as extremely valuable when applied to learning outcomes. Rather than taking constructivism as an instructional design theory, we suggest that the ideas of constructivism be applied to assessment. We ask the question "Does instruction prepare learners to construct knowledge once we no longer orchestrate specific instructional conditions to target specific learning mechanisms and outcomes?".

At the level of a single lesson, educators are able to offer guidance for specific instructional conditions. Once students leave the confines of formal instruction, however, teachers have no influence over specific instructional decisions and learning processes. In the wild, people need to construct knowledge using whatever resources, internal and external, are available. Because we cannot anticipate and decompose these future opportunities into specific learning mechanisms and situations, the granularity of constructivist theory now becomes appropriate.

A goal of much formal schooling is to provide students a foundation of knowledge on which they can build new knowledge once they leave school. For example, in interviews with school superintendents, we found a unanimous desire for their students to be prepared to learn and adapt so they can make choices on their own once they have left school (Schwartz, Bransford, & Sears, 2005). Except for very narrow training, people will need to continue learning new ideas and skills once they have left the confines of immediate instruction. People grow into their jobs, they change jobs, and the world changes around them.

As we consider constructivist learning outcomes, it is important to note that not all instruction should target the outcome of preparing students to construct new knowledge beyond immediate instruction. This is particularly true for highly stable domains where it is possible to cover nearly every possible combination of skills and performance conditions. For some domains, such as typing, the requisite skills and performance conditions are extremely stable. Keyboards and major keystroke combinations do not change, so there is little reason to prepare students to learn how to type. In this case, the instructional goal should be to make sure people develop good initial habits, so they can become increasingly efficient without having to undo their prior learning.

The typing example is informative when we think of assessment and potential mismatches with the goals of instruction. If we held typing instruction to constructivist outcome measures, for example by evaluating typing instruction based on whether students are prepared to learn the Dvorak keyboard, it would mis-measure the benefits of procedural training for stable environments.

Unfortunately, researchers in education frequently make an analogous mismatch when attempting to assess constructivist-inspired pedagogies by using non-constructivist assessments.

Most end-of-unit tests explicitly block students from constructing new knowledge during the test itself. These tests measure students' abilities at sequestered problem solving (SPS) rather than learning (Bransford & Schwartz, 1999). We say sequestered, because students are shielded from any resources that might help them learn during the test. SPS assessments are ideal when the goal is to determine student efficiency at retrieving and executing well-practiced knowledge and routines in familiar situations. For example, a good typing test would be an SPS test—how fast and accurately can people type. However, SPS assessments are not ideal when evaluating whether students have been prepared to construct knowledge based on what they have learned. During an SPS test, there are typically very slim opportunities and resources for constructing new knowledge.

A more appropriate test for constructivist outcomes is a preparation for future learning (PFL) assessment. In this type of assessment, students have an opportunity to learn during the test itself. Students who have been prepared to construct new knowledge in a domain will learn more during the assessment than those who have not been prepared to learn. PFL measures seem more in line with constructivist outcomes. We provide examples of PFL measures below, but a simple thought experiment can help signify their value. Imagine a firm wants to hire a financial analyst. Tom has just completed a 2-week course in Excel—his first exposure to spreadsheet software. Sig has not learned Excel, but he has previously taught himself to high levels of expertise with multiple spreadsheet packages over the past several years. The company decides whom to hire by using a paper and pencil test of basic Excel operations that just happen to have been covered in Tom's course. Tom would probably do better on this SPS test. However, we suspect Sig would be more likely to serve the company well in the long run. His deeper understanding of spreadsheet structure and capacity to learn independently will enable him to learn and adapt on the job, for example, when the company switches to a new software package or when the employees are asked to learn advanced features of Excel on their own.

When evaluating instruction it is important to use outcome measures that capture what we want students to achieve. Thus far, most high stakes and experimental assessments have used an SPS format. This format favors direct instruction and repetitive practice, because direct instruction's primary goal is to increase efficiency in well-specified tasks. The SPS format does not favor constructivist pedagogies, which ideally target constructivist outcomes. More importantly, the current lack of PFL assessments means that we cannot know whether constructivist-inspired pedagogies actually achieve constructivist outcomes any better than direct instruction. Simply put, our measures have been misguided. If we use constructivist assessments, like PFL, then we are in a better position to see which features and styles of instruction promote constructivist outcomes.

What Prepares People to Construct Knowledge from Direct Instruction?

One of the authors (DS) was a remote committee member for a graduate student who was enrolled at one of the premier universities in the United States. Standard tests no longer determined the student's fate. The author told the student at his dissertation proposal that he should explore his theories and conduct loose empirical research until the important structures and themes started to reveal themselves. Only then should he commit his time to exhaustive experimentation and a focused review of the literature. The student's primary advisor had written an important paper favoring direct instruction as a way to improve inquiry skills. In a convivial manner, the advisor stated to the author and graduate student, "You are asking him to follow your theory of education, but your theory is not proven in clinical trials." The author responded wryly, "He should do it your way then. You can tell him exactly what to do, and I am sure he can copy it perfectly. Afterwards, we can see how well the student fares on the job market."

The point of this story is not that one of the committee members was right or wrong, but rather that both pieces of advice, taken to their extremes, are somewhat foolhardy. Over the course of time, people will learn in many ways. Sometimes, it is important to explore and develop one's own ideas. Sometimes, it is important to receive direct guidance. The question is not which method is right; the question is what combination of methods is best for a given outcome.

Direct instruction can be very effective, assuming that people have sufficient prior knowledge to construct new knowledge from what they are being told or shown. In many cases, they do not. For example, a challenge in pre-service teacher education courses is that the students do not have sufficient prior knowledge of teaching in the classroom. In their teacher preparation courses, it is hard for the students to see the significance of the theories or map them into their future pupils' behaviors. Moreover, they lack a repository of personal instances that round out the specific examples used during instruction. Teacher educators have to work very hard to include multiple ideal examples, cases, and videos to help the pre-service teachers understand the significance of theories and lessons presented in the class readings and lectures. In this respect, it is much easier to work with in-service teachers. They can bring to mind relevant cases and juxtapose their own classroom wisdom with the ideas presented by the professor.

If students do not have sufficient prior knowledge to readily make sense of direct instruction, what is the best way to develop it? To address this question empirically, it is important to think in terms of PFL assessments. The empirical question is what types of experiences best prepare students to construct knowledge from direct instruction. A PFL assessment can be used to determine how well students have been prepared to learn, for example from a lecture, by making learning from the lecture part of the assessment.

A relevant series of studies comes from teaching students in an introductory cognitive psychology course (Schwartz & Bransford, 1998). We describe one of the studies. In this study, the students learned two clusters of cognitive concepts over several lessons. To gather baseline data, all the students received the same instruction to learn the cluster of concepts relevant to encoding theory (what aspects of information people store in long-term memory). The second cluster of concepts was about schema theory (how people organize knowledge). For the second cluster, students completed different treatments so we could compare their learning across instructional treatments and against their own baselines.

The students learned about the encoding concepts first. They received a packet of simplified data sets and experimental descriptions from classic encoding studies. Their task was to analyze the data and make graphs of the most important patterns in 80 minutes. They were not told the point of the experiments they analyzed; it was up to them to decide what patterns were important enough to graph. We will call this activity "Analyze." Two days later, the students heard a 30-minute lecture on encoding. The lecture reviewed the results from the studies and the broader theories that explained them. We will call this activity "Lecture." Thus, for the encoding concepts, all the students received an Analyze+Lecture treatment.

A week later, the students were separated into three treatments for learning about the schema concepts. Two groups completed Analyze activities for new schema data sets. A third condition received a text passage on the schema concepts. The third condition's task was to write a summary of the passage. We will call this the "Summarize" activity. The chapter included descriptions of the studies and graphs of results, and the chapter developed the broader theories that explained the results. All three groups had 80 minutes to complete their activities. Two days later, one of the Analyze groups received the Lecture, and the other Analyze group was asked to continue the Analyze activity. The Summarize group received the Lecture. Thus, in this phase of the study there was an Analyze+Lecture treatment that was comparable to the way all the students had learned the encoding concepts. There were also the two new learning treatments, Summarize+Lecture and Analyze+Analyze. (We describe what happens when students receive an analog of a Lecture+Analyze treatment in a later section.)

The following week, the students received an assessment to see how much they had learned. They received a written description of a novel experiment. Their task was to predict the outcomes of the experiment. The experiment had eight possible predictions, four based on the encoding concept cluster and four based on the schema cluster. Figure 3.1 shows how many of the concepts appeared in their predictions. The students did better predicting around concepts that they had learned by analyzing the data and then hearing about in the lecture. Students who had summarized the chapter and then heard the lecture did not do very well. The difference was not due to the Summarize students overlooking concepts during the Summarize activity. Graphs and summaries produced by the students during the Analyze and Summarize activities indicated whether or not they had covered a concept prior to the lecture. When Summarize students noted concepts in their summaries, they only made predictions based on those concepts 23% of the time. In contrast, when the Analyze+Lecture students showed concept-relevant patterns in their graphs, they were likely to make predictions using those concepts 60% of the time.

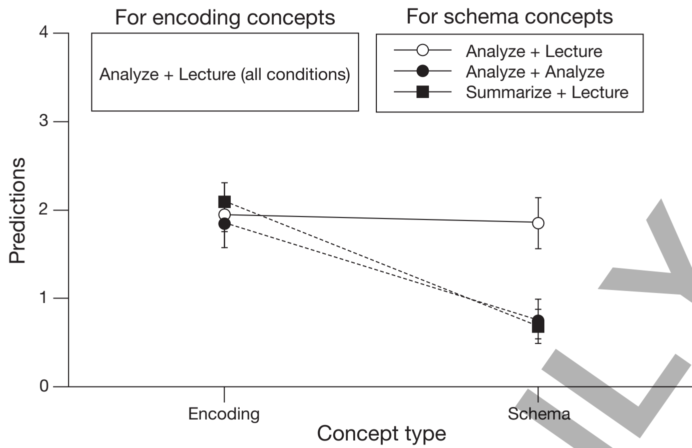

Figure 3.1 Students who explored and graphed data sets on memory experiments demonstrated superior abilities to predict outcomes of a novel experiment, but only if they had a chance to hear an organizing lecture afterwards (source: adapted from Schwartz & Bransford, 1998).

Importantly, students who analyzed the data, but never heard the lecture, also did not do very well. Although the Analyze+Analyze students had graphed the relevant patterns, they only made predictions based on the concepts they graphed 18% of the time. This latter result means that analyzing the data by itself did not lead to effective learning. Instead, the effect of analyzing the data was to prepare students to construct knowledge from the lecture.

From these results, a good way to prepare students for direct instruction is to give them targeted experiences with "exploratory" activities. It was the combination of exploration and telling that led to good learning outcomes. The exploratory activity or the verbal activities by themselves did not yield the desired learning. Monolithic theories that advocate all exploration or all direct instruction do not explain these results.

It is important to note, however, that the benefits of blending instruction only hold if the desired outcome is to have students transfer their knowledge to novel situations. Other studies in the series measured recognition memory. Students who heard the lectures without any analysis activity did very well (as did the students who did the analyses-plus-lecture activities). Students who wrote summaries, for example, were able to correctly affirm the encoding concept that people only remember the gist of a passage when they understand it. They just could not apply this declarative fact to the prediction task. If simple memory were the goal of instruction, then lectures may be fine for college students. However, the goal of this psychology course was for students to develop the ability to predict and interpret human behavior. In cases like this, SPS memory assessments are the wrong way to evaluate the quality of instruction or student learning.

Constructivist Activities Yield Constructivist Outcomes

It is not always possible to anticipate when a learning experience will arise. Instruction that targets constructivist outcomes should prepare students to recognize and take advantage of learning opportunities, even when there is no teacher available to say "learn this." In the preceding example, the sheer presence of the lecture format indicated to students that it was important to learn. A more stringent test of pedagogy would examine whether students are prepared to learn without overt prompts.

To examine what types of experiences prepare people to construct knowledge without overt prompting, Schwartz and Martin (2004) conducted a study on learning statistics. Unlike the psychology instruction study which looked at conceptual content, this study examined the acquisition of quantitative ideas and procedures. Multiple classes of ninth-grade students completed invention activities related to the topic of variance. They received data sets, and they had to invent a formula for computing a consumer "reliability index." For example, given data, they had to come up with a value that indicated which of two trampolines produced the most consistent bounce when a ball was dropped repeatedly from a fixed height. Few students invented a correct solution. However, the goal of the instruction was not for students to re-invent what took professional mathematicians many years to articulate. Rather, the goal was to prepare them to understand the formal solution created by mathematicians once it was presented in class. After students completed a series of invention activities, they were taught a formula for variance, and the students were given time to practice using it. Like the case of college students analyzing data on human memory, the invention activities prepared students to form a deep understanding from the content of the lecture. The ninth graders outperformed college students who had a semester of college statistics on a variety of measures ranging from their explicit long-term memory of the formulas to their ability to explain the structure of the formulas (e.g., why variance formulas divide by n). These findings complement the results of the analysis-plus-lecture activities in the preceding study.

The new question in the statistics study was whether students would be prepared to learn spontaneously without overt instruction. The way the experiment worked was that on the last day of several weeks of instruction the students worked in small groups of two to four to learn about standardized scores (e.g., grading on a curve). Half of the students completed a direct instruction treatment we will call "Tell and Copy." The students were told (and shown) a graphical technique for computing standardized scores. They then received a data set and had to answer a question using the technique to find standardized scores (e.g., who broke the world record by relatively more, the high jumper or the long jumper?). The teachers corrected student errors during this practice. The other half of the students completed an invention treatment we will call "Invent a Measure." They received the same data set and question, but they were told to invent their own solution to the problem. They did not receive any specific guidance on how to solve the problem, and they did not receive any feedback on their solutions. No students invented a workable answer, so this would at first seem to be a very inefficient form of instruction. However, by a subsequent PFL assessment (described next), the experience of trying to invent a solution revealed itself as very important. It prepared students to learn from a "hidden" solution later on.

The PFL assessment involved two items within a longer test that occurred about a week after instruction. One item, near the end of the test, was a difficult transfer problem. The target transfer problem did not look like any problems the students had studied. The surface features were novel, and the problem required an application of standardized scores not covered directly in the prior instruction. The other item was a worked example in the middle of the test. Students had to follow the worked example to solve a subsequent problem just below it. Embedded in the worked example was the procedure that would help solve the transfer problem. Nearly all of the students correctly followed the worked example. The question was whether they were prepared to learn from the worked example, such that they could apply it to the target transfer problem.

All the test packets included the target transfer problem, but only half of the tests in each condition included the worked example. This way, it was possible to determine whether students were solving the target transfer problem on the basis of the worked example. Figure 3.2 shows the results. Students who received the direct instruction on standardized scores were not prepared to learn from the worked example. The Tell and Copy students performed roughly the same on the post-test with or without the worked example in their tests. In contrast, students who had tried to invent a way to handle standardized scores were twice as

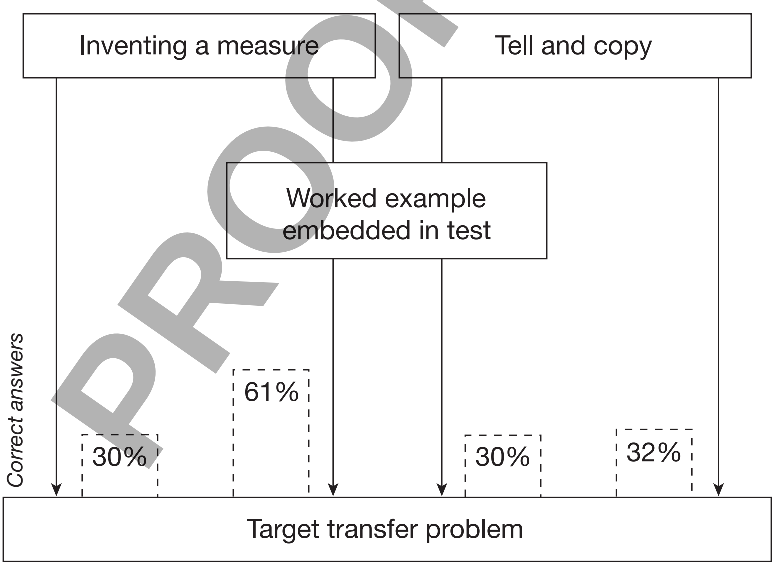

Figure 3.2 Students who tried to invent a solution to learn the concept of standardized scores were more prepared to learn spontaneously from worked example and transfer it to a subsequent problem than students who had been told and practiced a solution using standardized scores (source: adapted from Schwartz & Martin, 2005.)

likely to learn from the worked example and use it to solve the transfer problem. A subsequent study replicated the results with a new set of ninth graders who received the instruction from their regular classroom teachers (Schwartz & Martin, 2004).

The study provides three useful bits of information. The first is that constructivist-inspired activities can lead to constructivist outcomes better than direct instruction, at least in this instance. Students who had unsuccessfully tried to invent solutions to problems involving variance were more prepared to learn spontaneously without explicit support. The second bit of information is that the benefits of the invention activities would have been overlooked if the studies had not used PFL assessments. Students who tried to solve the transfer problem without the benefit of the worked example completed an SPS assessment because there were no resources for learning. By this SPS assessment there was little difference between direct instruction and invention. The third bit of information is that worked examples can create effective instruction, but only if students are prepared to construct useful knowledge from the examples. Students in the Tell and Copy condition were able to follow the worked example to solve the analogous problem below it, but they did not learn the significance of what the procedure could accomplish.

Issues of Efficiency and Variability in Learning

The efficiency of instruction has always been a central consideration in American pedagogy. Skinner (1986), for example, opposed discovery activities because they do not lead students to produce reinforceable behaviors as quickly as possible. More recently, Chandler and Sweller (1991) pointed to the inefficiency of having students waste time searching for information when it would be more cognitively parsimonious to tell or show them where to look. Even so, it would be a mistake to generalize that exploration is always an inefficient use of time and cognitive resources.

In the preceding studies, we demonstrated that there can be advantages to first letting students experience the complexities of a situation and then providing information that helps them understand expert techniques and concepts in light of their earlier successes, difficulties, and questions. Vollmeyer, Burns, and Holyoak (1996), in a more cognitive paradigm, showed the value of allowing exploration over immediate correction. Even the animal literature shows benefits of slow exploration over getting the right answer as quickly as possible. For example, Verbeek, De Goede, Drent, and Wiepkema (1999) created a set of feeding sites for a species of bird (titmice). Some birds took a slower, more exploratory approach to learning the feeding sites, and others learned the feeding sites more directly. The researchers then changed the feeding sites. The birds that had taken the more exploratory approach were more effective at learning the new feeding sites. In each of these cases, exploration permits the learners to induce more about the structure and variability of the learning space, which enables them to handle new problems in that space more effectively. Effective learning for variable settings is not just about knowledge of routines and concepts; it also demands knowledge about situations to which those routines and concepts apply. Perhaps it is possible to instruct people directly about the structure and variance of a situation, but it is not obvious that the learning processes associated with telling would be the same or as effective as exploration. Obviously they would not be equivalent for the birds.

When people are being instructed for highly stable conditions where efficiency is at a premium, such as typing, then exploring variability is less important. Isolated and stable practice is very useful. But, if there is a possibility that people will need to use their learning in new situations that instruction cannot fully anticipate, then "background" variability becomes important to include in some portion of the instruction. The seminal studies by Gick and Holyoak (1983) provide valuable evidence for this point, and their results can help to overcome the intuition that the best way to teach is to include only relevant information and exclude all "noise" from the instruction. In their studies, people learned how to solve story problems, and then later received a structurally similar story problem with different surface features. For example, the learning problem might have a medical cover story and the target problem might have a military cover story, but both scenarios depend on the same solution of dividing one's forces—radiation beams or troops—and having them converge simultaneously at the target. The question of interest was what conditions of initial learning would help people transfer the convergence solution to the target problem.

Gick and Holyoak explored multiple combinations of initial learning to see which would support transfer. All the combinations worked well for learning the basic solution, but they were not equally effective for supporting transfer. They found that the most effective treatment for achieving transfer was to have subjects read two examples with different cover stories plus an explanation. This combination was more effective than a single example; more effective than an explanation without an example; and, more effective than a single example plus an explanation. It was also more effective than using two examples with similar cover stories plus an explanation. The cover stories of the two different examples were incidental "noise" for the structure of the solution. Yet, by including two different cover stories, it helped the participants learn which features of the problems were relevant (the necessity of dividing a single large force) and which were irrelevant (the medical or military context). In this case, contextual heterogeneity was critical for helping students induce which aspects of the situations were relevant. Working with two similar examples did not help as much as working with two different examples. Figure does not exist without ground. If instruction removes all background variability for the sake of efficiency, students will not be prepared for new situations where they must discern on their own what is relevant and what is extraneous.

Efficient and Constructivist Outcomes Are Not Mutually Exclusive

Thus far, we have considered the issue of efficiency from the perspective of instruction—which combinations of instruction most efficiently help students learn. The same issues of efficiency also suffuse the assessment side of the instructional coin. Are outcomes of high efficiency in knowledge application incompatible with constructivist outcomes? There are merits to this question. One possibility is that people can over-rely on well-learned and efficient routines. They can miss opportunities for learning new ways of doing things that may be even more efficient. Luchins and Luchins' (1959) famous studies of set effect or *einstellung* (rigidity of behavior) demonstrated this problem. People learned a way to solve water jug problems, and they did not let go of this solution even when it became inefficient for new problems. The second possibility is that desired outcomes that emphasize future abilities to construct knowledge will come at the expense of learning efficient solutions, even if students are eventually shown and practice efficient solutions.

A recent study suggests that the former concern is more important than the latter. Premature efficiency is the enemy of subsequent new knowledge construction, whereas early innovation permits both efficiency and constructivist outcomes. Sears (2006) examined whether constructivist outcomes that emphasize abilities to learn are incompatible with outcomes of efficient knowledge application. The study involved college students learning the logic behind the chi-square statistic (i.e., expected versus observed results). Students received a sequence of carefully designed cases. For example, one case focused on determining whether dice were loaded according to the expected probabilities. Students had to invent a procedure for computing a value that indexed their chances of being loaded. Another case asked students to invent a procedure to index whether different groups exhibited different preferences for food, when the expected probabilities are unknown. Figure 3.3 provides this example.

One factor in the study was whether students worked alone or whether they worked in pairs. The second factor involved the sequence of instruction. In one treatment, students were told the relevant procedure for each case, they practiced on that case, plus one more. (This condition would be the analog of a Lecture+Analyze condition had we included it in the studies on learning cognitive psychology.) In the other treatment, students tried to invent a way to handle each case, and then they were told the relevant procedure and practiced on the additional case. This instructional manipulation was very subtle—all the conditions received the exact same cases and the exact same procedural solutions—the only difference was the order they received the materials. The study was conducted in a laboratory so it was also possible to ensure that the time on task was identical in all conditions.

Compute an index to indicate if there are different preferences

	Candy	Chocolate
Children	6	14
Adults	16	4

	Apples	Oranges
Pigs	14	6
Horses	16	4

Figure 3.3 Students needed to derive a single numerical index (value) for each matrix by using the same formula. Students read how to compute the index first, or they tried to invent the index and then they read about the standard solution.

After completing the activities, the students received a post-test. All the students completed the post-test working individually. The test included SPS questions that assessed students' abilities to efficiently apply the procedures to relatively similar problems. The test also included a PFL assessment. The PFL assessment took the same form as the study with high school students learning about variance. Embedded in the test was a worked example that provided hints for how to extend the chi-square logic to determine reliability (per Cohen's Kappa). The question was whether students would spontaneously learn enough from the worked example so they could subsequently solve a very difficult problem that required extending the worked example to a new situation involving instrument reliability.

Figure 3.4 summarizes the relevant results. The left panel shows that students in all four conditions learned the standard procedure to about the same level according to an SPS test. Thus, the invention activities did not interfere with learning and subsequently applying the efficient solution. The right panel shows the results of the PFL assessment. Students who were told the procedure before each case did not do very well by the PFL assessment. They had not been prepared to construct knowledge given the opportunity. The students who tried to invent a solution before hearing the standard approach did better. The PFL benefit was particularly strong for the students who worked in pairs to invent solutions during the initial instruction (they did not work in pairs when completing the test or the worked example in the PFL assessment).

Sears' study provides several informative results. One result is that the seemingly inefficient activity of trying to invent solutions does not entail inefficiency in subsequent performance, *if* students have a subsequent chance to learn the

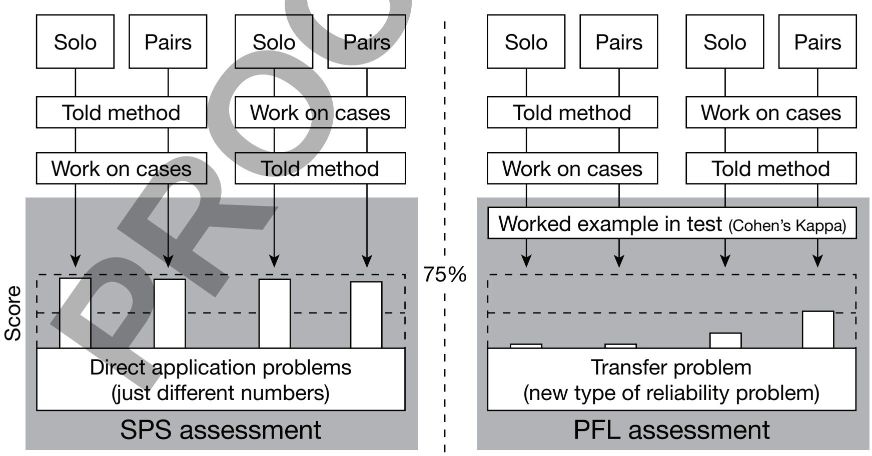

Figure 3.4 Students who tried to invent solutions before learning the canonical solution exhibited equally effective command of the procedures as students who were directly told how to solve the problems and then practiced. However, students who worked in pairs to invent solutions were far superior in their preparation to learn new methods from a worked example embedded in the text and apply them to new problems (source: adapted from Sears, 2006).

standard solution. As in the preceding studies, invention activities prepare students to learn canonical solutions and explanations quite effectively, and in this case, without an appreciable cost in overall instructional time. A second result is that being told an efficient procedure before working with the cases blocked student learning as measured by the PFL assessment. Knowing the procedure beforehand interfered with exploratory behaviors, because the students focused on the procedure rather than the situation presented in the problem. A third informative result is that the use of group discussion has specific benefits. Here we see the benefits of PFL assessments for identifying effective elements of instruction. Working individually or in groups did not make much of a difference for SPS assessments that measure efficiency. However, for constructivist outcomes, group work was beneficial—but only if the groups had an opportunity to invent solutions together. Working in groups in the context of direct instruction had no discernible benefits over working individually. These results make sense. When given direct instruction on how to solve problems, the pairs simply handed off their results to one another to check answers. In contrast, in the invention condition, the paired students talked with one another about their ideas, and this helped them explore their thoughts about the learning space more completely.

Conclusions: What Types of Activities Produce Constructivist Outcomes?

The aim of this chapter was to demonstrate the value of considering constructivist outcomes when evaluating the effectiveness of instruction. We proposed a style of assessment called Preparation for Future Learning (PFL) that seems better suited to the goals of constructivist-inspired pedagogies, which is to enable students to construct knowledge. In a PFL assessment, students receive an opportunity to construct knowledge during the assessment. The studies demonstrated that the PFL measures were sensitive to prospective aspects of constructivist learning that a standard assessment would have missed.

While we believe that PFL assessments are powerful tools and that their incorporation into mainstream educational institutions should be seriously considered, we fully acknowledge that there are significant challenges to a widespread conversion. It is a major issue, for instance, whether it will be possible to scale PFL assessments in this era of SPS testing. Establishing reliability at a level that is on a par with current standardized assessments will require a great deal of effort and collaboration among educational institutions and the organizations that oversee them. At a philosophical level, these entities would need to acknowledge the capacity for future learning as an explicit goal of instruction. We should also note that the effects described in the studies above were mostly likely due to the *content* knowledge the students acquired through the activities of exploring the problem space and inventing solutions. It would likely require longer and more persistent instructional interventions to affect the broader dispositions and the metacognitive capacities of students that regulate their abilities to construct knowledge on their own.

The studies presented in this chapter were specifically designed to demonstrate that PFL assessments, which are consistent with constructivist outcomes, reveal effects missed by SPS assessments. Now that this evidence has been established, at least so far, we hope that others will consider the inclusion of PFL assessments in their own research. It should be relatively easy to add PFL assessments to most instructional studies. Future applications of PFL assessments do not have to use the specialized research designs employed here, which were designed to demonstrate the validity of PFL assessments.

In the examples provided here, the PFL assessments all involved learning from verbal materials such as a lecture or worked example. However, one can imagine PFL assessments that are more interactive and take other forms. For example, technological environments make it possible for students to receive feedback to guide their learning at test. Presumably, good prior instruction would help students learn how to use feedback to construct knowledge. The work of Chen and Klahr (1999) provides an excellent instance. Students learned to conduct experiments in a computer environment, and the outcome measures included how well students arranged and used feedback from their own experiments to keep learning. Chen and Klahr called the most successful condition "direct instruction," but this is a gloss on the intricacy of their intervention which included many elements of exploration coupled with direct explanation.

Given that the constructivist measures revealed unique outcomes in the studies above, we can begin to map backward to discern what learning processes caused the effects. The preceding instructional designs were all variants on a common genre of instruction. This genre, which we describe below, does not include many of the elements that other educators consider important and include in their constructivist-inspired instruction; for example, the instruction presented above was not personally relevant to the students in any interesting way. Therefore, our observations are not meant to say that this genre of instruction is the only or the best model of instruction. Rather, we want to identify the elements and processes that seemed responsible for the current results. Our conclusions are necessarily speculative, because these studies were designed to isolate outcomes of learning not causes.

The successful activities described here share several common design ingredients (for more details, see Schwartz, Martin, & Nasir, 2005). One element was the use of contrasting cases. In each of the activities, students received materials that highlighted important features by using carefully selected contrasts. For example, in the study by Sears (2006), students received materials that juxtaposed cases where the expected values could be known based on probability (e.g., dice) versus cases where the expected values had to be inferred (e.g., by using the marginal means). Other cases contrasted ratio differences versus absolute frequency differences, and so forth. These contrasts helped students notice the important structural features and sources of variability in the learning space, as well as recognize what sources of variability are irrelevant. Although we do not have a hard rule, it has been our experience that adolescents and adults can handle cases that target three to four conceptually central contrasts at time.

When students engage in the inquiry and exploratory activities that comprise much of constructivist instruction, they are also engaging contrasting cases. For example, they may notice that two different actions lead to two different effects. A risk of poorly developed inquiry activities is that there can be too many contrasts, some less useful than others. While a broad range of possible contrasts will uncover many interesting and useful student ideas, too many contrasts make it difficult for students to discern which variables and interactions are most important. Moreover, in large classes, different students may follow the implications of different contrasts, which will make it difficult to "pull it together" for all the students in a class. In our approach, we pre-figure the contrasts to simplify the instructional task.

A second important feature was that the students were asked to invent representations for the cases, whether symbolic procedures or graphs. This was important for four reasons. The first, as demonstrated by Sears, is that students will not notice the structures highlighted by the contrasts if they are told at the outset how to use the correct procedure. They will focus on the procedure rather than the situational structures that make the procedure useful. Inventing the procedure focuses them on the situation and the procedural issues.

The second reason is that invention prepares students to appreciate the "why" captured inside the canonical solution. By working to invent solutions themselves, they begin to understand the issues that led to the design of the expert theory or procedure.

The third reason for having students do representational activities is that the goal of much school instruction is to help students learn to understand and use the compact symbolic representations and theories that experts use to organize complexity. Having students work towards these representations sets the stage for learning the canonical accounts.

A final reason for the invention activities is that students enjoy them, and there appears to be more engaged thinking and positive effects as a result. Students treat these activities as original production activities that promote creative thinking through the exploration of possible solution paths and representational artifacts. The solutions that students produce are sometimes suboptimal, but in general, students are not wrong in their inventions. Rather, their inventions simply do not handle all the cases or generalize to cases yet to be seen. When confronted with their "partial accuracy" students come to appreciate their own work, the work of others, and the standard solution.

Returning to the overall sequence of instruction, the third important feature that we have emphasized here is the eventual delivery of a comprehensive account of the cases. The goal is to prepare students to understand the account. The activities we have described are not discovery activities in the sense of expecting students to discover the canonical solutions on their own. In addition, activities that require "correct" student discovery can place a significant burden on instructors to artfully steer activity without "spilling the beans." Hills (2007), for example, reports that constructivist pedagogies can increase teacher anxiety. By removing the "pressure" of finding the right answer for a lesson to be successful, students and teachers are liberated to explore the learning space more fully. The canonical solution can be delivered after the exploration.

A Final Thought

There are multiple learning mechanisms and outcomes, and different situations can elicit them to varying degrees. Instructional theory can only be improved by understanding each of the ways that people learn, how to engage the relevant processes, and how to measure the types of knowledge and performance they yield. Yet, among all the different ways of learning, two clusters have been consistently called out and pitted against each other: constructivist-type learning versus direct-instruction type learning. It did not have to be this way. The instructional debate could instead have centered on other learning processes, for example, spatial versus social. For that matter, the enduring argument could have been about individual differences versus the "average" student, or a host of other issues. Nonetheless, the issue of constructivism versus direct instruction has dominated American educational debates on instructional technique for a very long time (Duffy & Jonassen, 1992).

The debate comes in many forms: passive versus active learning; shaping behavior versus discovery; inquiry versus authority; student centered versus teacher centered; school-as-life versus school-as-preparation-for-life. We suspect that the underlying fuel for much of this debate are (tacit) issues of free choice, expression, and agency as applied to students in classrooms. Through a loose analogy, direct instruction has been associated with controlling students, and constructivism has been associated with self-determination. We are not up to engaging this debate on the basis of evidence. One reason is that—writ large—the merits of being controlled and being self-determined is a normative, cultural question more than it is an empirical question about learning. As Dewey (1916) points out, these issues are deeply related to larger underlying questions in our society. Scientific data cannot prove the principles that define the outcomes we hold most dear. Data can only help us determine how to achieve those outcomes.

Yet, even if we limit ourselves to a narrow empirical question about content learning, we still cannot answer what types of instruction lead to the outcome of helping students construct knowledge (and conceivably, thereby be in a better position to make the choices that determine their lives). Empirical research has been using the wrong outcome measures of instruction. We have being using non-constructivist assessments in an era of constructivist beliefs.

Question: Sweller. *The "learning to learn" goal that is the focus of your chapter is highly desirable and has been pursued for several generations. In this chapter it is seen, correctly I believe, as an aspect of constructivism. My concerns are that I do not believe the goal can be reconciled with any viable cognitive architecture and more importantly, there is no body of supporting evidence that the goal can be attained. My questions concern the experiments claimed to support the basic proposition. As far as I could ascertain, while all of the experiments described were concerned with and indeed demonstrated enhanced learning under various constructivist regimes, none of them demonstrated enhanced learning outside of the content area being taught. If we claim to be preparing students for future* *learning, should we not be able to demonstrate enhanced learning in areas unrelated to those in which the future learning techniques were taught? I do not believe our cognitive architecture supports such teachable/learnable strategies but you presumably do.*

Reply: Schwartz et al. There are two parts to this question. The first is the presumption of a correct theory of cognitive architecture. If true, then the work of cognitive science is done, and all that is left is to draw out a few implications. We are not of this mind. The second part of your question, which we like much more, asks us to differentiate two elements of preparedness for future learning and transfer. One element is the "content" knowledge of a domain that helps people make sense of new related content. This is what our studies developed for the high school students—they learned the deep structure of the important concept of variability that carries through the domain of statistics. It is why they were prepared to transfer these ideas to learn about the more difficult concept of normalizing data. (In this respect, our work differs from the older "learning to learn" literature that focused on abstract skill instruction, for example in logic, that could conceivably transfer to any domain, but psychologically did not.) At the same time, the students were learning to think about a very general phenomenon, namely variance. Unlike learning a procedure about a specific device, there is some hope that they learned powerful content that can affect how they reason about the very many situations where variance is an important feature.

We think it is an incorrect read of the empirical literature to assume that people cannot learn strategies, concepts, and dispositions that transfer to improve learning across a broad class of situations. The evidence is quite clear that people can learn to read well, think in statistical terms, reason by analogy, control variables in research, and so forth. These are not domain general thinking skills in the sense of logic, which can be applied to all topics. Rather, they occupy the middle ground between what Newell and Simon (1972) called (a) weak methods—problem-solving methods that apply to all situations; and, (b) strong methods—methods that arise from knowledge about a specific domain. Perhaps this middle level should be called "protean" methods. They have the potential for flexible use within a domain and related fields that share similar underlying empirical structures.

A good example of a protean method for problem solving and learning might be the construction of visualizations to aid in the management of information. Creating visual representations is a fine way to help learn the structure inherent in complex information, and it can work across many domains. Lee Martin (Martin & Schwartz, in review) found that advanced graduate students in science are more likely to make visual representations to organize novel information than science undergraduates, even when the task has nothing to do with the topic of their graduate studies and the task could be completed without visualizations. Thus, the graduate students exhibited a relatively far transfer of a learning and problem-solving method that was not tied to any specific content, at least in the sense of content about biology or computer science or engineering.

What is notable about Martin's finding is that both the graduate and undergraduate students were very capable of making appropriate visualizations—the undergraduate students just did not bother. Given that it took several years of graduate school for the students to learn it was worth the time to make visualizations of data, it seems that protean learning strategies and their likely transfer take some time to develop. We suspect that given enough time, learners who go through our invention activities would develop some protean skills and dispositions associated with innovating new ideas, which is a type of learning that is quite different from replicating what one is told.

Question: Sweller. *In the research you describe, prior to being presented a lecture or text to read, learners were given several relevant cases that they were asked to analyze in order to attempt to find patterns. Schwartz and Bransford (1998) found that analyzing cases proved superior to summarizing the text while Schwartz and Martin (2004) found that analyzing cases was superior to being explicitly taught how to carry out a relevant calculation. It was concluded that exploration was superior to explicit instruction. My questions concern the appropriateness of the control groups used in these studies. If multiple factors are varied simultaneously, as they were in these experiments, does this procedure not break the "vary one thing at a time" rule essential to all randomized, controlled experiments, resulting in it being impossible to determine exactly what caused the effects? Can the authors of this chapter throw any light on why the obvious control groups seem never to be used?*

Reply: Schwartz et al. This is an important question. Thank you for the opportunity to clarify our approach to cognitive and instructional experimentation, which we see as different endeavors. There are three quick answers. One answer is that the first two studies in the chapter were on measurement and not about isolating singular causes. Until we measure the right thing, studies on causes or conditions of learning are irrelevant. Measurement drives science, and as the current climate of testing in the United States indicates, it also drives instruction.

The second answer is that we very much like the idea of titrating the specifics of our instruction in systematic research, and there are many people who are currently working on this problem. For example, the study by Sears, described at the end of the chapter, is a nice example where information and time on task were held constant, and the only difference was the order of the materials. We believe this is what you asked for in your question.

The third answer is that we do not favor the idea of simultaneously trying to (a) titrate what psychological interactions cause our instruction to be effective, and (b) conduct efficacy research to show that our instruction is more effective than some variant of standard practice. This ends up watering down one or both models of instruction, because people try to make the instruction similar, except for one thing. There are very many examples of this mistake, where researchers end up comparing two sub-optimal models of instruction that have the sole merit of differing on only one variable. For instance, comparisons of learning from diagrams versus animations run into this kind of problem, because people make the animations just like the diagram except that it moves (the one difference); this limits the types of interactions that animations can naturally provide such as pausing, replaying, slow motion, and so forth. Proving what is better and proving the mechanisms that lead to particular outcomes are very different endeavors.

Reaction to Schwartz et al.'s reply: Sweller. *I have two reactions to your reply to the questions. First, you seem to dismiss the importance of using cognitive architecture in instructional design because it presumes "a correct theory of cognitive architecture." Should we really dismiss cognitive architecture from instructional design so easily? Second, I really do not think I am wrong about varying one thing at a time in instructional, cognitive, or any other experiments. That does not prevent factorial experiments to look at interactions nor does it prevent us from investigating the effects of "pausing, replaying, slow motion, and so forth" on animation. It is just that if you vary all of these at once, you do not know which one or which combination caused any observed effects. Running properly controlled experiments that look at individual factors and interactions between factors is essential—and routine—irrespective of whether we are looking at psychological or instructional issues.*

Response to Sweller: Schwartz et al. We certainly did not mean to imply that conducting one-thing-at-a-time experiments is a bad thing to do. We do it a lot. As we have started to move into neuroscience, it has been humbling to see what "control" means in this context. For instance, in a normal behavioral experiment, we would never care about whether the learner turns the worksheet slightly while reading. In the context of a brain scan, different angles of stimulus presentation would be a terrible loss of control. That said, there are other types of productive research designs besides changing only one thing at a time, including fieldwork, epidemiological studies, and so on. Our point is that in the context of classic factorial designs it is important to separate efficacy research from causal research, and it is important to further separate causal research about external conditions of learning from causal research that tests theories.

Why is it important to keep these separate? It has something to do with the rhetoric of the constructivism debate. Can we compare direct instruction and constructivist-inspired instruction by only changing one variable? Imagine you create a model of direct instruction to teach employees about procedures for working with electrical equipment. Now imagine that somebody thinks they can do it better using a more constructivist approach. If they take an efficacy approach and compare their constructivist instruction to your direct instruction approach, you might complain they did not change only one thing at a time, per causal research. On the other hand, if they only study two variations of their own constructivist-inspired instruction, you will complain that they did not prove their method is more effective than yours. No wonder there is so little research that compares direct instruction and constructivism. People should be clear on whether they are testing the features of their own instruction or comparing their instruction against the best version of the other camp. The latter would not vary one variable at a time.

There is also a second, deeper issue: whether the one-variable-at-a-time research is in fact testing a causal theory or only testing whether one feature is more effective than another for a specific model of instruction. For instance, in your own excellent and compelling empirical studies you have confronted this challenge and explicitly acknowledged this issue. Your instructional results are consistent with the theory of cognitive load, but they do not directly test the internal causal mechanisms that give rise to this theory. The evidence is circumstantial with respect to the theory. To demonstrate that working-memory overload is responsible for a decline in performance, it is necessary to do more intrusive designs that involve things like double dissociations, interference paradigms, parametric loads on working memory, working memory covariates, attention tracking, and so on. Moreover, it would also be important to control secondary psychological effects like anxiety, which might confound the isolation of working memory as the causal mechanism. Suddenly one would be doing psychological research that has less and less ecological validity for conditions of instruction.

So, what is the point here? Changing one feature at a time in the context of instructional research is a good way to find out what works for your brand of instruction. It is not a good way to compare different instructional paradigms. It is rarely a good way to gather hard-nosed evidence to prove theories about internal learning mechanisms, unless one is willing to leave the realm of effective instruction to implement the very tight and often unnatural learning conditions that can isolate psychological processes in a precise way.

Question: Fletcher. *I appreciated this chapter a lot—especially the focus on PFL and beyond that real data! On constructivist issues! And that's not to overlook some clever designs to produce the data. My question relates to your focus on PFL as an assessment issue. Isn't it just a transfer task—something we know how to do? Is the assessment issue simply a matter of deciding whether or not to do it? It seems to me that an important outcome of your work is that you focus on a constructivist outcome that we actually know how to measure, allowing us to get serious about designing instructional environments that promote it.*

Reply: Schwartz et al. Our work is done! Thank you! Yes, the point is that if we can change our assessments, we will be in a much better position to learn how to design instruction that promotes constructivist outcomes. There is a subtlety in here, however. Most transfer tests have taken a sequestered problem-solving format (SPS), where people can only rely on their prior knowledge to solve the target problem. The typical metrics in an SPS transfer task include speed, accuracy, low variability, first-time correct transfer. SPS transfer measures are often a good way to separate instruction that leads to understanding versus memorization, so they are very useful to include as an assessment. However, SPS measures are about the efficiency with which people can replicate prior behaviors in a new context. In contrast, PFL transfer measures examine whether people can adapt and learn given a new situation. There has not been a tremendous amount of instructional or psychological research that uses PFL measures, but it is increasing. So, at this time, there is no off-the-shelf book of PFL assessments or instructional methods, although we have our beliefs about how to do these things. To make further headway, it will be important to address two key problems of PFL transfer. The first is the knowledge problem: What types of experiences help people develop the knowledge that can transfer to new settings to learn? The second, and more insidious of the two, is the inertia problem: What types of experiences make people feel it is worth the trouble to use what they have learned in new ways? This has not been addressed effectively by the cognitive literature, and we suspect that many demonstrations of failed transfer happen because people did not care enough to see if there were other ways to solve a given problem (cf. Pea et al., 2007). To help address these two problems through research, it would be wonderful if other people started using PFL transfer measures, so the field can start finding out what prepares people to construct knowledge when there is not an instructor telling them what to do.

Question: Fletcher. *I heartily agree that prior knowledge has a big impact on the amount and speed of learning in direct instruction, but doesn't prior knowledge have a far greater impact on constructivist-oriented instruction? Isn't it possible to employ direct instruction given far less prior knowledge on the part of the learner than we need to employ constructivist-oriented instruction? Or does the impact of prior knowledge depend more directly on the different types of instructional objectives likely to be sought with direct instruction, aiming lower in learning taxonomies, versus constructivist instruction, aiming higher in learning taxonomies?*

Reply: Schwartz et al. It is interesting that you predict that direct instruction works better for low prior knowledge. We would have predicted just the opposite. If the goal is to teach kids to tie their shoes, then direct instruction seems like a very good approach, if direct instruction includes opportunities to ask questions, try it out, and get useful feedback. But, if the kids have never seen a shoe, then it might be worth letting them explore shoes and laces a bit. Prior knowledge is not a monolithic construct, and there are different types of prior knowledge. So, we agree with your second option: Different instructional goals interact with different types of prior knowledge, and this interaction requires different instructional techniques.

Let us expand a bit by asking first, do experts or novices learn better from direct instruction? Take the case of giving a lecture. If it is a professional talk you count on the prior knowledge of the audience to make sense of what you are telling them. Despite slips and omissions, the audience can still get something out of the talk. Moreover, they will know the right questions to ask so they can clarify what they are learning. Now imagine you give the same professional talk to novices. There is a good chance they will not gain very much from the talk because they do not have sufficient prior experience of your particular topic. Even if you do a lot of work to ensure that you set up their prior knowledge just right, if you omit some critical information or use a slightly imperfect example, they will not be able to recover. They probably won't even know where to start asking questions.

Now, one might object that this contrast is unfair, because the lecture for a novice should be simpler than a lecture for an expert. Exactly! Experts can learn by being told more effectively than novices, if the content is in their domain.

If we look beyond the immediate lecture, direct instruction done poorly can also have a secondary effect. This involves the "practice" that comes after direct instruction, for example, when doing a problem set. Experts already know something about the structure of the domain, so they can benefit by just practicing the procedure. In contrast, the novices do not. Ideally, during the practice, novices will start to learn something about the structure of the problem space. But, direct instruction (and a lifetime of schooling) tends to focus students' attention to the told-solution procedures, not problem situations, so students learn answers to a problem space they never come to understand.

Next, we ask if novices or experts learn more effectively from "constructivist" activities. Again, the experts win. As one instance, the first author not-so-humbly submits he is better at designing effective psychological inquiry to learn something than first-year undergraduates. Thus, there is a main effect, where experts are better than novices when learning in their domain, regardless of the two types of learning experiences. This main effect is likely to swamp any of the subtle differences in the relative benefits of one type of instruction over another for novices versus experts.

Nevertheless, the relative comparison does help to highlight the challenge of constructivism for novices. Novices do not know how to learn very effectively by these methods. Good inquiry takes some time to develop and differs somewhat across domains (e.g., lawyers engage in a different type of inquiry than geologists do). This means that instructional designers have the substantial burden of making instructional materials that can guide novices to construct knowledge effectively through inquiry, discovery, and what have you. But this does not mean that it is a burden that should be avoided. We suggest that instructional designers make a mistake when they give in to the contention that making constructivist materials for novices is so difficult that it is better to just directly tell students what to do, because we know how to do that well enough. We also think there are many examples of effective constructivist instruction out there, even if the researchers have not compared their instruction to a variant of direct instruction. After all, if the goal is to develop effective constructivist pedagogies, then it seems like an inefficient use of precious resources to mount a direct-instruction comparison. The "what works" standards for this type of racehorse comparison—constructivism versus direct instruction—require a minimum 3-month intervention in schools with very many students and difficult-to-implement sampling schemes. Little of the research that people cite in favor of direct instruction has met these standards of evidence.

Question: Clark. *Thanks for an articulate and entertaining chapter. It seemed balanced, insightful, and focused on moving the discussion to topics more comfortable to advocates of constructivism such as motivation and instructional goals. First a question to clarify your view on the main topic of the debate:*

Since you acknowledge that constructivism has generally not succeeded at supporting learning better than guided instruction ("too large and general," "not at the right level for describing specific instructional decisions"), should we close the door on that argument and move to discuss interventions that influence motivation to *learn and reopen the dialogue about the conditions that promote the farther transfer of knowledge?*

Reply: Schwartz et al. We want to head off the lurking possibility of a presumptive question. If "guided instruction" should be translated into "direct instruction" in this question, then I think we should also point out that direct instruction has generally *not* succeeded at supporting learning better than constructivism. To our knowledge, the studies favoring direct instruction tend to be small-scale, use limited measures and time horizons, pick "skill acquisition" or simple concepts as the learning goals, and distort the constructivist control conditions. This makes sense. How would a person who studies direct instruction know how to make a good constructivist lesson and, of course, vice versa? Perhaps what we need is to find the world's greatest constructivist instructional designer and the world's greatest direct instructional designer and put them head-to-head in a big clinical trial. It is interesting to anticipate their negotiation on the terms of the head-to-head comparison. We guess that the negotiation would stall on the outcomes to be measured. By the end, both designers would concede that each form of instruction is generally better suited to particular outcomes. From there, they could argue on philosophical, economic, or other grounds for why their likely outcomes are more important. If "guided instruction" means a learning situation where there has been a careful arrangement of situations to support learning, whether by discovery, telling, or whatever, then we heartily agree. Well-arranged conditions for learning are going to be better than "sink or swim" versions of constructivism for most outcomes.

We assume that your point is really that terms like "guided instruction," "constructivism," and even the term "direct instruction" are too vague for specific instructional decisions. We agree completely. General theories of knowledge acquisition are too monolithic for moment-to-moment instructional decisions across a variety of contexts. We prefer Gagne's original approach of matching specific learning experiences for specific learning outcomes, and the field has learned a lot since his time that could be very useful in its precision. Even so, we are not sure we want to completely close the door on the debate. As we argued, the significance of constructivism for instruction is that it reminds us that there is more at stake for learners than being able to execute a skill when cued to do so. And, of course, direct instruction reminds us that free-form constructivism is inefficient for some outcomes. We also agree that some form of motivation needs to be let back into the cognitive lexicon, and that transfer is important for domains where students will need to adapt.

Question: Clark. *Isn't it possible that farther transfer is possible when instructivist-guided support employs what has been called varied (or variable) practice and haven't most of the reviews of past research on this issue concluded that the evidence only supported gradual and limited transfer despite claims to the contrary?*

Reply: Schwartz et al. Transfer, transfer, transfer ... It is nearly as contentious as constructivism versus direct instruction in some corners of academia. It is too big to handle here, though it is worth noting that if people could not learn for transfer, then they would be awfully stupid every time they walked into a new situation. A repeated finding—one that predates the cognitive revolution—is that a key to transfer is the opportunity to learn the deep structure beneath the surface elements of a situation. This way, the learners can recognize that structure in a new situation that may have different surface features and thereby transfer. "Variable practice" is one approach to this problem, because the hope is that students will learn what is common (i.e., the deep structure) amid the variability of problems and such.

The trick is how to get people to notice the deep structure in situations. Common models of direct instruction can inadvertently block learners from noticing that structure. For example, we just finished gathering data that compared two treatments. In the procedure-given treatment, children were told a formula (density), and they had to apply it to a set of cases to compute the answers. In the invent-the-procedure treatment, children were told to invent a formula to apply to the cases (which they were able to do). The next day, we asked the children from both treatments to remember the cases the best they could. The students in the invent-the-procedure condition recreated the deep structure of the cases. In contrast, the students in the procedure-given condition recreated the formulaic answers. They did not remember the structure of the cases. Moreover, they showed more memory for incidental surface features than the invent-the-procedure condition. Thus, a risk of direct instruction is that learners will focus on the procedure instead of the situation, and they will only notice the eye-catching surface features. This is exactly what prevents people from exhibiting transfer, because they never learn to recognize the structure beneath the surface features. It is notable that many of the instances of failed transfer involved direct instruction; now, we have some evidence why. The study participants learned the solution procedure, but they did not learn about the structure of situations for which that procedure might be useful.

Reaction to Schwartz's reply: Clark. *Transfer is one of the two issues that seem to be raised most often by constructivists when defending discovery and criticizing strong guidance (the other argument has to do with "ill-defined domains" of knowledge). One of the reasons that the transfer issue is so contentious may relate to the fact that transfer studies are most often designed and conducted by people who share a bias about transfer. It is not totally surprising that most of us find evidence for our own theory when we design the treatments that represent not only our approach but also the views of people who disagree with us. I doubt that I'm the best person to design a constructivist treatment in a study examining questions about the transfer potential of guided instruction. What seems to be called for is a series of collaborative studies where people who disagree about what works collaborate on design and implementation and agree ahead to jointly publish the results. It would be productive if we identified and tried to resolve our differences on thorny core issues such as the operational definition (or utility) of different types of transfer and knowledge domains; the metrics for measuring transfer; and the impact of constructivist or guided instruction on transfer. While we will still find opportunities to disagree, the* *disagreements will be based on evidence from studies that we jointly designed and conducted.*

References

Barron, B. (2004). Learning ecologies for technological fluency: Gender and experience differences. *Journal of Educational Computing Research, 31*(1), 1–36.

Boaler, J. (2002). *Experiencing school mathematics: Traditional and reform approaches to teaching and their impact on student learning*. Mahwah, NJ: Erlbaum.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of Research in Education* (24, pp. 61–101). Washington, DC: American Educational Research Association.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293–332.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*, 1098–1120.

Dewey, J. D. (1916). *Democracy and education*. New York: Macmillan.

Duffy, T. M., & Jonassen, D. H. (1992). Constructivism: New implications for instructional technology. In T. M. Duffy & H. H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation* (pp. 1–16). Hillsdale, NJ: Erlbaum.

Gagne, R. (1985). *The conditions of learning* (4th ed.). New York: Holt, Rinehart & Winston.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38.

Hills, T. (2007). Is constructivism risky? Social anxiety, classroom participation, competitive game play and constructivist preferences in teacher development. *Teacher Development, 11*, 335–352.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*, 99–107.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*, 661–667.

Lillard, A., & Else-Quest, N. (2006, September). Evaluating Montessori education. *Science, 313*, 1893–1894.

Luchins, A. S., & Luchins, E. H. (1959). *Rigidity of behavior: A variational approach to the effect of einstellung*. Eugene, OR: University of Oregon Books.

Martin, L., & Schwartz, D. L. (in press). Prospective adaptation in the use of representational tools. *Cognition & Instruction*.

Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59*, 14–19.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Okita, S. Y., Bailenson, J., & Schwartz, D. L. (2007). The mere belief of social interaction improves learning. In D. S. McNamara & J. G. Trafton (Eds.), *The proceedings of the 29th meeting of the Cognitive Science Society* (pp. 1355–1360). August, Nashville, TN.

Pea, R. D., Goldman, S., Martin, L., Blair, K. P., Booker, A., Esmonde, I., & Jimenez, O. (2007). Situations and values in family mathematics. In *Proceedings of CSCL-2007 (Computer-supported collaborative learning)* (pp. 26–35). Mahwah, NJ: Erlbaum Associates.

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning & Memory, 2*(1), 88–94.

Russell, S. H., Hancock, M. P., & McCullough, J. (2007, April). Benefits of undergraduate research experiences. *Science, 316*, 548–549.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition & Instruction, 16*, 475–522.

Schwartz, D. L., Bransford, J. D., & Sears, D. L. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–51). Greenwich, CT: Information Age Publishing.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for learning: The hidden efficiency of original student production in statistics instruction. *Cognition & Instruction, 22*, 129–184.

Schwartz, D. L., Martin, T., & Nasir, N. (2005). Designs for knowledge evolution: Towards a prescriptive theory for integrating first- and second-hand knowledge. In P. Gardenfors & P. Johansson (Eds.), *Cognition, education, and communication technology* (pp. 21–54). Mahwah, NJ: Erlbaum.

Sears, D. (2006). *Effects of innovation versus efficiency tasks on collaboration and learning.* Unpublished dissertation. Stanford University, Stanford, CA.

Seger, C. A., Prabhakaran, V., Poldrack, R., & Gabrieli, J. D. E. (2000). Neuronal activity differs between explicit and implicit learning of artificial grammar strings: An fMRI study. *Psychobiology, 28*, 283–292.

Skinner, B. F. (1986). Programmed instruction revisited. *Phi Delta Kappan, 68*, 103–110.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 592–604.

Verbeek, M. E. M., De Goede, P., Drent, P. J., & Wiepkema, P. R. (1999). Individual behavioural characteristics and dominance in aviary groups of great tits. *Behaviour, 136*, 23–48.

Vollmeyer, R., Burns, B., & Holyoak, K. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science, 20*, 75–100.