

Comparing Four Instructional Techniques for Promoting Robust Knowledge

J. Elizabeth Richey · Timothy J. Nokes-Malach

Published online: 1 June 2014

© Springer Science+Business Media New York 2014

Abstract Robust knowledge serves as a common instructional target in academic settings. Past research identifying characteristics of experts' knowledge across many domains can help clarify the features of robust knowledge as well as ways of assessing it. We review the expertise literature and identify three key features of robust knowledge (deep, connected, and coherent) and four means of assessing these features (perception, memory, problem solving, and transfer). Focusing on the domains of math and science learning, we examine how four instructional techniques—practice, worked examples, analogical comparison, and self-explanation—can promote key features of robust knowledge and how those features can be assessed. We conclude by discussing the implications of this framework for theory and practice.

Keywords Analogical comparison · Worked examples · Self-explanation · Practice · Expertise

Over the past several decades, cognitive and learning scientists have investigated a variety of instructional techniques to support learning from problem solving. These techniques include practice (Anzai and Simon 1979; Barron et al. 1998; Chen 1999; Nokes and Ohlsson 2005; Phye 1990, 2001), worked examples (Atkinson et al. 2003; Paas and Van Merriënboer 1994; Sweller and Cooper 1985; Ward and Sweller 1990), analogical comparison (Gentner et al. 2003; Gick and Holyoak 1983; Gick and Paterson 1992; Robins and Mayer 1993; Star and Rittle-Johnson 2009), and prompting self-explanation (Chi et al. 1989; Chi et al. 1994; Rittle-Johnson 2006). Although a great deal of work has investigated each technique individually, much less has been done to examine the similarities and differences across techniques. Such an analysis raises a number of interesting questions. What instructional features are shared across techniques? Do these techniques lead to the acquisition of the same type of knowledge or are there important differences in terms of what is learned? Are the techniques equally effective in facilitating the acquisition of robust knowledge? Are these techniques hypothesized to trigger similar or different cognitive processes? Would a combination of techniques have a cumulative effect?

J. E. Richey (✉) · T. J. Nokes-Malach

Department of Psychology and Learning Research and Development Center, University of Pittsburgh, 3939
O'Hara Street, Pittsburgh, PA 15260, USA
e-mail: jes1235@pitt.edu

Learning has been defined and operationalized in dramatically different ways across the vast body of literature examining learning outcomes of instructional techniques. As many papers fail to provide explicit definitions for terms such as “deep” or “conceptual” learning, the way they are operationalized in experiments often serves as the only avenue for comparing differences in meaning across research. The lack of clear, consistent terminology for learning outcomes makes it very difficult to see whether different types of instructional techniques lead to similar learning and transfer. Furthermore, this lack of a coherent theoretical framework hampers progress in identifying patterns in the features of instruction across techniques. To compare techniques, we must first define the knowledge targets of instruction in such a way that they can be standardized across theories and research.

For the current work, we identify “robust knowledge” as the general learning objective for each instructional technique. We propose a framework for acquiring robust knowledge based on Koedinger et al.’s (2012) knowledge–learning–instruction (KLI) framework, which examines the relationship between observable instructional and assessment events and inferred learning events and knowledge (Fig. 1). Our review is split into three sections in accordance with the goals of the work. We first describe a framework for acquiring robust knowledge, propose three features that characterize it, and describe assessment events that are likely to identify those features. Next, we review four common instructional techniques—practice, worked examples, analogical comparison, and prompting self-explanation—and examine the evidence for whether each promotes the separate features of robust knowledge. Finally, we discuss the theoretical implications of this analysis and potential directions for future research.

By analyzing a variety of techniques that have been shown to promote robust knowledge, we aim to identify critical features shared by those techniques as well as important features that may not be shared. The identification of such features could provide insight into instructional principles that go beyond any single technique and may help to bridge laboratory discoveries

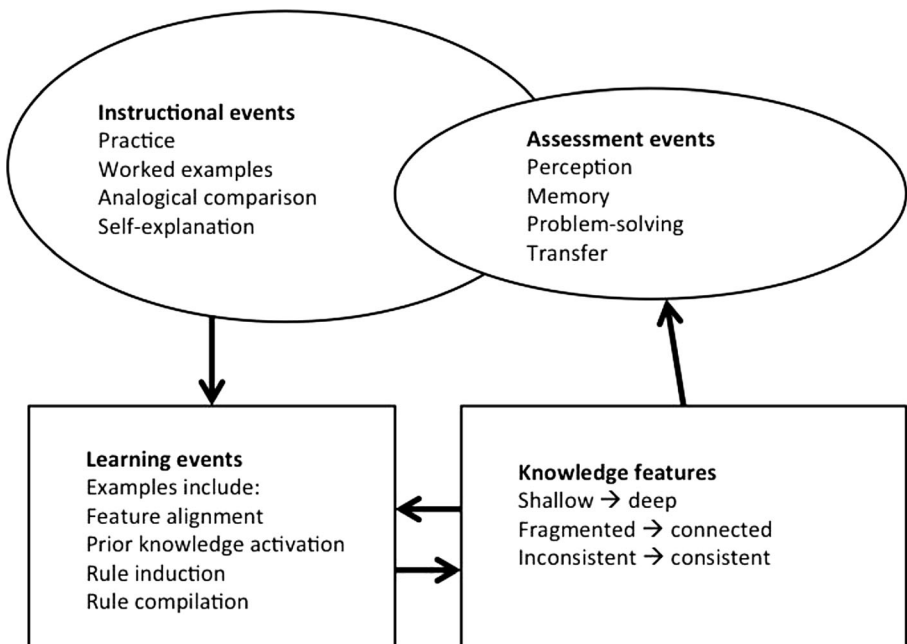


Fig. 1 Representation of robust knowledge acquisition, adapted from KLI framework (Koedinger et al. 2012)

to instructional interventions tested in classroom settings. Such an approach will also help to provide a conceptual framework for how these different interventions relate to one another and offer evidence of which techniques are best suited to promote particular learning and transfer outcomes.

A Framework of Robust Knowledge

The KLI framework was designed to integrate instruction, knowledge structures, and cognitive processes as core features of a unified, domain-independent theoretical framework (Koedinger et al. 2012). It incorporates four key elements: observable instructional events (e.g., practice problems and self-explanation prompts), inferred learning events (e.g., proceduralization of problem-solving steps and generation of inferences), inferred knowledge components that contain application conditions and responses (e.g., $1+1=2$, $F=ma$), and observable assessments that can serve as a means of inferring knowledge components and also as instructional events in themselves (e.g., measuring accuracy and reaction time to solve a geometry problem). By representing the relationship between instructional techniques and knowledge components, the KLI framework can be used to identify which instructional techniques are best suited for acquiring specific knowledge components (Koedinger et al. 2012).

Our review proposes a complementary analysis that examines robust knowledge at a larger grain size than the knowledge components proposed in the KLI framework. While knowledge components capture fine-grained, domain-specific knowledge content, we examine higher-level features of robust knowledge, which we identify and elaborate in the following sections. We structure our review of four instructional techniques within a KLI framework to articulate the specific changes in knowledge features that each instructional technique has been shown to produce. This framework also integrates assessments that can be used to detect changes as learners acquire robust knowledge. Based on this framework, we make specific predictions about the features of robust knowledge produced by each instructional technique, the opportunities for improved assessment of that knowledge, and future directions for better understanding the relationships between instructional techniques, acquired knowledge, and assessment outcomes.

Identifying the Key Features of Robust Knowledge: An Analysis of Experts

We draw on research from the expertise literature to identify features of robust knowledge. Experts have often been cited as a standard for measuring robust knowledge and learning (Nokes et al. 2010; Sternberg 1998). For example, Sternberg (1998) argued that the skills underpinning academic success could be viewed as developing expertise, whereas Bransford et al. (1986) suggested that explicitly focusing on the type of domain-specific knowledge that distinguishes experts from novices could bolster efforts to teach thinking and problem solving.

The expertise literature is informative for our framework of knowledge, learning, and instruction for several reasons. First, expertise has been the focus of extensive research, leaving us with a deep understanding of experts' knowledge and behaviors (Chi et al. 1988; Chi and Ohlsson 2005; Nokes et al. 2010; Ericsson and Smith 1991). Research on expertise highlights the importance of the underlying organization that supports robust knowledge. It places emphasis on the structure of knowledge and the role of long-term memories in making knowledge more flexible and adaptive (Schwartz et al. 2005). Differences in experts' behaviors are driven by differences in schemas stored within long-term memory structures. Schemas are hierarchical knowledge structures with deep features and connected relations between

variables that support expert behaviors. They organize knowledge through the relationships between different variables and values that fill those variables (Chi and Ohlsson 2005; Ohlsson 1993; Thorndyke 1984), and they can include application conditions and problem-solving procedures for applying knowledge. As a result of these representational differences in experts' knowledge structures, there are also behavioral differences in how experts approach problems and carry out their computations.

Second, instructors and teachers often want students to become expert-like, in the sense that they know some of what experts know (acquire disciplinary content knowledge) and behave in ways similar to experts (acquire disciplinary problem-solving skills). A number of different kinds of assessments have been used to identify expert-novice differences (Chi 2006), and we argue that these assessments can be applied more broadly to assess robust knowledge and identify the features of knowledge resulting from different instructional techniques.

We can use our understanding of expertise as both a target for what and how instructional content is taught and an indicator of the knowledge structures that students acquire as they learn. Although instructional techniques often use expertise as an implicit target, we can gain additional understanding by comparing the learning outcomes across techniques and relating them to the critical features of robust knowledge. This will provide a framework for interpreting the four instructional techniques of interest.

Features of Robust Knowledge

Across many studies of expertise, several key features of robust knowledge emerge. First, experts have conceptual and strategic knowledge advantages demonstrated in their more elaborate, abstract planning and more efficient strategies. Second, there are perceptual and memory advantages facilitated by the volume and connectedness of experts' knowledge. Their schemas allow them to perceive deep features and maintain more information about a problem state in working memory. Third, experts demonstrate highly accurate and consistent performance. These differences suggest that three key characteristics of robust knowledge are its tendency to be *deep*, *connected*, and *coherent* (Chi and Ohlsson 2005), characteristics that we define in greater detail in the following sections.

It is important to note that we are not claiming that all expert knowledge is deep, connected, and coherent. There are certainly circumstances in which experts generate novice-like behavior for particular tasks. It is also likely that experts have some knowledge that is incidental, isolated, and superficial, similar to novices' knowledge. Our point here is to emphasize the critical features that differentiate experts from novices and discuss how instruction might help students better acquire those features.

Deep Deep knowledge consists of the critical features or relations necessary to understand a problem and its solution (Chi and Ohlsson 2005). This knowledge can be articulated with respect to several different levels of analysis including the problem, task, domain, or discipline. Within a particular problem domain, there tends to be a set of deep features that are common to the structure of many problems in that domain. By contrast, superficial features are not broadly applicable across problems or relevant for solutions. For example, differences between novices' and experts' classification of physics problems suggest that experts connect perceptual information to deep knowledge (e.g., Newton's Laws) while novices connect it to superficial features such as whether the problem involves sliding blocks or pulleys (Chi et al. 1981; Novick 1988). Based on Chi et al.'s (1981) examination of a third group whose experience placed them in between experts and novices, this ability to see principle-based features appears to change incrementally as expertise develops.

There is no general rule for differentiating deep and superficial knowledge across domains without taking into account a domain-specific content analysis and the goals of the task (Lynch et al. 2000; Proffitt et al. 2000). However, the difference between deep and superficial elements has been demonstrated across a variety of domains, including computer programming, physics, geometry, medicine, literature, and dance (Adelson 1981; Allard and Starkes 1991; Chi et al. 1981; Hardiman et al. 1989; Koedinger and Anderson 1990; McKeithen et al. 1981; Schmidt and Boshuizen 1993; Schoenfeld and Herrmann 1982; Zeitz 1994). Deep knowledge also promotes forward-working strategies by allowing experts to identify principles and develop solutions based on these principles, rather than working backward from the goal (Simon and Simon 1978; Sweller et al. 1983).

Connected Connected knowledge consists of relations linking separate pieces of information to one another. Connections can relate abstract principles to specific problem features (Nokes-Malach et al. 2013), or they can relate principles to one another within or across domains. They can also link problem-solving steps within a problem, which may be represented declaratively (e.g., the steps can be verbalized) or procedurally (e.g., the steps can be executed in sequence with few cognitive resources and may not be verbally accessible). In addition, connections can relate steps or features across problems within the same domain or across domains. Chase and Simon's (1973) seminal work on chess expertise showed that experts tend to perceive underlying patterns based on knowledge connecting critical features. Although both experts' and novices' semantic networks are grouped by domain, with interconnected knowledge components, Chi and Koeske (1983) found that experts' semantic networks are more strongly and cohesively connected. Specifically, they found that when a child discussed a set of dinosaurs with which he was very familiar (more expert), compared with a set with which he was less familiar (more novice), he made many more links between dinosaurs that belonged in the same group (e.g., large plant eaters). Connecting concepts to procedures can make procedural knowledge more flexible and powerful by helping a learner to correct errors, apply procedural skills in novel situations, and strategically modify steps to overcome obstacles (Hiebert and Lefevre 1986; Ohlsson and Rees 1991).

Coherent Coherent knowledge is consistent and free of contradictions. Some contradictions lie in factual inaccuracies; for example, if a person believes gravity is simply the pull of the earth on other objects, this inaccurate belief would be contradicted by the information that every object exerts an equal and opposite force on the earth. Other cases only *appear* to be contradictions because of misconceptions or a lack of deep understanding. For example, that the gravitational force the earth exerts on an object is proportional to the object's mass might appear to contradict the fact that all objects fall to the earth at the same rate of acceleration in a vacuum. When the processes are better understood, however, the apparent contradiction is eliminated because the learner recognizes that an object with greater mass also has more inertia, resulting in the same acceleration due to gravity regardless of mass. Knowledge contradictions are extremely common as a person acquires information and replaces or modifies inaccurate beliefs, and the process of recognizing and resolving those contradictions can be extremely difficult (for a review of levels of misconceptions, see Chi 2008).

Coherent knowledge influences experts' behaviors by directing their attention to new information that is consistent with their existing knowledge framework. This can have implications for whether they see superficial or deep features as more consistent with their existing understanding. Novick (1988) found that experts were able to see superficial analogical matches as conflicting with their prior knowledge and instead chose deeper, structural matches. By contrast, novices went with the surface match over the structural match,

suggesting they are not able to override superficial consistency to see structural consistency. Research on misconceptions has repeatedly shown that novice learners are also poor at detecting inconsistencies in their knowledge and, when faced with contradictory evidence, are likely to ignore or reinterpret the evidence rather than revising their own ideas (Chinn and Brewer 1993; Dunbar et al. 2007; Posner et al. 1982).

In this section, we have identified three important characteristics of robust knowledge: that it is deep, connected, and coherent. This provides a set of features by which one can evaluate successful instructional interventions. We now turn to methods of assessment for identifying features of robust knowledge.

Assessments of Robust Knowledge

The memory structures that make robust knowledge deep, connected, and coherent give rise to differences in performances on four broad types of assessment events: those that target perception, memory, problem solving, and transfer. We describe evidence for the indicators of robust knowledge with each type of assessment event. Table 1 summarizes and characterizes the major behavioral features and hypothesized knowledge structures of experts, who have robust knowledge, and novices, who do not.

Differences in Perception

Experts perceive domain-relevant problems differently than novices. Experts tend to focus on structural information such as the underlying principles and relations contained in a problem,

Table 1 Behaviors and hypothesized knowledge structures of experts and novices

| Novice | Expert |
|--|--|
| Features | |
| Surface: Problem encoding typically includes superficial features not critical to understanding the problem. | Deep: Encoding typically includes critical features necessary to solve and understand the problem. |
| Fragmented: Knowledge about concepts is not connected, and new information is not connected to prior knowledge. | Connected: Knowledge is connected between problem-solving steps, across problems or concepts in domain, and across domains. |
| Inconsistent: Knowledge contains contradictions and may include robust misconceptions. | Coherent: Knowledge is consistent and free of contradictions. |
| Assessments | |
| Perception: Focuses on salient surface features of the problem. | Perception: Focuses on structural and conceptual features of the problem. |
| Memory: Recall constrained by working memory capacity, susceptible to memory encoding and retrieval errors. | Memory: Ability to recall many critical details of domain-relevant material. |
| Problem solving: Use of domain-general methods such as means-ends analysis, trial-and-error, and working backward from the goal to solve new problems. Error-prone, long solution times, and inconsistent performance. | Problem solving: Use of forward-working strategies and many domain-specific reasoning strategies and approaches for particular problem types. High accuracy, short solution times, consistent performance on routine problems. |
| Transfer: Inflexible knowledge based on surface details limits transfer to near problems with same surface details. | Transfer: Flexible knowledge to support transfer to new problems in the domain. |

while novices typically focus on salient surface information, even when the superficial details are not relevant to how a problem is solved or how variables relate to one another (Kellman and Garrigan 2009). Consequently, experts are more adept at encoding abstract, relational patterns of information from domain-relevant problems. This ability to see and encode the relevant problem information has been referred to as “feature-focusing” (Koedinger et al. 2012). Similar evidence regarding experts’ perception of deep features has been found for math (Schoenfeld and Herrmann 1982) and computer programming (Adelson 1981), among other domains. Students’ abilities to sort problems by structure can be improved through exercises such as practice matching variables and selecting common structure descriptions (Cummins 1992; Quilici and Mayer 2002) and training to analyze physics problems like experts (Dufresne et al. 1992).

Differences in Memory

Experts develop more efficient processes for storing information by using the structures in long-term memory to assist working memory (Ericsson and Kintsch 1995). The complex, interrelated structures that organize experts’ knowledge afford more robust and accessible long-term memories. For example, Chase and Simon (1973) demonstrated that expert chess players could reliably recall three to four times as many chess pieces as novices when studying various chessboard game scenarios. The authors hypothesized that the expert players had 10,000 to 100,000 chess piece configurations stored in long-term memory, enabling them to rapidly encode and recall the various configurations from real game scenarios. Experts’ vast memories of possible configurations captured important features of the domain, such as how different chess pieces interact with each other, constraints on how they can move, and strategies for defeating an opponent. Similar evidence for interconnected memory structures has been found in many domains including math (Koedinger and Anderson 1990), computer programming (McKeithen et al. 1981), medicine (Schmidt and Boshuizen 1993), and literature (Zeit 1994).

Differences in Problem Solving

In some domains, experts tend to do more planning than novices and, consequently, spend less time executing a series of problem steps. Voss et al. (1983) demonstrated a domain-general problem-solving behavior of experts by having three groups of expert faculty members (USSR policy experts, policy experts with no USSR specialization, and chemists) and a group of novice undergraduates enrolled in a USSR policy course address an ill-structured political science problem about the USSR. The authors found that while USSR policy experts had significantly more knowledge related to the specific context of the USSR, all three expert faculty groups demonstrated the same expert problem-solving behaviors. Compared with the novice group of undergraduates, they spent more time evaluating the questions, developed more elaborate problem representations, and proposed a small number of more complex and well developed solutions, rather than a large number of simple, poorly developed solutions.

Another problem-solving difference between experts and novices arises from the strategies they use in formal domains like mathematics and science. Novices tend to use means-end strategies, attempting to reduce the distance between the initial problem state and the goal state through the use of available operators, or working backward from the goal state to the initial problem state (Larkin et al. 1980). Experts, on the other hand, frequently work forward from the initial state to the goal state of a problem (Simon and Simon 1978). Koedinger and Anderson (1990) modeled experts’ tendencies to focus on abstract steps connected to

principles when planning geometry proofs, while often omitting steps tied to details less critical to the solution.

Differences in Transfer

Transfer assessments can help to distinguish robust knowledge from shallower or more fragile forms of knowledge (Day and Goldstone 2012; Hickey and Pellegrino 2005). We define transfer as the ability to use knowledge or skills acquired in one situation to solve novel problems in another (Nokes-Malach and Mestre 2013). Transfer assessments can test participants' abilities to understand a topic at a conceptual level (Brown and Kane 1988), adapt to changing conditions (Judd 1908), or learn from a new resource (Schwartz et al. 2005). Transfer assessments are often described as being near or far, with the distance typically referring to how similar the transfer scenario is to the original learning scenario (Barnett and Ceci 2002). Barnett and Ceci (2002) propose nine dimensions across which transfer distance can be measured, including factors concerning the content of what is transferred (e.g., procedures or concepts) and the context in which transfer occurs (e.g., time, modality, domain, etc.).

Experts possess a detailed, interrelated body of knowledge that can be applied flexibly to solve new problems and reach beyond that which has been explicitly taught (Hiebert and Lefevre 1986). Hatano and Inagaki (1986) proposed that "routine experts" cannot flexibly apply their knowledge to new problems, while "adaptive experts" can. This difference is hypothesized to arise from declarative, conceptual understanding of the content in the expert's domain. However, there is relatively little evidence of expert transfer outside their domain of expertise. Some domain-general strategies appear to transfer, such as ways of representing ill-structured problems and working toward solutions (Schunn and Anderson 1999; Voss et al. 1983), but the specific knowledge components of experts' schemas are generally not thought to support performance in novel domains (Ericsson and Charness 1994; Gick 1986).

Instructional Techniques that Facilitate Robust Learning

We now turn to the literature on instructional techniques hypothesized to facilitate robust learning. We examine a set of instructional techniques associated with a broad range of learning events. Koedinger et al. (2012) suggest a general taxonomy of learning events consisting of memory and fluency-building processes, induction and refinement processes, and understanding and sense-making processes. While different instructional techniques commonly facilitate different types of learning events, the authors note that learning events are not uniquely associated with different knowledge components and learning outcomes (e.g., it is not the case that only sense-making processes are associated with transfer). As all three classes of learning events can support the creation of robust knowledge, we chose to examine a set of instructional techniques that would, taken together, cover all three learning processes. Our representative sample includes practice (primarily associated with memory and fluency-building processes), worked example study (primarily associated with induction and refinement processes), analogical comparison (primarily associated with understanding and sense-making processes), and prompting self-explanation (primarily associated with understanding and sense-making processes) (Koedinger et al. 2013; Koedinger et al. 2012).

Prior work has investigated different theories for each technique, such as skill acquisition for practice, cognitive load for worked examples, analogy for comparison, and self-explanation and conceptual change for prompting explanations. Although this strategy has led to

significant progress in testing aspects of each of those theories, it does not easily lead to the development of a general theory of instruction that incorporates multiple techniques and relates the different techniques to one another and their roles in acquiring robust knowledge. Part of the problem stems from the absence of a common theoretical language shared across all of the techniques. It is possible, for example, that the different techniques promote similar types of knowledge but use different names for them. This is important for knowing which technique is needed for promoting a particular type of knowledge outcome (deep, consistent, or coherent). Based on these features, different techniques might be best suited for targeting different points in the learning trajectory.

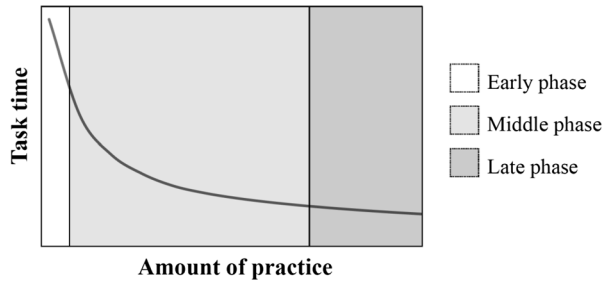
We focus our review on problem-solving instruction in the domains of math and science, with an emphasis on applications to classroom environments, although we also cite some foundational work conducted outside of traditional academic domains. We chose this focus because all of the techniques have been investigated in these domains, and reviewing work in the same domains makes it easier to compare across techniques. We describe each technique, provide examples, review the central findings, and relate the findings back to the features of robust knowledge. After introducing all four techniques, we analyze the commonalities and differences across them and, through this analysis, identify factors that promote features of robust knowledge. We also discuss limitations of each technique, highlight the knowledge features best supported by each one, and describe how they might work together to promote robust knowledge.

Practice

We define practice as repeated problem solving or task performance completed without any additional instructional techniques. “Practice” is sometimes used interchangeably with “problem solving”; other times, a practice-only condition is labeled “conventional problem solving.” For our purposes throughout this paper, we describe the process of repeated problem solving without an additional instructional intervention as practice. When repeated problem solving is interleaved with worked examples, prompts to reflect or explain, feedback, or some other form of instruction, we identify it as such. We use problem solving more generally to refer to any process of working from an initial problem state to a solution state.

Practice is often incorporated into instruction in math and science classes in the form of problem-solving activities (Richland et al. 2012). Practice is a common pedagogical feature built into most disciplines for the development of expertise (Brown et al. 1989), and there is a very large literature on practice in the cognitive and educational sciences (Chi and Ohlsson 2005; Ericsson et al. 1993; Newell and Rosenbloom 1981; VanLehn 1996). Research over the past 30 years suggests that practice can foster skill acquisition, but there are many factors involved in determining its effectiveness. We begin by describing the mechanisms of practice within a three-phase model of skill acquisition in which practice plays a key role (Fitts 1964; Newell and Rosenbloom 1981; Nokes et al. 2010; VanLehn 1996). Anderson and colleagues have proposed production compilation as the mechanism through which practice transforms declarative knowledge into a set of steps and procedures for solving a problem (Anderson 1987; Singley and Anderson 1985). This mechanism provides an account for how people can use declarative knowledge to solve a problem early in the skill acquisition process, and it models the types of errors and solution times people generate across the stages of skill acquisition (Anderson et al. 1997; Neves and Anderson 1981; Taatgen and Anderson 2002). Figure 2 illustrates the characteristic changes in time on task as a function of practice across the three phases of skill acquisition.

Fig. 2 Changes in task time based on amount of practice across the three phases of skill acquisition. This curve has been described as the power law of practice



Mechanisms of Practice

During the early phase of skill acquisition, learners primarily compile information about the skill, acquiring declarative knowledge from a lecture or demonstration, a worked example, an analogous problem, or some other source. The learners then use this declarative knowledge to attempt to solve the problem or perform the task. This phase is characterized behaviorally by long solution times and a high likelihood of making errors (see Fig. 2). Research has shown that learners during this phase often incur a high cognitive load and make errors due to memory retrieval failures (Chandler and Sweller 1991; Paas 1992; Sweller et al. 1998).

In the intermediate phase, learners refer less often to declarative sources of instruction for step-by-step guidance during problem solving and depend more on the procedures that they have been practicing (Anderson et al. 1997). In this phase, learners can address gaps or errors in their understanding, but the primary focus is attaining more accurate and efficient problem solving and task performance. This phase and the next phase are characterized by production compilation, in which two repeating steps or productions are concatenated into a single production rule (Taatgen and Lee 2003). The single production rule takes less time to execute, thereby providing a mechanism to account for the dramatic changes in reaction time typically observed during this phase, with learners getting faster and faster at performing the task or solving the problem (e.g., Singley and Anderson 1985).¹

During the late phase, learners continue to improve their speed and accuracy until the task becomes fully proceduralized, allowing them to perform a specific sequence of steps with no errors (VanLehn 1996). This phase is often described as automatic in that once a procedure is initiated, it is hard to interrupt or adapt to new or changing conditions, and it can be difficult for the learner to describe the individual steps of the procedure. At this stage, learners have often forgotten the declarative rules or examples they relied upon when practice began (Anderson et al. 1997). It is also hypothesized that executing these skills in the late phase requires very few cognitive resources (i.e., working memory). One outcome of this phase is extremely fast and accurate performance. Transfer from this stage of practicing a task to a new task is constrained by the degree to which the productions for the practiced task are also applicable to the new task (Singley and Anderson 1985). Anderson and colleagues have proposed the use-specificity principle, which hypothesizes that acquired procedural rules are goal-specific and will not transfer to tasks that have different goals, even if those tasks address the same content knowledge (e.g., Delaney et al. 1998; Singley and Anderson 1989).

¹ By contrast, Logan and colleagues have argued that practice enables learners to progress from an early phase characterized by reliance on general algorithms to a later phase characterized by retrieval of declarative memories of specific problems (Compton and Logan 1991; Logan 1988).

As described above, practice plays a large role in transitioning learners from declarative to procedural knowledge and enables them to solve similar problems accurately and efficiently with little cognitive effort (Singley and Anderson 1989). The three phases of skill acquisition have been shown to produce very consistent patterns of performance across many different tasks and domains, including geometry and physics (Neves and Anderson 1981; Newell and Rosenbloom 1981; VanLehn 1999). Expert performance emerges only after significant deliberate practice in the domain (Ericsson et al. 1993; Ericsson and Charness 1997), which entails repeated practice but also focus on the task at hand, effort to improve, and attention to timely feedback.

Limitations of Practice

This work supports the view that practice can help novices acquire the perceptual-memory components of expert performance, leading to highly accurate and consistent performance on routine problems in the domain (Neves and Anderson 1981; Newell and Rosenbloom 1981; VanLehn 1999). However, much research also suggests that procedural knowledge acquired through practice cannot be adapted easily to solve new problems that require different goals or have different structures (Singley and Anderson 1985). This outcome is consistent with problem-solving phenomena such as the “Einstellung effect,” in which an over-practiced, successful solution procedure is applied to a new problem that shares some similar features with the practice problems but also differs in critical ways such that applying the procedure results in either an incorrect or inefficient solution to the new problem (Luchins 1942). This result is often described as a rote or blind application of prior knowledge without reflection, which in some contexts can lead to poor performance. Examples are often given in mathematics where students learn a procedure through practice and then apply that same procedure to a new problem for which the procedure is not appropriate (e.g., Sherman and Bisanz 2009).

Factors Influencing Practice

Although work on proceduralization suggests that transfer is limited to contexts in which the same application conditions are present, other research and approaches have identified factors that can affect what is learned and transferred. These include investigations of the role of cognitive factors such as prior knowledge and cognitive load during practice as well as instructional factors such as the variability and the types of goals used in the practice problems. In addition, other approaches have examined procedural-to-declarative knowledge acquisition through practice. Below, we briefly describe how each of these approaches and factors can affect what is learned.

Declarative and Procedural Interactions While practice is generally viewed as helping the learner progress from relying on declarative knowledge to procedural knowledge, the two may work in a reciprocal fashion. For example, Capon and Kuhn (2004) found that an instructional condition that combined direct instruction with problem-solving practice improved learners' conceptual knowledge of a math concept after a delay more than direct instruction that did not include practice. In turn, conceptual understanding can drive the development of procedural skills both by improving learners' understanding of instructed procedures and supporting their ability to modify those procedures (Hiebert and Wearne 1996). Rittle-Johnson et al. (2001) argue that conceptual and procedural understanding relate through problem representations: conceptual knowledge improves attention to the key features of a problem, which in turn supports the use of procedures, while procedural knowledge frees up cognitive resources and may help learners identify misconceptions, which in turn can improve conceptual knowledge.

Prior Knowledge Prior knowledge also plays a critical role in what is learned from practice. Capon and Kuhn (2004) hypothesized that problem-based instruction encourages the activation of prior knowledge and its integration with new information, compared with lecture and discussion-based instruction. Ohlsson and Rees (1991) also have argued that principled knowledge can constrain the number of possible problem states a learner considers, thus guiding procedural learning. Some researchers have argued that the minimal level of guidance offered with some practice regimens cannot support learning until learners have acquired sufficient knowledge to guide themselves, based on limitations of working memory, the role of working memory in creating long-term knowledge, and research comparing guided and unguided instruction (Kirschner et al. 2006).

Cognitive Load Practice can facilitate transfer by freeing up cognitive resources, although this does not address the issue of the procedures themselves being specific to the conditions encoded from the original learning task (Cooper and Sweller 1987; Paas 1992). Significant practice can reduce cognitive load and free up working memory, providing an opportunity for reflective cognitive processes such as comparison or self-explanation. Through these forms of reflection (described in greater detail in later sections), a learner can abstract a general solution principle and schema from the problem.

Variability of the Practice Problems The variability of the practice problems also has implications for what is learned. High variability of practice problems can provide an opportunity for acquiring more abstract knowledge representations (Nokes and Ohlsson 2005). This can occur if the learner attempts to construct a problem representation or procedure that applies to all of the problems in the practice set. For example, Chen (1999) found that practice supported young participants' abstract schema induction only when they solved problems with variant solution procedures. When the participants practiced only one invariant solution procedure, most derived a specific schema bound to the procedure they would practice, instead of an abstract schema representing the general solution principle. The very specific schemas derived from invariant practice make transfer challenging, as learners demonstrate great difficulty modifying highly practiced procedures tied to specific problem conditions (Chen 1999). This is consistent with work suggesting that one mechanism for schema induction is analogical comparison, in which the learner is exposed to multiple examples that share the same general principles but have different superficial features (Gick and Holyoak 1983; Novick and Holyoak 1991). We discuss this specific instructional technique in a later section.

Goals The number and types of goals learners receive while practicing can also affect what is learned. Receiving a higher density of subgoals on practice problems improves speed on trained problems and reduces reliance on means-end strategies favored by novices (Sweller 1983). The same series of experiments showed that as a consequence of increasing rule induction, high-density subgoals also increase speed on transfer tasks. Sweller (1988) found that nonspecific goals led to better recall of trigonometry problems, and particularly the recall of structurally significant features, compared with specific goals. Performance on the problems, however, did not differ by goal type. Burns and Vollmeyer (2002) presented evidence that nonspecific goals encourage hypothesis testing, while specific goals promote goal focusing. These studies suggest that frequent but nonspecific goals encourage expert-like behaviors such as structural feature-focusing and lead to more flexible, abstract knowledge.

Relating Practice to Robust Knowledge Features

Practice can potentially lead to the acquisition of deep features (VanLehn and van de Sande 2009), depending on whether schema induction takes place based on the structure and type of practice. Learning outcomes depend on a myriad of factors including one's prior knowledge, cognitive load, goals, amount of practice, and variability of problem structure and content. By exposing the learner to many examples, practice provides an opportunity to learn to complete a procedure efficiently and with few errors; however, a well-practiced procedure is typically applied under conditions very similar to those under which it was learned, resulting in little far transfer (Singley and Anderson 1985, 1989). Proceduralized knowledge is specific, accurate, and fast but not flexible. Given the focus on attaining procedural knowledge, learners engaged in practice can easily focus on superficial features of the examples. A learner might recognize deep features through induction after significant practice, or the learner might engage in reflection across problems once proceduralization has reduced the cognitive load of the task. Transfer from practice seems limited to applying the same procedure to similar problems in new contexts, however, suggesting that neither of these processes is common (Salomon and Perkins 1989; Singley and Anderson 1989).

Standard practice regimens typically create procedural connections among steps within a problem but fewer connections across problems or between concepts and problem steps, as evidenced by participants' tendency to forget initial examples or principles once a task has been well practiced (Anderson et al. 1997). Practice primarily supports the acquisition of procedural knowledge connections within a problem through the process of proceduralization (Anderson 1993), as the learner transitions from declarative to procedural knowledge. Practice does little to improve declarative connections between concepts, as it promotes a reliance on nonverbal procedural knowledge (Singley and Anderson 1989).

Practice can lead to very high procedural consistency in the sense that a learner will tend to execute a well-practiced process whenever the application conditions are present and sometimes will over-apply the procedure, as discussed previously (e.g., Luchins 1942). Although it has not been well studied, practice may not affect conceptual coherence because it does not afford the reflection opportunities necessary to resolve conceptual incoherence. While practicing a series of steps, a learner rarely has reason to compare steps for inconsistencies. Furthermore, once those steps have been compiled into a series of productions, the learner loses the declarative access to the steps necessary for comparison. For these reasons, we argue that practice alone is unlikely to resolve misconceptions or inefficient problem-solving strategies and may at times exacerbate them (e.g., the Einstellung effect). In fact, students can become adept problem-solvers while still demonstrating misconceptions (Brown and Hammer 2008).

In summary, practice supports procedural learning, leading to fast, accurate performance, and long-term retention (see Table 2 for a summary of features and evidence of knowledge acquired through practice). While practice is a critical instructional pathway for proceduralizing knowledge, the degree to which a learner can acquire robust knowledge or recognize and correct errors in understanding is limited by the amount and type of practice.

Worked Examples

Worked examples refer to instruction that gives example problems with solutions (Atkinson et al. 2000; Cooper and Sweller 1987; Paas 1992; Sweller and Cooper 1985). They can range from sample problems with answers to step-by-step solutions with justifications or explanations accompanying each step (see Atkinson et al. 2000, for a review). Worked examples aim

Table 2 Robust knowledge features supported by practice

Depth

- + Supports the acquisition of condition-specific procedural features (Anderson 1987; Anderson et al. 1997; Singley and Anderson 1985; Taatgen and Lee 2003)
- 0 Declarative features lost with practice (Anderson et al. 1997)

Connectedness

- + Supports the creation of procedural connections between steps within a problem (Anderson 1987; Anderson et al. 1997; Singley and Anderson 1985, 1989)
- 0 Little evidence for supporting between-problem and between-domain connections (Anderson et al. 1997; Singley and Anderson 1985)

Coherence

- 0 Little evidence for promoting coherence and reducing misconceptions

to demonstrate correct problem-solving steps that learners can then apply on their own to similar problems. In laboratory studies, students who first studied sample algebra problems with solutions performed better on subsequent problems than if they were told to practice solving the same examples without solutions (Cooper and Sweller 1987; Sweller and Cooper 1985). Worked examples are also more efficient than principle- and definition-based hints offered in a computer learning environment (Ringenberg and VanLehn 2006). They have been studied most commonly in math and science domains, although this technique could be applied to illustrate any correct procedure or problem solution.

Worked examples are often most effective when combined with other instructional techniques such as practice (Kalyuga et al. 2001; Renkl and Atkinson 2003) and self-explanation (Atkinson et al. 2003; Renkl 2002). Worked examples are rarely used on their own, given the powerful results of combining worked examples with other techniques, although exclusive worked example study has still been shown to produce greater learning than practice (Paas and Van Merriënboer 1994). For the purposes of this review, we define worked examples as examples interleaved with some amount of practice. Later in this section we discuss key modifications to worked examples that have proven successful for increasing learning and transfer, and in later sections, we discuss evidence focused on combining different instructional techniques.

Mechanisms of Worked Examples

Worked examples support the development of robust knowledge by reducing the learner's reliance on means-end analysis and trial-and-error approaches (Owen and Sweller 1985; Sweller and Levine 1982; Sweller et al. 1982; Sweller et al. 1983; Sweller 1988). When learners begin solving problems, they are often unsure of which steps or strategies to apply. They frequently attempt to solve problems by inferring the next steps and keeping those intermediary states in mind. This approach leads to means-end analysis or trial-and-error approaches, which can result in a failure to discover the correct steps, strategies, or answers. Many novices confronted with a novel problem turn to means-ends analysis to reduce the gap between the problem and the solution (Larkin et al. 1980). The process of means-end analysis can interfere with learning, at least in part because of the large amount of information a learner must remember when manipulating a number of nonautomated operators while considering the problem, the goal, and the distance between the two (Sweller et al. 1983). Even when such approaches lead to the correct answer, they require much more time and thus are much less efficient than being told the correct steps.

Directly related to this issue, students who must infer and keep track of multiple problem states while solving problems can “overload” working memory and make mistakes. Sweller and colleagues’ (Chandler and Sweller 1991; Sweller et al. 1998) cognitive load theory builds on well-documented evidence of the limits of working memory, which determines the amount of information people can maintain in their minds at a given time without storing any of it in long-term memory (Ericsson and Kintsch 1995). The cognitive load theory suggests that this limit imposes natural constraints on a learner’s processing capacity. Cognitive load is created by factors inherent to the problem (e.g., complexity) and the learner (e.g., prior knowledge). Learners and instructors can alter a task to reduce the amount of load stemming from learning behaviors such as processing information, constructing representations, and automatizing procedures, as well as a wide range of sources not directly related to the task such as the format of the problem, the presence of a distractor task, or emotional input such as anxiety or fear (Chandler and Sweller 1991; Sweller et al. 1998). The cognitive load theory suggests that a large cognitive load can disrupt learning by reducing the cognitive resources available for recognizing and committing to memory the key structural features of a problem, as these are often more difficult to recognize than superficial features and thus may require deeper processing to identify. By providing the correct steps and strategies, worked examples provide students an opportunity to focus their limited processing capacity on encoding those steps and strategies into long-term memory.

Worked examples can reduce cognitive load by decreasing the size of the search space and diminishing the need for a learner to construct procedural solution steps, represent subgoals, or maintain multiple problem states. Reducing cognitive load should facilitate learning by leaving more mental resources available for recognizing critical problem features. Worked examples also reduce the likelihood of initially learning an incorrect solution or strategy and having that solution interfere with later problem solving (Paas 1992).

Limitations of Worked Examples

Although worked examples improve near transfer and promote more efficient learning, there is mixed evidence concerning their utility in facilitating far transfer. Sweller and Cooper (1985) initially argued that worked examples may do little to facilitate far transfer, which depends on schema abstraction. The authors found that worked examples improved efficiency and accuracy on similar algebra problems, compared with a practice-only condition, but they did not improve performance on dissimilar problems that required the same operations but differed in problem structure. More recent work has also shown that worked examples alone are not a reliable technique for promoting far transfer (Renkl et al. 2002). However, factors such as extended example study, variability of examples, backward fading, self-explanation prompts, and comparison can all improve far transfer from worked examples, and these factors will be discussed in greater detail below (Atkinson et al. 2003; Cooper and Sweller 1987; Nokes-Malach et al. 2013; Paas and Van Merriënboer 1994).

Factors Influencing Worked Examples

Despite the many potential benefits of using worked examples in place of rote practice in early skill acquisition, there are a number of factors that can influence how effectively worked examples promote learning through these mechanisms.

Goals Worked examples give students an opportunity to engage in more constructive cognitive processing than practice alone. One example of this has been examined through worked

examples with explicit subgoals, which represent intermediate problem states along the path to a solution. Labeling clusters of problem-solving steps as subgoals provides clues for applying relevant prior knowledge, limits the search space for correct procedures, and creates an opportunity for self-explanation (Catrambone and Holyoak 1990). Groups receiving labels on clusters of worked example steps perform better on novel problems than no-label groups (Catrambone 1996). Catrambone (1998) argued that meaningless or abstract labels require more effort during encoding but lead to better memory and acquisition of deep features by reducing focus on superficial features associated with the labeled subgoals and encouraging self-explanation of the abstract labels. This is consistent with work reviewed in the previous section suggesting that including nonspecific subgoals in practice improves learners' focus on structural information, hypothesis testing, and schema acquisition (Burns and Vollmeyer 2002; Sweller 1988).

Example Design Based on the cognitive load theory, Ward and Sweller (1990) explored a variety of worked examples and found that they were beneficial only to the extent that the worked examples directed attention to the solution path and did not increase cognitive load beyond that of a similar practice problem. While their work replicated previously demonstrated benefits of single-source worked examples over conventional problems, the authors found that worked examples were actually *less* effective than practice problems if the worked examples required students to split their attention between multiple sources of information. The authors hypothesized that the process of having to integrate multiple sources of information dramatically increased cognitive load, cancelling out the benefits of worked examples and actually exceeding the cognitive load created by the same problems in unsolved form. When the worked examples were reformatted to provide a single source of information, students studying the worked examples once again outperformed students who had practiced with traditional problems, as well as those who studied the split-source worked examples.

Prior Knowledge Cooper and Sweller (1987) conducted an extensive comparison of worked example study and practice using algebra problems. For learners with less prior knowledge given shorter problem or example sets, only those who studied worked examples showed evidence of schema acquisition, while practice and worked examples were equally effective for learners with greater prior knowledge or longer problem or example sets. While under the right conditions both practice and worked example study supported both near and far transfer, worked examples led to more efficient learning and transfer and proved more useful for students with less prior knowledge.

Explanations Providing too much information, as in the form of instructional explanations accompanying worked example steps, can also disrupt learning, possibly by reducing the activation of prior knowledge and the opportunity to engage in spontaneous self-explanation (Richey and Nokes-Malach 2013). Even if students are engaged in studying worked examples, evidence suggests that the act of generation can be a valuable learning process because it activates prior knowledge and brings it in direct relation to the problem at hand (Berthold and Renkl 2009; Chi 2000). A meta-analysis of instructional explanations used in worked examples showed that instructional explanations are minimally helpful and may not be more effective than self-explanation of worked examples (Wittwer and Renkl 2010).

Fading If solutions appear obvious or a student is unwilling to engage more deeply than the minimal level required by a problem, worked examples may suppress activation of prior

knowledge and inference generation by simply providing all the answers, which in turn reduces learners' opportunities to build connections and identify inconsistencies in their knowledge. Worked examples may be especially helpful for novices, but with experience standard practice problems may become more beneficial (Kalyuga et al. 2001).

Alternating between worked examples and practice problems or removing steps from worked examples and requiring students to fill in the missing pieces can encourage more active engagement while still reducing cognitive load. Known as "fading," this technique is most successful when worked examples are interwoven with problem-solving activities and when the scaffolding provided by the worked examples is reduced, or "fades," with successive problems (Atkinson et al. 2003). Fading the level of support provided with successive problems leads to more flexible knowledge and better performance on near-transfer tasks (Atkinson et al. 2003; Renkl et al. 2002). Even with steps faded, however, Atkinson et al. (2003) found that students failed to show greater performance on far-transfer items, suggesting that worked examples with faded steps still did not support the induction of abstract schemas that could be applied in superficially different situations. To achieve far transfer, the authors had to employ self-explanation prompts mixed within the faded worked examples.

Variability As with practice, increasing variability in worked examples can also support transfer (Paas and Van Merriënboer 1994). The authors argue that worked examples free up additional resources for extracting the shared, relevant features across examples, making variability that highlights those features particularly beneficial.

Relating Worked Examples to Robust Knowledge Features

Worked examples operate through a focus on acquiring declarative knowledge, rather than procedural, with students learning structural features within a problem such as principles, subgoals, and sometimes explanations of the application conditions for a principle or concept (Burns and Vollmeyer 2002; Renkl 1997; Sweller 1988). These constitute some features of deep knowledge. As worked examples provide solution steps, however, they also provide an opportunity for a learner to engage in superficial study by mimicking the steps provided on new problems without necessarily engaging in constructive practices such as inference generation (Koedinger and Alevan 2007; McLaren et al. 2008; Pirolli and Anderson 1985). The way a student engages with worked examples plays a large role in determining the depth of knowledge acquired from them (Renkl 1997).

Worked examples can promote connectedness between pieces of declarative knowledge by helping to highlight features and principles shared across problems (Fong and Nisbett 1991). Each step is explicitly presented, providing the student with declarative information to learn the steps and the order in which to apply them (Pirolli and Anderson 1985). Most worked examples increase declarative knowledge of problem-solving steps but they do not lead to proceduralization the way practice does because worked examples do not entail such extensive repetition; to achieve the speed and accuracy of a proceduralized skill, learners must transition from worked examples to practice problems (Renkl and Atkinson 2003). Worked examples promote connections between pieces of knowledge to the extent that the learner engages in explaining or justifying worked example steps using declarative knowledge (Renkl 1997, 2005). By reducing cognitive load, worked examples may leave more working memory capacity available for generative activities such as self-explaining, which in turn could help

learners to form connections across concepts. In addition, the learner is more likely to notice relationships among subgoals, procedures, and concepts, leading to greater connectedness among these critical elements. There is little evidence suggesting worked examples promote connections across domains.

Little work has examined whether worked examples improve knowledge coherence. Seeing correct steps should prevent learners from acquiring incorrect solutions, but there is little evidence that worked examples help learners to recognize errors in their understanding or prompt them to abandon incorrect problem-solving steps in favor of correct ones. Worked examples often incorporate more conceptual information compared with practice problems, but they do not provide a direct opportunity for recognizing contradictions between the newly acquired information and prior knowledge. Renkl (2014) suggests that worked example-problem pairs might help learners to recognize gaps in their knowledge while working through problems, and then facilitate opportunities to fill those gaps with the subsequent examples. However, no work has explicitly examined this mechanism. If worked examples do not provide opportunities to integrate new information into prior knowledge, then they could decrease coherence by introducing new, conflicting information without causing the learner to revise existing, inaccurate knowledge. Empirically, we know of no evidence testing the effects of worked examples on misconceptions.

In summary, worked examples reduce cognitive load, allowing students to focus on learning solution steps and deep features. Worked examples support the acquisition of declarative knowledge about steps and some deep concepts that support near transfer (see Table 3 for a summary of the features of knowledge acquired through worked example study). However, by providing clear solution paths, worked examples potentially create the opportunity for learners to rely on superficial processing, which may reduce some of the benefits mentioned above. While there is some evidence of worked example study supporting far transfer to problems with different or adapted structures from the ones in the examples, this far transfer seems to be supported through additional constructive activities such as self-explanation.

Analogical Comparison

There is a huge literature on analogy's role in problem solving, learning, and education. Analogy is the mapping of shared features or relationships between two examples, cases, or problems (Gentner 1983, 2002). A large amount of prior work has focused on understanding the processes of analogical problem solving including retrieval, mapping, and inference and

Table 3 Robust knowledge features supported by worked examples

Depth

- + Supports the acquisition of the declarative structural features of a problem including principles, subgoals, and application conditions (Burns and Vollmeyer 2002; Renkl 1997; Sweller 1988)

Connectedness

- + Supports the creation of connections within a problem, linking declarative problem steps to one another (Renkl 1997, 2005)
- + Provides an opportunity to link concepts to problem steps (Fong and Nisbett 1991)
- 0 Little evidence for supporting between-domain connections

Coherence

- 0 May help learners recognize gaps in their knowledge, but little evidence exists for this outcome or for the reduction of misconceptions (Renkl 2014)

the various cognitive and instructional factors that affect each stage of the process. Analogy allows learners to transfer prior knowledge to new situations, such as the use of familiar examples to solve novel problems, and it can serve as a powerful pedagogical tool. Analogy can also take the form of a comparison activity, which involves comparing between two examples in an effort to better understand the cases and concepts involved in those examples. Comparison focuses on the mapping and learning components of problem solving while eliminating the issue of retrieval, and it can facilitate schema acquisition from the comparison.

For the purposes of this paper, we focus the remainder of our discussion on analogical comparison, in which the learner compares two provided examples, then applies information derived from that comparison to a novel problem or uses it to construct an abstract representation (Gick and Holyoak 1983). We chose this focus because this instructional technique has been shown to facilitate learning and transfer in both math and science domains.

Mechanisms of Analogical Comparison

Analogical comparison involves two basic processes: mapping between analogs and drawing inferences from the comparisons. Similarities play a major role in helping learners identify mappings between analogs, and these similarities can be structural (those that are critical to the shared concepts and the comparison) or superficial (those that are noncritical to the underlying shared concepts). Learners are sensitive to differences between superficial and structural similarities, and past work has shown that people generally have a preference for the latter when asked to identify which is more important for the utility of an analogy (Clement and Gentner 1991; Markman and Gentner 1993). However, learners also have been shown to have a strong bias to recall similar surface features when retrieving an analog for comparison (Blanchette and Dunbar 2000; Reed et al. 1990). Analogical comparison instruction directly addresses the issue of superficial analog retrieval by providing both cases for comparison. The learner can thus focus on the comparison process of aligning cases, mapping between their features, and drawing inferences. In educational settings, the instructor often provides the cases.

In Gentner's structure-mapping theory, analogy involves finding relations and correspondences between structures (Gentner 1983). At first, learners might create representations based on domain-specific rules that are not particularly helpful outside of that domain's context. During comparison, however, learners have the opportunity to align features across cases and encode the commonalities shared between them. If the cases are structured to highlight higher-level relations, the learner can create more abstract and generalizable structural representations (Gentner 1983). Most theories of analogical comparison agree that far transfer, the application of existing knowledge to a novel situation in a different semantic domain, depends on abstract schemas (Novick and Holyoak 1991; Reeves and Weisberg 1994). Direct comparison between two cases may help a student to solve a similar problem but will probably not help solve a novel problem with nonalignable features (Novick and Holyoak 1991). To make the shared features between two cases applicable to more distant problems, the learner must adapt and abstract the shared features or principles between two cases. In other words, learners must extract deep knowledge about the principles' underlying relationships in the cases so they can then apply those principles to matching variables in new problems.

Limitations of Analogical Comparison

Analogical comparison generates robust knowledge in the form of abstract schemas and by linking key features to principles (Novick and Holyoak 1991). However, empirical evidence

suggests it is not always enough to simply offer analogs and instruct students to compare them (Reed 1989; Richland et al. 2007). The superficial similarities that most novices gravitate toward will, at best, help them solve only those problems that happen to share both superficial and structural features (Holyoak and Koh 1987). Ross and Kilbane (1997) provided further evidence that superficial details affect the mapping process, sometimes improving structural mapping and other times interfering with it depending on problem structure. Simply increasing the number of examples is also unlikely to improve success, as Scheiter and Gerjets (2006) found that the benefit of multiple examples depends on learning conditions. Having insufficient time or instructions has a larger negative effect on learning from three examples than from one example, and under such nonoptimal conditions, three examples are not better than one.

Factors Influencing Analogical Comparison

Through analogical comparison, a learner can extract a principle or set of key relationships from familiar problems and use them to solve a new problem. Comparing across examples can be an effective way to help a learner identify the critical features of a problem set, and this process can highlight the structural elements and relations that are most critical to the problems. Structural comparisons can help a learner to better understand and remember the conceptual correspondences between two cases, and knowledge of those correspondences can be transferred to novel problems that rely on similar structural principles. Gick and Holyoak (1983) demonstrated that learners were more likely to recognize and successfully apply a solution to a novel problem when they extracted the solution through comparison across multiple examples, as opposed to reading only one example. Gentner et al. (2009) have also shown that schemas abstracted from analogical comparison can in turn promote backward retrieval of additional analogous examples from memory. However, a number of factors in the design and instruction of analogical comparison can influence learning outcomes.

Scaffolding Comparison A number of studies in math and science have shown that people perform better on assessments of problem solving, conceptual understanding, and transfer when they are encouraged to compare across cases or examples, as opposed to seeing those examples sequentially (Catrambone and Holyoak 1989; Cummins 1992; Nokes-Malach et al. 2013; Rittle-Johnson and Star 2007; Star and Rittle-Johnson 2009; see Alfieri et al. 2013, for a review). Similarly, Cummins (1992) found that instructed comparison of examples improves a learner's focus on structure and equation selection, while sequential processing increases focus on superficial features.

A recent meta-analysis found that analogical comparison of cases in both classroom and laboratory settings leads to better learning outcomes than instruction using sequential cases, single cases, or nonanalogous cases, $d=0.50$ (Alfieri et al. 2013). Providing insight into the mechanisms responsible for the learning benefits, the authors found that positive outcomes were associated with instructions encouraging learners' effortful search for similarities but not differences between cases. Given novices' difficulty distinguishing superficial and structural features, it suggests that analogical comparisons must be carefully constructed to emphasize structural similarities (Guo et al. 2012). The meta-analysis also revealed that providing abstract principles for learners *after* they completed the comparison led to the greatest learning gains. This suggests benefits from inference generation during the comparison process and potentially from knowledge correction by comparing the abstracted principle with the provided principle to correct inconsistencies in prior knowledge.

In the classroom, analogical comparison can be supported through instructional techniques that encourage active reasoning and knowledge generation (Richland et al. 2007). Although analogies are frequently employed as instructional devices in classrooms, simply presenting an analogy to students is often ineffective (Richland et al. 2007). Based on an analysis of classroom videos of mathematics instruction from around the world, the authors discovered that high achievement was associated with analogies being presented with support cues such as hints and prompting questions. A number of techniques have been found to promote comparison, including labeling relations (Gentner and Medina 1998), using software to suggest alignments (Kolodner 1997), prompting subjects to rate similarity (Markman and Gentner 1993), asking directed questions (Catrambone and Holyoak 1989), providing guided-analogy prompts (Kurtz et al. 2001), and prompting descriptions of commonalities (Gick and Holyoak 1983; Loewenstein et al. 1999).

Bridging Analogies Brown and Clement (1989) employed analogical reasoning to help students overcome physics misconceptions; they proposed the use of “bridging” analogies to help students understand the analogical relationships between different elements. For example, to help students understand that a table exerts an upward force on a book resting on the table, contrary to the common misconception that the only force in the situation is that of the book exerting a downward force, the authors used an analogy of a book resting on a spring that compresses under the book. Although a physics expert would see the two situations as similar, physics students often reject the analogy because they believe the spring scenario is somehow different. The authors succeeded in getting students to accept the analogy after introducing the bridging analogy of a very thin, flexible table bending under the weight of a book. The authors note the challenge of getting students to accept an analogy that contradicts what they already (incorrectly) believe they understand, and they argue that simply presenting analogies for students will often be insufficient to override misconceptions. Instead, they suggest students may need to be guided through the process of analogical reasoning, using techniques such as bridging analogies to overcome misconceptions.

Relating Analogical Comparison to Robust Knowledge Features

Successful analogical comparison increases a learner’s understanding of deep features and the relationships among features, as these features and relationships must be abstracted to build schemas (Gentner 1983). Dellarosa (1985) found that comparison across isomorphic problems improved participants’ performance on a classification task based on solution procedures, indicating that they perceived the key, deep features of the problems in a way more similar to experts, but there was no effect when applying the procedures to solve problems. Carefully constructing comparisons to highlight key features or including scaffolding to support the learner through the comparison can improve the likelihood that analogical comparison will support deep knowledge acquisition (Gentner et al. 2003; Gentner 1983; Kurtz et al. 2001). Alternatively, poorly constructed or unscaffolded analogical comparisons can allow the learner to focus on shared superficial features across analogs, resulting in no deep feature learning (Kurtz et al. 2001).

Unlike practice and worked examples, analogical comparison emphasizes learning relations not only within a problem but also across problems and sometimes across domains. As a result of the emphasis on structural relationships, analogies help learners form connections in knowledge that the other techniques do not, particularly between structural features, principles, and the application conditions in which they can be applied (Novick and Holyoak 1991). For

example, a student instructed to compare two superficially different examples of projectile motion might come to understand they are really governed by the same principles and, consequently, the application conditions for those principles would then be connected to both sets of values. If learners focus on structural features, they may be able to derive abstract, general principles from the comparison, and those principles can also be connected to the specific features of the cases. Scaffolding, such as prompts to recognize similarities and differences across examples, can support the learner in forming connections (Gentner et al. 2003; Kurtz et al. 2001). Simply providing instructions to compare examples also encourages learners to connect elements, resulting in greater learning (Loewenstein et al. 1999).

Analogical comparison can serve as an opportunity for supporting coherence by allowing learners to recognize that different cases are governed by the same principle (e.g., Gick and Holyoak 1983) or to identify and revise flaws in their mental models (Gadgil et al. 2012). It has even been employed specifically as a method for reducing misconceptions (Brown and Clement 1989). However, incoherence and gaps in knowledge will disrupt a learner's ability to recognize similarities across examples, thus pushing the learner to focus on more superficial features (Booth and Koedinger 2008). As with worked examples, increasing knowledge coherence through analogical comparison also requires that learners integrate new information into their prior knowledge (Gadgil et al. 2012; Rittle-Johnson et al. 2009). Additionally, analogies that depend on superficial features may increase misconceptions by leading the learner to infer relationships that may not exist in all examples or promoting a focus on superficial relationships (Markman and Gentner 1993).

In summary, analogical comparison eliminates the challenges of retrieving an appropriate analogical match. Comparison between two cases that share structural features can help a learner abstract general principles and form connections between the features and principles (see Table 4 for a summary of the features of knowledge acquired through analogical comparison). Learners often fail to see the opportunity for analogical comparison when cases are presented sequentially, and they may focus on superficial similarities or misalign structural features even when directly instructed to compare. Scaffolding that facilitates fruitful comparisons can improve the utility of analogical comparison, although the learner may still require additional support to integrate the new principles into prior knowledge. Selection of cases, placement of comparisons within the learning process, and supporting materials highlighting key principles can also affect learning outcomes.

Table 4 Robust knowledge features supported by analogical comparison

Depth

- + Supports the acquisition of structural features and relations (Dellarosa 1985; Kurtz et al. 2001)
- + Allows for abstraction of general principles (Gentner et al. 2003; Gentner 1983; Gick and Holyoak 1983)
- 0 Little evidence for creating proceduralized knowledge

Connectedness

- + Scaffolds connections between features and relations both within and across problems (Gentner et al. 2003; Kurtz et al. 2001)
- + Can lead to connections between problem features and abstract principles (Novick and Holyoak 1991; Nokes-Malach et al. 2013)

Coherence

- + May support coherence by addressing misconceptions (Brown and Clement 1989)
-

Self-explanation

Self-explanation is the process of generating explanations of some target instructional material. It can be spontaneous, as in the case of students who study a text or example and then pause to explain it in their own words (Chi et al. 1989; Renkl 1997), or it can be prompted, as in the case of materials that intersperse prompts instructing students to reflect and explain key ideas (Atkinson et al. 2003; Chi et al. 1994; Nokes et al. 2011).

Mechanisms of Self-Explanation

Nokes et al. (2011) highlights two types of self-explanations shown to support learning: those that focus learners on filling in gaps in the provided text or examples (e.g., Conati and VanLehn 2000; Hausmann and Chi 2002), and those that focus on revising errors and inconsistencies in prior knowledge (Chi et al. 1994; Hausmann and VanLehn 2007). Gap-filling self-explanations draw inferences to justify provided content or fill in missing information, and they are especially useful when the learner has little relevant prior knowledge. Mental-model revising self-explanations also rely on inference generation, but they focus on connecting prior knowledge to provided content and addressing conflicts between the two (Nokes et al. 2011). Self-explanation of new strategies may also operate through the mechanism of improving subgoal understanding and management, as opposed to better recall or increased likelihood of selecting a given strategy (Crowley and Siegler 1999).

Self-explanation can encourage learners to identify and elaborate on the critical features of problems, including the underlying principles (Atkinson et al. 2003; Chi and VanLehn 1991), the conditions for applying those principles (Chi et al. 1989), and the logic and subgoals for applying them (Catrambone 1998; Crowley and Siegler 1999). These critical features tend to apply across problems within a domain. By recognizing and understanding these features, a learner is more likely to successfully transfer knowledge to a novel problem (Atkinson et al. 2003).

Critically, providing detailed instructional explanations does not produce similar positive effects (Richey and Nokes-Malach 2013), possibly because the instructional explanations reduce self-explanations (Schworm and Renkl 2006). Even when self-explanations are incomplete or inaccurate, they can be more effective than studying provided instructional explanations as the latter does not promote the same kind of prior knowledge activation, inference generation, and revision of existing knowledge (Hausmann and VanLehn 2007). The act of generating explanations is important, and it seems that students don't derive the same benefits from being told explanations (Crowley and Siegler 1999) or paraphrasing instructional explanations (Hausmann and VanLehn 2007).

Limitations of Self-Explanation

Only some learners self-explain without any prompting. Chi et al. (1989) found that the students who scored best on a series of physics problems after studying physics materials were the ones who generated the most self-explanations. Furthermore, the content of the self-explanations that students generate is strongly predictive of learning. Chi et al.'s (1989) analysis of protocols offered a number of clues about the mechanisms that may have supported the high-performing students' success. They found that the high-performing students focused on conditions for executing a problem step, thus building an understanding of the structural features associated with a correct step rather than using superficial features as cues. High-performing students often drew inferences to expand their understanding, while the low-performing students' explanations often focused on restating problem elements (e.g., reading

a force diagram and stating which forces were acting on which objects). High-performing students also gained many more knowledge components about the physics laws they were studying than the low-performing students, who gained very few. Furthermore, low-performing students spent more time re-reading examples while problem solving, suggesting they had extracted less information from the examples compared with the high-performing students. In other research, explanations focused on principles and positive monitoring have been associated with better performance (Ainsworth and Burcham 2007).

Factors Influencing Self-Explanation

Given the high variability in the types of self-explanations that students generate, and the impact of that variability on learning outcomes, a number of factors have been investigated with the aim of increasing students' fruitful self-explanations.

Prompting Given that spontaneous self-explanation is associated with positive learning outcomes but occurs infrequently (Chi et al. 1989), researchers have pursued the question of whether prompting self-explanation can produce similar effects. Applying self-explanation prompts to an expository text, Chi et al. (1994) found that prompted students showed greater learning gains than unprompted students on a variety of conceptual measures. The performance benefits for students prompted to self-explain were most pronounced on the most difficult questions, which required students to induce information from the provided content. This suggests that self-explanation may be particularly beneficial for inference generation and sense-making, which is consistent with the KLI taxonomy (Koedinger et al. 2013). Similarly, Atkinson et al. (2003) found that self-explanation prompts incorporated in faded worked examples led to greater near and far transfer.

Type of Explanations Self-explanations can be trained or prompted, but there are still big individual differences in the length, depth, and effectiveness of the explanations learners generate. These differences translate into differences in performance outcomes (Roy and Chi 2005) and reflect the effortful processes (e.g., knowledge revision and inference generation) that good self-explanations require on the part of the learner (Chi 2000). While completeness and accuracy are not necessary to make self-explanations fruitful (Hausmann and VanLehn 2007), the focus and content matter. Simply increasing the volume of self-explanations is not necessarily beneficial, though there are benefits to increasing the number of explanations used to generate inferences, integrate new information into prior knowledge, or revise prior misconceptions (Ainsworth and Burcham 2007). In another example of explanations not focusing on the most fruitful elements, Rittle-Johnson (2006) found that children instructed to self-explain math problems focused explanations on procedures and explaining why an answer was correct or incorrect. They rarely talked about concepts, which may explain why the self-explanation condition performed better on procedural assessments but no better on conceptual questions compared with a condition that did not receive self-explanation prompts.

Type of Material Given that the benefits of self-explanation stem as much from the information that is not provided and thus must be constructed, Roy and Chi (2005) argue that self-explanation is more fruitful with materials that require more integration and explanation on the part of the learner, such as diagrams or multimedia lessons, compared with materials that require less work on the part of the learner, such as texts. However, Ainsworth and Burcham (2007) found that although less coherent text encourages people to make more self-explanations, these additional self-explanations did not promote more robust learning.

Explanations of less coherent text included more goal-driven explanations, which were not associated with performance, and false explanations, which were negatively associated with performance (Ainsworth and Burcham 2007).

Relating Self-Explanation to Robust Knowledge Features

Self-explanation can encourage a learner to focus on deep, structural features within problems or concepts when attempting to fill gaps in novel material (Conati and VanLehn 2000; Hausmann and Chi 2002) and revise flawed mental models (Chi 2000). Some learners spontaneously engage in self-explanation, but others produce relatively sparse self-explanations even when prompted (Chi et al. 1989, 1994). Explanations that focus on certain superficial features, such as describing surface elements and re-reading examples, lead to less learning and transfer than explanations focused on structural features such as principles and application conditions (Ainsworth and Burcham 2007; Chi et al. 1989; Rittle-Johnson 2006). The materials provided for self-explanation can increase deep learning outcomes by requiring more integration and explanation on the part of the learner (Roy and Chi 2005), but they must be coherent enough to promote fruitful explanations (Ainsworth and Burcham 2007).

Self-explanation encourages connectedness as learners attempt to relate new information to prior knowledge (Chi et al. 1989) and examples to principles (Chi and VanLehn 1991). Chi et al. (1989) found that students who frequently self-explained were more likely to generate explanations relating steps to underlying principles from the text, thus focusing on the connections between examples and principles. They also generated more connections to their prior knowledge. These connections can prompt learners to draw inferences that fill gaps in existing knowledge (Chi et al. 1994; Chi and VanLehn 1991; VanLehn and Jones 1993). Connections can be formed among problem features (Nokes-Malach et al. 2013), between problem features and principles (Aleven et al. 2003), or between what is being learned and prior knowledge (Chi et al. 1989). Connections between problem features and principles can help the learner recognize the appropriate principle to apply based on a novel problem's features, thus improving transfer performance. Inferences generated to connect pieces of knowledge can also fill in gaps in the learner's understanding. When these connections are generated through self-explanation, they are typically explicit, declarative connections.

Multimedia support may help students in generating self-explanations that connect new content to prior knowledge (Roy and Chi 2005), although it can also provide additional resources, such as a glossary, that might permit students to copy explanations without building connections to prior knowledge (Aleven et al. 2003). Critically, most of the focus of self-explanation in problem solving is on filling gaps and making inferences within problems. Rarely does self-explanation target comparison across problems, which might create more connections across problems and domains.

While generating inferences and forming connections, learners also have the opportunity to notice inconsistencies in their understanding or contradictions between their prior knowledge and the new information. Chi et al. (1989) found students who generated more self-explanations made more monitoring statements in general than students who self-explained less frequently and also that a significantly greater portion of these monitoring statements focused on a lack of understanding. While the more-frequent self-explainers' statements were roughly split between identifying a failure to understand and declaring a satisfactory understanding, the less-frequent self-explainers' statements usually confirmed their understanding and rarely identified misunderstandings. Students who recognized their comprehension limits as they were self-explaining also learned more than those who did not. Based on the frequency

with which students fail to recognize that they do *not* understand something, the authors argue that monitoring statements identifying comprehension failures are particularly important for correcting misconceptions and creating more coherent knowledge. As learners identify and fill gaps in their understanding through self-explanation, they may engage in mental model revision to correct errors in their understanding.

Self-explanation is the only technique for which there is much evidence for resolving misconceptions and fragmented knowledge. Self-explanation encourages the learner to identify and correct contradictions within prior knowledge or between prior knowledge and new information. While generating inferences and forming connections, learners also have the opportunity to notice contradictions between their prior knowledge and the new information. Consequently, self-explanation also affords learners the opportunity to address incoherence, and students who recognize their comprehension limits as they are self-explaining learn more than those who do not (Chi et al. 1989).

In summary, self-explanation is subject to large individual differences, both in spontaneous and prompted occurrences. Self-explanation is highly effective when it focuses on deep principles, connecting new information to prior knowledge, and identifying and resolving contradictions in prior knowledge (see Table 5 for a summary of the features of knowledge acquired through self-explanation). Self-explanations focused on more superficial features of material provide little instructional value, however, and it is challenging to create scaffolding that will encourage all students to engage in fruitful self-explanations.

Comparing and Contrasting Instructional Techniques

We have shown that four common instructional techniques—practice, worked examples, analogical comparison, and self-explanation—can all be understood as addressing some of the critical features of robust knowledge structures: being deep, connected, or coherent. We now integrate these reviews of instructional techniques to suggest which techniques are best suited to facilitate different features of robust knowledge.

Table 5 Robust knowledge features supported by self-explanation

| | |
|---------------|---|
| Depth | |
| + | Supports the acquisition of structural features including principles, goals and subgoals, and application conditions (Ainsworth and Burcham 2007; Chi et al. 1989; Crowley and Siegler 1999; Rittle-Johnson 2006) |
| 0 | Little evidence for creating proceduralized knowledge |
| Connectedness | |
| + | Supports the acquisition of declarative connections within problem steps as well as between principles and problem features (Aleven et al. 2003) |
| + | Supports connections between new information and prior knowledge (Ainsworth and Burcham 2007; Chi et al. 1989; Hausmann and VanLehn 2007) |
| + | Some evidence for supporting declarative between-problem connections, based on improvements in procedural transfer measures (Rittle-Johnson 2006) |
| Coherence | |
| + | Supports coherence by highlighting inconsistencies and gaps in learners' knowledge (Ainsworth and Burcham 2007; Chi et al. 1989; Chi et al. 1994) |
| + | May help learner recognize inaccurate conceptions (Chi 2000) |

Deep Knowledge

Deep knowledge depends on understanding critical features or relations (Chi and Ohlsson 2005), and it supports experts' abilities to classify examples based on principles (Adelson 1981; Allard and Starkes 1991; Chi et al. 1981; Hardiman et al. 1989; Koedinger and Anderson 1990; McKeithen et al. 1981; Schmidt and Boshuizen 1993; Schoenfeld and Herrmann 1982; Zeitz 1994) and employ principle-driven, forward-working strategies (Koedinger and Anderson 1990; Larkin et al. 1980; Simon and Simon 1978). Practice leads to highly accurate, efficient, proceduralized knowledge, but it does not generally promote deep knowledge of principles unless it is used in combination with other techniques. Worked examples make principles more explicit, though sometimes the deep knowledge acquired from worked example study is connected to the superficial features of the examples. Analogical comparison can highlight deep principles across cases, and scaffolding can reduce the likelihood that a learner may instead focus on superficial commonalities. Self-explanation provides an opportunity for the learner to identify underlying principles, and prompts can encourage more fruitful explanations, particularly if they target the appropriate learning goals.

Connected Knowledge

Connections exist in both conceptual and procedural knowledge: they relate concepts to one another and to instances of those concepts, and they link sequential steps within a procedure. We have identified multiple levels of connectedness, and we argue that different instructional techniques target connections at different levels. Practice promotes proceduralization within skills and tasks, which connects nonverbal representations, but there is little evidence that it promotes any connections across problems, domains, or principles. Worked examples promote connections between declarative steps, and depending on the level of instructional explanation included in the worked examples, it may also create an opportunity for connecting problem steps to principles. Analogical comparison facilitates connections between features and relationships both within problems and across problems. It can also promote connections to abstract principles and, depending on the cases selected for comparison, it can lead to between-domain connections. Depending on the materials and the focus of the learner's explanations, self-explanation can support connections between new information and prior knowledge; declarative problem steps; or features and principles (Nokes et al. 2011).

Coherent Knowledge

Coherent knowledge, which is free of contradictions, allows experts to behave in structurally consistent ways when applying their knowledge to solve problems. There is little evidence that practice or worked examples promote coherence or help learners resolve misconceptions. While worked examples model accurate problem solving and thus should help learners identify errors in their declarative knowledge of problem-solving steps, we know of few studies that have directly examined this outcome. Some evidence suggests that analogical comparison can be used to revise science misconceptions, at least when one of the analogs represents the learner's flawed model, but further research is needed to examine the analog features necessary for improving conceptual coherence (Gadgil et al. 2012). Self-explanation has the most empirical support for improving coherence and addressing misconceptions, as it provides an opportunity for the learner to make direct comparisons between instructional materials and prior knowledge. Even with prompted self-explanation, however, this process is effortful and many learners fail to engage in it productively.

Assessments of Robust Knowledge

Robust knowledge improves perception of structural and conceptual features (Adelson 1981; Chi et al. 1981; Kellman and Garrigan 2009; Schoenfeld and Herrmann 1982). Little work has assessed the effectiveness of practice, worked examples, or self-explanation using perceptual assessments. It has frequently been used to assess the knowledge acquired through analogical comparison in the form of problem-sorting tasks (Booth and Koedinger 2008; Cummins 1992; Dellarosa 1985; Dunbar et al. 2007), ratings of perceived similarity (Kurtz et al. 2001), or learners' selections of good analogical matches (Reed et al. 1990).

Working memory capacity and encoding and retrieval errors often disrupt memory performance, but robust knowledge improves a learner's ability to recall many details of domain-relevant material by relying on long-term memory structures (Chase and Simon 1973; Ericsson and Kintsch 1995). While memory tests are not often used to assess learning from practice, worked examples, or self-explanation in math and science, practice can have a powerful effect on memory span (Ericsson et al. 1980) and the reduction in cognitive load created by worked examples can increase the number of problem details a learner recalls (Sweller 1988). Learners' recall of details from analog problems (Robins and Mayer 1993) has been examined to assess learning from analogical comparison.

Robust knowledge encourages the use of forward-working strategies and domain-specific reasoning strategies when problem solving, resulting in high accuracy, short solution times, and consistent performance on routine problems (Koedinger and Anderson 1990; Simon and Simon 1978). The effects of practice on problem solving have been carefully examined through computational models (Anderson 1993) as well as observation and self-reported descriptions of problem-solving steps (Anzai and Simon 1979; Chen 1999), solution times (Delaney et al. 1998), and self-reported mental effort or cognitive load (Paas 1992). Worked examples have been tested using performance on isomorphic problems (Sweller and Cooper 1985), self-reports of mental effort or cognitive load (Kalyuga et al. 2001; Paas and Van Merriënboer 1994; Paas 1992), a learner's application of subgoals (Catrambone and Holyoak 1990), talk-alouds collected during problem solving (Catrambone 1996), and groupings of solution steps (Catrambone 1996). Problem-solving skills derived from analogical comparison can be assessed through a learner's ability to map between analogs (Kurtz et al. 2001; Novick and Holyoak 1991), the amount of variety in a learner's solution procedures (Rittle-Johnson and Star 2007), and mapping errors between problems (Ross and Kilbane 1997). The effects of self-explanation on problem-solving strategies have been examined through learners' judgments of whether sufficient information is provided (Aleven et al. 2003), accuracy on learning materials (Hausmann and VanLehn 2007), frequency of help requests (Hausmann and VanLehn 2007), and variety of problem-solving strategies (Rittle-Johnson 2006).

The depth and flexibility of robust knowledge supports transfer to new problems with different surface features, as well as transfer across time and contexts (Brown and Kane 1988; Day and Goldstone 2012; Hickey and Pellegrino 2005; Judd 1908; Schwartz et al. 2005). Practice has been well studied through measures of transfer to problems with similar structural features (Anderson 1993; Chen 1999; Sweller 1983) as well as problems with different structural features or goal states (Burns and Vollmeyer 2002). Worked examples have also been assessed using transfer to problems with similar and different structural features (Atkinson et al. 2003; Catrambone 1998; Paas and Van Merriënboer 1994; Paas 1992), problems solved after a delay (Fong and Nisbett 1991; Kalyuga et al. 2001), and problems solved in a different context (Catrambone and Holyoak 1989). Transfer from analogical comparison is primarily assessed through the learner's transfer of a solution or principle to novel problems (Gick and Holyoak 1983; Nokes-Malach et al. 2013; Novick and

Holyoak 1991; Reed 1989). The effects of self-explanation on transfer have been examined by assessing the learner's ability to induce functions of components (Chi et al. 1994), generalize strategies (Crowley and Siegler 1999), solve new problems that differ in structure or domain (Hausmann and VanLehn 2007; Nokes-Malach et al. 2013), and solve problems given after delay (Hausmann and VanLehn 2007; Rittle-Johnson 2006).

In summary, perceptual and memory assessments have been underutilized for measuring robust knowledge acquired through practice, worked examples, and self-explanation. Expertise literature suggests a number of assessments for identifying and characterizing components of robust knowledge, and these assessments could provide greater insights into the specific features of robust knowledge acquired from different instructional techniques.

Implications for Theory

A great deal of past research has examined each instructional technique reviewed here, leading to significant progress in understanding the theoretical and practical underpinnings of these techniques. However, there is much work to be done in developing a comprehensive theory of instruction that incorporates these techniques and makes general claims about the common knowledge features acquired through each one. To address these issues, we adopted a theoretical approach based on the KLI framework and employed expert knowledge as a model for identifying robust knowledge features (Koedinger et al. 2012). Research on expertise suggests that robust knowledge consists of three target attributes: that the knowledge is deep, connected, and coherent. This research also suggests that robust learning results in changes in perception, memory performance, problem solving, and transfer. We identified which of these features were promoted by each technique and the ways they have been assessed. Although our review focused on four instructional techniques that covered a range of learning events (i.e., memory and fluency-building processes, induction and refinement processes, and understanding and sense-making processes), this framework could also be applied for categorizing and reasoning about additional types of instruction, such as structured inquiry or argumentation.

The review has also identified new features of these techniques. One example is the different notions of connectedness: within-problem procedures, within the declarative steps of a problem, between-problem connections, between-domain connections, and connections between new information and prior knowledge. These different types of connections matter for what is learned and what is transferred. For example, connections between procedural steps support the fast, accurate performance typical of well-practiced strategies, while connections between new information and prior knowledge are particularly useful for identifying and addressing misconceptions. Identifying these features may also guide the design of new types of instruction by providing a key set of target features to promote.

While many of the studies reviewed here focused on math and science learning with high school and college students, some examined learning in elementary and middle school classrooms. Rather than classifying studies by the age of participants, we argue based on the KLI framework that examining the knowledge components learners possess at the start of instruction may be a more fruitful approach to understanding how these instructional techniques might lead to different learning outcomes. In other words, while different age groups might demonstrate different outcomes from the same instructional techniques, we believe these effects are driven by differences in knowledge components typical for those age groups. A systematic examination of the knowledge components students possess prior to engaging in each instructional technique could provide greater insights into *when* in the learning process

each technique is most useful. For example, if connecting prior knowledge to new information is a critical component of self-explanation, a learner who lacks relevant prior knowledge components may benefit more from an alternative instructional technique.

Implications for Practice

The instructional techniques we have reviewed suggest several basic practices that make learning and transfer from problem solving more effective and efficient. Across all instructional techniques, carefully selected problems that focus on critical features appear to be best. Scaffolding the learner's comparison across problems or providing instructionally appropriate self-explanation prompts can also help, although the worked example and self-explanation literatures suggest that leaving some gaps in information is more likely to encourage constructive learning practices and, in turn, deeper, more connected knowledge (Nokes et al. 2011; Renkl et al. 2002; Roy and Chi 2005).

Given the robust evidence that self-explanation supports the development of knowledge that is connected and coherent, it is not surprising that self-explanation has been frequently employed in conjunction with other instructional techniques. A number of studies have found promising learning effects by applying self-explanation to worked example study (Atkinson et al. 2003; Nokes-Malach et al. 2013; Renkl 2002; Schworm and Renkl 2007), so much so that Renkl (2005) included prompting or training self-explanation among a core set of principles for structuring effective worked examples. In fact, many of the classic studies examining the effects of self-explanation have used worked examples as the target for learners' explanations (e.g., Chi et al. 1989). Other empirical studies of prompting self-explanation of worked examples have shown mixed results, however, suggesting that combining multiple instructional practices may not have an additive effect (Rittle-Johnson 2006). When employing self-explanations as a way to enhance learning from worked examples, many of the same principles previously described for effective self-explanation apply, such as focusing on principles (Renkl 1997, 2002).

Interleaving worked examples and practice problems has also produced powerful learning effects, with evidence suggesting that students benefit from the scaffolding of worked examples early in the learning process but also the reduced structure of practice problems as their skill improves (Renkl and Atkinson 2003). Analogical comparison often features worked examples as the cases being compared (e.g., Nokes-Malach et al. 2013; Rittle-Johnson et al. 2009). Comparing self-explanation of worked examples, analogical comparison of worked examples, and worked example study with practice problems, Nokes-Malach et al. (2013) found that while the analogical comparison condition performed less well than the others on a near-transfer task, the self-explanation and analogical comparison conditions led to greater far-transfer performance than the worked example and practice condition. Participants' transfer performance suggested that they created connections between abstract principles, prior knowledge, and examples, while the group that read worked examples and completed practice problems acquired knowledge that was tied to features of specific examples.

Future Directions

Although these instructional techniques have been extensively investigated, there are still important phenomena that have not been explored. For example, the effects of practice, worked examples, and analogical comparison on knowledge coherence have not been

thoroughly investigated. Future work should examine whether these techniques provide direct opportunities for learners to recognize and revise misconceptions in their knowledge and, if not, how they might be modified to do so. Our review also suggests that some types of assessment have been underutilized. While perception and memory tests are not typically included in assessments of students' learning from practice, worked examples, and self-explanation, it is likely that these assessments could provide new insights into the depth, connectedness, and coherence of the knowledge students acquire through these techniques. Future work should make use of all types of assessments to create a more complete picture of the ways these instructional techniques impact robust knowledge.

There is extensive work directly comparing some of these techniques, such as worked examples vs. rote practice (e.g., Cooper and Sweller 1987; Sweller and Cooper 1985), analogical comparison vs. sequential study of worked examples (e.g., Rittle-Johnson et al. 2009; Rittle-Johnson and Star 2007; Star and Rittle-Johnson 2009), and self-explanation of examples vs. worked example study (e.g., Atkinson et al. 2003; Renkl 2002). However, one particularly noticeable gap in the literature deals with direct comparison of self-explanation and analogical comparison. Only a handful of studies have compared these two (Gadgil et al. 2012; Nokes-Malach et al. 2013), and further comparison could significantly expand our understanding of the levels of transfer each promotes, the role of prior knowledge, and the effectiveness of each in addressing misconceptions.

These studies mark only the beginning of comparing different combinations of instructional techniques to one another on the same task to test similarities and differences in what is learned. Future work should examine whether different instructional combinations predict the learning and transfer outcomes described here or whether there are interactive effects of combining techniques. Our work demonstrates the utility of incorporating a concrete framework for comparing across techniques. Koedinger et al. (2012) proposed the KLI framework to encourage systematic analysis and testable hypotheses about the vast number of instructional techniques that have emerged from decades of research. Our review adopts the utility of clearly distinguishing and articulating the relationship between knowledge, learning, and instruction, and it demonstrates that such an approach can be applied at different grain sizes (e.g., the knowledge component level of KLI and the knowledge feature level of the current review). Further work examining evidence within instructional techniques and comparing across techniques should maintain such a structured framework, as it highlights the importance of considering learning goals, assessments, and knowledge features when selecting an instructional approach.

Acknowledgments This work was supported by Grant SBE0836012 from the National Science Foundation to the Pittsburgh Science of Learning Center (<http://www.learnlab.org>). We gratefully acknowledge Christian Schunn, Joel Chan, and Jooyoung Jang for their feedback on this work, Jose Mestre and Brian Ross for insights into the ideas discussed, and Daniel Robinson and three anonymous reviewers for their very helpful suggestions and comments on the article.

References

- Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. *Memory & Cognition*, 9(4), 422–433. doi:10.3758/BF03197568.
- Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction*, 17(3), 286–303. doi:10.1016/j.learninstruc.2007.02.004.
- Aleven, V. A. W. M. M., Koedinger, K. R., & Popescu, O. (2003). A tutorial dialog system to support self-explanation: Evaluation and open questions. In *Proceedings of the 11th International Conference on Artificial Intelligence in Education* (pp. 39–46).

- Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: a meta-analytic review. *Educational Psychologist, 48*(2), 87–113. doi:10.1080/00461520.2013.775712.
- Allard, F., & Starkes, J. L. (1991). Motor-skill experts in sports, dance, and other domains. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 126–152). Cambridge: Cambridge University Press.
- Anderson, J. R. (1987). Skill acquisition: compilation of weak-method problem situations. *Psychological Review, 94*(2), 192–210. doi:10.1037/0033-295X.94.2.192.
- Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48*, 35–44. doi:10.1037/0003-066X.48.1.35.
- Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 932–945. doi:10.1037/0278-7393.23.4.932.
- Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review, 86*(2), 124–140. doi:10.1037/0033-295X.86.2.124.
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181–214. doi:10.3102/00346543070002181.
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774–783. doi:10.1037/0022-0663.95.4.774.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: a taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637. doi:10.1037//0033-2909.128.4.612.
- Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., & Bransford, J. D. (1998). Doing with understanding: lessons from research on problem- and project-based learning. *Journal of the Learning Sciences, 7*(3–4), 271–311. doi:10.1080/10508406.1998.9672056.
- Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology, 99*(1), 100–114. doi:10.1037/a0013247.
- Blanchette, I., & Dunbar, K. N. (2000). How analogies are generated: the roles of structural and superficial similarity. *Memory & Cognition, 28*(1), 108–124. doi:10.3758/BF03211580.
- Booth, J. L., & Koedinger, K. R. (2008). Key misconceptions in algebraic problem solving. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 571–576). Austin: Cognitive Science Society.
- 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* (Vol. 24, pp. 61–100). Washington, D.C: American Educational Research Association.
- Bransford, J. D., Sherwood, R., Vye, N. J., & Rieser, J. (1986). Teaching thinking and problem solving: research foundations. *American Psychologist, 41*(10), 1078–1089. doi:10.1037/0003-066X.41.10.1078.
- Brown, D. E., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: abstract transfer versus explanatory model construction. *Instructional Science, 18*(4), 237–261. doi:10.1007/BF00118013.
- Brown, D. E., & Hammer, D. (2008). Conceptual change in physics. In S. Vosniadou (Ed.), *International handbook on research in conceptual change* (pp. 127–154).
- Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: learning to learn and learning from example. *Cognitive Psychology, 20*(4), 493–523. doi:10.1016/0010-0285(88)90014-X.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42. doi:10.2307/1176008.
- Burns, B. D., & Vollmeyer, R. (2002). Goal specificity effects on hypothesis testing in problem solving. *The Quarterly Journal of Experimental Psychology, 55A*(1), 241–261. doi:10.1080/02724980143000262.
- Capon, N., & Kuhn, D. (2004). What's so good about problem-based learning? *Cognition and Instruction, 22*(1), 61–79. doi:10.1207/s1532690Xci2201_3.
- Catrambone, R. (1996). Generalizing solution procedures learned from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(4), 1020–1031. doi:10.1037/0278-7393.22.4.1020.
- Catrambone, R. (1998). The subgoal learning model: creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127*(4), 355–376. doi:10.1037/0096-3445.127.4.355.
- Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(6), 1147–1156. doi:10.1037//0278-7393.15.6.1147.
- Catrambone, R., & Holyoak, K. J. (1990). Learning subgoals and methods for solving probability problems. *Memory & Cognition, 18*(6), 593–603. doi:10.3758/BF03197102.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332. doi:10.1207/s1532690Xci0804.

- Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.
- Chen, Z. (1999). Schema induction in children's analogical problem solving. *Journal of Educational Psychology*, 91(4), 703–715. doi:10.1037/0022-0663.91.4.703.
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology, Vol. 5* (pp. 161–238). Mahwah: Lawrence Erlbaum Associates.
- Chi, M. T. H. (2006). Laboratory methods for assessing experts' and novices' knowledge. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 167–184). Cambridge: Cambridge University Press.
- Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *International handbook on research in conceptual change* (pp. 61–82). Hillsdale: Erlbaum.
- Chi, M. T. H., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology*, 19(1), 29–39. doi:10.1037/0012-1649.19.1.29.
- Chi, M. T. H., & Ohlsson, S. (2005). Complex declarative learning. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 371–399). New York: Cambridge University Press. doi:10.1207/s15327809jls0101_4.
- Chi, M. T. H., & VanLehn, K. A. (1991). The content of physics self-explanations. *Journal of the Learning Sciences*, 1(1), 69–105. doi:10.1207/s15327809jls0101_4.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152. doi:10.1207/s15516709cog0502_2.
- Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hillsdale: Lawrence Erlbaum Associates, Inc.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145–182. doi:10.1207/s15516709cog1302_1.
- Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439–477. doi:10.1207/s15516709cog1803_3.
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science instruction. *Review of Educational Research*, 63(1), 1–49. doi:10.3102/00346543063001001.
- Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science*, 15(1), 89–132. doi:10.1016/0364-0213(91)80014-V.
- Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. *Memory & Cognition*, 19(2), 151–158. doi:10.3758/BF03197111.
- Conati, C., & VanLehn, K. A. (2000). Toward computer-based support of meta-cognitive skills: a computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education*, 11, 389–415.
- Cooper, G. A., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, 79(4), 347–362. doi:10.1037//0022-0663.79.4.347.
- Crowley, K., & Siegler, R. S. (1999). Explanation and generalization in young children's strategy learning. *Child Development*, 70(2), 304–316. doi:10.1111/1467-8624.00023.
- Cummins, D. D. (1992). Role of analogical reasoning in the induction of problem categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 1103–1124. doi:10.1037/0278-7393.18.5.1103.
- Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: connecting findings and theories of transfer of learning. *Educational Psychologist*, 47, 153–176. doi:10.1080/00461520.2012.696438.
- Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy-specific nature of improvement: the power law applies by strategy within task. *Psychological Science*, 9(1), 1–7. doi:10.1111/1467-9280.00001.
- Dellarosa, D. (1985). *Abstraction of problem-type schemata through problem comparison (Tech. Rep. No. 146)*. Boulder.
- Dufresne, R. J., Gerace, W. J., Hardiman, P. T., & Mestre, J. P. (1992). Constraining novices to perform expertlike problem analyses: effects on schema acquisition. *Journal of the Learning Sciences*, 2(3), 307–331. doi:10.1207/s15327809jls0203_3.
- Dunbar, K. N., Fugelsang, J. A., & Stein, C. (2007). Do naive theories ever go away? Using brain and behavior to understand changes in concepts. In P. Shah & M. Lovett (Eds.), *Thinking with data* (pp. 193–206). New York: Erlbaum.
- Ericsson, K. A., & Charness, N. (1994). Expert performance: its structure and acquisition. *American Psychologist*, 49(8), 725–747. doi:10.1037/0003-066X.50.9.803.

- Ericsson, K. A., & Charness, N. (1997). Cognitive and developmental factors in expert performance. In P. J. Feltovich, K. M. Ford, & R. R. Hoffman (Eds.), *Expertise in context: Human and machine* (pp. 3–41). Cambridge: MIT Press.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245. doi:10.1037/0033-295X.102.2.211.
- Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits* (p. 344). New York: Cambridge University Press.
- Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*(4448), 1181–1182. doi:10.1126/science.7375930.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. doi:10.1037//0033-295X.100.3.363.
- Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 243–285). New York: Academic Press.
- Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General*, *120*(1), 34–45. doi:10.1037/0096-3445.120.1.34.
- Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. H. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, *22*(1), 47–61. doi:10.1016/j.learninstruc.2011.06.002.
- Gentner, D. (1983). Structure-mapping: a theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170. doi:10.1016/S0364-0213(83)80009-3.
- Gentner, D. (2002). Analogical reasoning, psychology of. In *Encyclopedia of Cognitive Science*. London: Nature Publishing Group.
- Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*(2–3), 263–297. doi:10.1207/s15516709cog1003_2.
- Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: a general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393–405. doi:10.1037/0022-0663.95.2.393.
- Gentner, D., Loewenstein, J., Thompson, L., & Forbus, K. D. (2009). Reviving inert knowledge: analogical abstraction supports relational retrieval of past events. *Cognitive Science*, *33*(8), 1343–1382. doi:10.1111/j.1551-6709.2009.01070.x.
- Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist*, *21*(1), 99–120. doi:10.1207/s15326985ep2101&2_6.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38. doi:10.1016/0010-0285(83)90002-6.
- Gick, M. L., & Paterson, K. (1992). Do contrasting examples facilitate schema acquisition and analogical transfer? *Canadian Journal of Psychology*, *46*(4), 539–550. doi:10.1037/h0084333.
- Guo, J.-P., Pang, M. F., Yang, L.-Y., & Ding, Y. (2012). Learning from comparing multiple examples: on the dilemma of “similar” or “different”. *Educational Psychology Review*, *24*(2), 251–269. doi:10.1007/s10648-012-9192-0.
- Hardiman, P. T., Dufresne, R. J., & Mestre, J. P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory & Cognition*, *17*(5), 627–638. doi:10.3758/BF03197085.
- Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In *Child development and education in Japan* (pp. 262–272). New York: Freeman. doi:10.1002/ccd.10470.
- Hausmann, R. G. M., & Chi, M. T. H. (2002). Can a computer interface support self-explaining? *Cognitive Technology*, *7*(1), 4–14.
- Hausmann, R. G. M., & VanLehn, K. A. (2007). Explaining self-explaining: A contrast between content and generation. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *13th international conference on artificial intelligence in education* (pp. 417–424). Amsterdam: IOS Press.
- Hickey, D. T., & Pellegrino, J. W. (2005). Theory, level, and function: Three dimensions for understanding transfer and student assessment. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 251–294). Greenwich: Information Age Publishing.
- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale: Lawrence Erlbaum Associates.
- Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction*, *14*(3), 251–283. doi:10.1207/s1532690xc1403_1.
- Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, *15*(4), 332–340. doi:10.3758/BF03197035.
- Judd, C. H. (1908). The relation of special training to general intelligence. *Educational Review*, *36*, 28–42.
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, *93*(3), 579–588. doi:10.1037/0022-0663.93.3.579.

- Kellman, P. J., & Garrigan, P. (2009). Perceptual learning and human expertise. *Physics of Life Reviews*, 6(2), 53–84. doi:10.1016/j.plrev.2008.12.001.
- 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. doi:10.1207/s15326985ep4102_1.
- Koedinger, K. R., & Alevan, V. A. W. M. M. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239–264. doi:10.1007/s10648-007-9049-0.
- Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: elements of expertise in geometry. *Cognitive Science*, 14(4), 511–550. doi:10.1207/s15516709cog1404_2.
- Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, 36(5), 757–798. doi:10.1111/j.1551-6709.2012.01245.x.
- Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, 342(22), 935–937. doi:10.1126/science.1238056.
- Kolodner, J. L. (1997). Educational implications of analogy: a view from case-based reasoning. *American Psychologist*, 52(1), 57–66. doi:10.1037/0003-066X.52.1.57.
- Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences*, 10(4), 417–446. doi:10.1207/S15327809JLS1004new_2.
- Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208(4450), 1335–1342. doi:10.1126/science.208.4450.1335.
- Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin Review*, 6(4), 586–597. doi:10.3758/BF03212967.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95(4), 492–527. doi:10.1037/0033-295X.95.4.492.
- Luchins, A. S. (1942). Mechanization in problem solving: the effect of Einstellung. *Psychological Monographs*, 54(6), 1–95. doi:10.1037/h0093502.
- Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition*, 28(1), 41–50. doi:10.3758/BF03211575.
- Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, 25(4), 431–467. doi:10.1006/cogp.1993.1011.
- McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, 13(3), 307–325. doi:10.1016/0010-0285(81)90012-8.
- McLaren, B. M., Lim, S.-J., & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 2176–2181). Austin: Cognitive Science Society.
- Neves, D. M., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In *Cognitive skills and their acquisition* (pp. 57–84). Hillsdale: Erlbaum.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale: Erlbaum.
- Nokes, T. J., & Ohlsson, S. (2005). Comparing multiple paths to mastery: what is learned? *Cognitive Science*, 29(5), 769–796. doi:10.1207/s15516709cog0000_32.
- Nokes-Malach, T. J., & Mestre, J. (2013). Toward a model of transfer as sense-making. *Educational Psychologist*, 48(3), 184–207. doi:10.1080/00461520.2013.807556.
- Nokes, T. J., Schunn, C. D., & Chi, M. T. H. (2010). Problem solving and human expertise. In E. Peterson, E. Baker, & B. McGraw (Eds.), *International encyclopedia of education, Volume 5* (Vol. 5, pp. 265–272). Oxford: Elsevier.
- Nokes, T. J., Hausmann, R. G. M., VanLehn, K. A., & Gershman, S. (2011). Testing the instructional fit hypothesis: the case of self-explanation prompts. *Instructional Science*, 39(5), 645–666. doi:10.1007/s11251-010-9151-4.
- Nokes-Malach, T. J., VanLehn, K. A., Belenky, D. M., Lichtenstein, M., & Cox, G. (2013). Coordinating principles and examples through analogy and self-explanation. *European Journal of Psychology of Education*, 28(4), 1237–1263. doi:10.1007/s10212-012-0164-z.
- Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 510–520. doi:10.1037/0278-7393.14.3.510.
- Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(3), 398–415. doi:10.1037/0278-7393.17.3.398.
- Ohlsson, S. (1993). Abstract schemas. *Educational Psychologist*, 28(1), 51–66. doi:10.1207/s15326985ep2801.

- Ohlsson, S., & Rees, E. (1991). The function of conceptual understanding in the learning of arithmetic procedures. *Cognition and Instruction*, 8(2), 103–179. doi:10.1207/s1532690xci0802_1.
- Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology*, 77(3), 272–284. doi:10.1037/0022-0663.77.3.272.
- Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429–434. doi:10.1037/0022-0663.84.4.429.
- Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology*, 86(1), 122–133. doi:10.1037/0022-0663.86.1.122.
- Phye, G. D. (1990). Inductive problem solving: schema inducement and memory-based transfer. *Journal of Educational Psychology*, 82(4), 826–831. doi:10.1037/0022-0663.82.4.826.
- Phye, G. D. (2001). Problem-solving instruction and problem-solving transfer: the correspondence issue. *Journal of Educational Psychology*, 93(3), 571–578. doi:10.1037//0022-0663.93.3.571.
- Pirolli, P. L., & Anderson, J. R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology*, 39(2), 240–272. doi:10.1037/h0080061.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: toward a theory of conceptual change. *Science Education*, 66(2), 211–227. doi:10.1002/sce.3730660207.
- Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(4), 811–828. doi:10.1037/0278-7393.26.4.811.
- Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology*, 16(3), 325–342. doi:10.1002/acp.796.
- Reed, S. K. (1989). Constraints on the abstraction of solutions. *Journal of Educational Psychology*, 81(4), 532–540. doi:10.1037/0022-0663.81.4.532.
- Reed, S. K., Ackinlose, C. C., & Voss, A. A. (1990). Selecting analogous problems: similarity versus inclusiveness. *Memory & Cognition*, 18(1), 83–98. doi:10.3758/BF03202649.
- Reeves, L. M., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin*, 115(3), 381–400. doi:10.1037/0033-2909.115.3.381.
- Renkl, A. (1997). Learning from worked-out examples: a study on individual differences. *Cognitive Science*, 21(1), 1–29. doi:10.1207/s15516709cog2101_1.
- Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction*, 12(5), 529–556. doi:10.1016/S0959-4752(01)00030-5.
- Renkl, A. (2005). The worked-out-example principle in multimedia learning. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 229–246). New York: Cambridge University Press.
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38(1), 1–37. doi:10.1111/cogs.12086.
- Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: a cognitive load perspective. *Educational Psychologist*, 38(1), 15–22. doi:10.1207/S15326985EP3801_3.
- Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: smooth transitions help learning. *The Journal of Experimental Education*, 70(4), 293–315. doi:10.1080/00220970209599510.
- Richey, J. E., & Nokes-Malach, T. J. (2013). How much is too much? Learning and motivation effects of adding instructional explanations to worked examples. *Learning and Instruction*, 25, 104–124. doi:10.1016/j.learninstruc.2012.11.006.
- Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science*, 316, 1128–1129. doi:10.1126/science.1142103.
- Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the conceptual structure of mathematics. *Educational Psychologist*, 47(3), 189–203. doi:10.1080/00461520.2012.667065.
- Ringenberg, M., & VanLehn, K. A. (2006). Scaffolding problem solving with annotated, worked-out examples to promote deep learning. In M. Ikeda, K. D. Ashley, & T.-W. Chan (Eds.), *Proceedings of the 8th international conference on intelligent tutoring systems* (Vol. 4053, pp. 624–634). Berlin: Springer. doi:10.1007/11774303.
- Rittle-Johnson, B. (2006). Promoting transfer: effects of self-explanation and direct instruction. *Child Development*, 77(1), 1–15. doi:10.1111/j.1467-8624.2006.00852.x.
- Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, 99(3), 561–574. doi:10.1037/0022-0663.99.3.561.
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: an iterative process. *Journal of Educational Psychology*, 93(2), 346–362. doi:10.1037/0022-0663.93.2.346.

- Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, *101*(4), 836–852. doi:10.1037/a0016026.
- Robins, S., & Mayer, R. E. (1993). Schema training in analogical reasoning. *Journal of Educational Psychology*, *85*(3), 529–538. doi:10.1037/0022-0663.85.3.529.
- Ross, B. H., & Kilbane, M. C. (1997). Effects of principle explanation and superficial similarity on analogical mapping in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 427–440. doi:10.1037//0278-7393.23.2.427.
- Roy, M., & Chi, M. T. H. (2005). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 271–286). New York: Cambridge University Press.
- Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: rethinking mechanism of a neglected phenomenon. *Educational Psychologist*, *24*(2), 113–142. doi:10.1207/s15326985ep2402_1.
- Scheiter, K., & Gerjets, P. (2006). When less is sometimes more: Optimal learning conditions are required for schema acquisition from multiple examples. In *Proceedings of the 27th annual conference of the cognitive science society* (pp. 1943–1948). Mahwah: Erlbaum.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993). On acquiring expertise in medicine. *Educational Psychology Review*, *5*(3), 205–221. doi:10.1007/BF01323044.
- Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning Memory and Cognition*, *8*(5), 484–494. doi:10.1037/0278-7393.8.5.484.
- Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science*, *23*(3), 337–370. doi:10.1207/s15516709cog2303_3.
- Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–51). Greenwich: Information Age Publishing.
- Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: when instructional explanations reduce self-explanations. *Computers & Education*, *46*(4), 426–445. doi:10.1016/j.compedu.2004.08.011.
- Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, *99*(2), 285–296. doi:10.1037/0022-0663.99.2.285.
- Sherman, J., & Bisanz, J. (2009). Equivalence in symbolic and nonsymbolic contexts: benefits of solving problems with manipulatives. *Journal of Educational Psychology*, *101*(1), 88–100. doi:10.1037/a0013156.
- Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 325–348). Hillsdale: Lawrence Erlbaum Associates.
- Singley, M. K., & Anderson, J. R. (1985). The transfer of text-editing skill. *International Journal of Man-Machine Studies*, *22*(4), 403–423. doi:10.1016/S0020-7373(85)80047-X.
- Singley, M. K., & Anderson, J. R. (1989). *The transfer of a cognitive skill*. Cambridge: Harvard University Press.
- Star, J. R., & Rittle-Johnson, B. (2009). It pays to compare: an experimental study on computational estimation. *Journal of Experimental Child Psychology*, *102*(4), 408–426. doi:10.1016/j.jecp.2008.11.004.
- Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: what makes an expert student? *Instructional Science*, *26*, 127–140. doi:10.1023/A:1003096215103.
- Sweller, J. (1983). Control mechanisms in problem solving. *Memory & Cognition*, *11*(1), 32–40. doi:10.3758/BF03197659.
- Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science*, *12*(2), 257–285. doi:10.1207/s15516709cog1202_4.
- Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*(1), 59–89. doi:10.1207/s1532690xci0201_3.
- Sweller, J., & Levine, M. (1982). Effects of goal specificity on means-ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(5), 463–474. doi:10.1037/0278-7393.8.5.463.
- Sweller, J., Mawer, R. F., & Howe, W. (1982). Consequences of history-cued and means-end strategies in problem solving. *The American Journal of Psychology*, *95*(3), 455. doi:10.2307/1422136.
- Sweller, J., Mawer, R. F., & Ward, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General*, *112*(4), 639–661. doi:10.1037//0096-3445.112.4.639.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. doi:10.1023/A:1022193728205.
- Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say “broke”? A model of learning the past tense without feedback. *Cognition*, *86*(2), 123–155.
- Taatgen, N. A., & Lee, F. J. (2003). Production compilation: a simple mechanism to model complex skill acquisition. *Human Factors*, *45*(1), 61–76. doi:10.1518/hfes.45.1.61.27224.

- Thorndyke, P. W. (1984). Applications to schema theory in cognitive research. In J. A. Anderson & S. M. Kosslyn (Eds.), *Tutorials in learning and memory* (pp. 167–191). San Francisco: Freeman.
- VanLehn, K. A. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, 47(1), 513–539. doi:10.1146/annurev.psych.47.1.513.
- VanLehn, K. A. (1999). Rule-learning events in the acquisition of a complex skill: an evaluation of Cascade. *Journal of the Learning Sciences*. doi:10.1207/s15327809jls0801_3.
- VanLehn, K. A., & Jones, R. M. (1993). What mediates the self-explanation effect? Knowledge gaps, schemas or analogies? In M. Polson (Ed.), *Proceedings of the fifteenth annual conference of the cognitive science society* (pp. 1034–1039). Hillsdale: Erlbaum.
- VanLehn, K. A., & van de Sande, B. (2009). Acquiring conceptual expertise from modeling: The case of elementary physics. In K. A. Ericsson (Ed.), *Development of professional performance: Toward measurement of expert performance and design of optimal learning environments* (pp. 356–378). Cambridge: Cambridge University Press.
- Voss, J. F., Tyler, S. W., & Yengo, L. A. (1983). Individual differences in the solving of social science problems. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition, Vol. 1*. New York: Academic P.
- Ward, M. R., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, 7(1), 1–39. doi:10.1207/s1532690xci0701_1.
- Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review*, 22(4), 393–409. doi:10.1007/s10648-010-9136-5.
- Zeitz, C. M. (1994). Expert-novice differences in memory, abstraction, and reasoning in the domain of literature. *Cognition and Instruction*, 12(4), 277–312. doi:10.1207/s1532690xci1204_1.