37 Multimedia Learning with Cognitive Tutors

KENNETH KOEDINGER AND VINCENT ALEVEN

COGNITIVE TUTORS AS VEHICLES FOR ADVANCING LEARNING THEORY AND PRACTICE

Besides being highly effective computer-based learning environments (Koedinger, Anderson, Hadley, & Mark, 1997; Pane, Griffin, McCaffrey, & Karam, 2014; Ritter, Anderson, Koedinger, & Corbett, 2007), Cognitive Tutors are a vehicle for advancing theories of learning and instruction. Whereas many multimedia principles (e.g., Mayer, 2021) focus on up-front declarative instruction, Cognitive Tutors are especially relevant for elaborating and testing theories of learning-by-doing, including variations such as active learning (Deslauriers, McCarty, Miller, Callaghan, & Kestin, 2019), deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993), testing effect (McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011), and some interpretations of constructivism (cf., Chi & Wylie, 2014; Deslauriers, Schelew, & Wieman, 2011). Like other educational technologies, they are ideal for testing instructional principles because instructional design manipulations are precisely defined and systematically delivered. Further, Cognitive Tutors are not just a place to embody and test domain-general learning principles, but also imply a set of data- and theory-driven methods both for developing cognitive models of the domain-specific learning goals and for appropriate selection and adaptation of domain-general principles. In this chapter we focus on learning theory advances that have been facilitated by Cognitive Tutors in the past 14 years since we wrote a related review (Koedinger & Aleven, 2007).

Cognitive Tutors are a kind of AI-based learning environment used widely in US K-12 mathematics education as an integrated part of a blended instructional approach that also includes a range of classroom activities. Cognitive Tutors support learning-by-doing by posing challenging, elaborate problems, sometimes with multiple representations, and by supporting students with two forms of AI-based personalized guidance (see Figure 37.1). First, within problems, they provide step-level hints and feedback that adapt to the student's solution path and errors (see Figure 37.2). In particular, Cognitive Tutors are designed to guide students through problems with a range of possible solution paths; they allow students freedom regarding which path to take and adjust their hints and feedback to the student choice. For example, the tutor's hints (given at a student's request) will always suggest a next step consistent with the student's solution path.

As a second form of adaptive guidance, Cognitive Tutors support personalized mastery learning, as a way of adapting to students' knowledge growth. In this approach, the target knowledge for each unit in the tutor curriculum is viewed as a related set of knowledge components (KCs), where a KC as defined by Koedinger, Corbett, and Perfetti (2012, p. 764), is "an acquired unit of cognitive function or structure that can be inferred from performance on a set of related tasks." Each KC roughly captures the requisite knowledge needed for a single step in a tutor problem. The tutoring system continuously assesses each student's mastery of the KCs targeted in the instruction (see the "Skill Meter," bottom right, Figure 37.1), based on their performance in the tutor, using a statistical or Bayesian model (cf., Corbett & Anderson, 1995). Based on this assessment, the tutor tailors the selection of the next problem for the student to work on, until the student masters all targeted KCs.

Both forms of adaptive instruction have been empirically shown to contribute to the effectiveness of Cognitive Tutors in helping students achieve better learning outcomes (Anderson, Conrad, & Corbett, 1989; Corbett & Anderson, 2001; Corbett, McLaughlin, & Scarpinatto, 2000). We note that step-level guidance in problem-solving activities is characteristic in intelligent tutoring systems and helps contribute to their effectiveness (VanLehn, 2011, 2016). Such guidance tends to be absent in other educational technologies, which can lead to simple practice problems and limited effectiveness.

From the start, Cognitive Tutors have been grounded in cognitive theory, cognitive task analysis, and cognitive modeling, in a manner that supports the software's adaptive capabilities. Cognitive Tutors have contributed to cognitive theory, for example, by helping to solidify the use of production rules as formalism for modeling

https://doi.org/10.1017/9781108894333.046 Published online by Cambridge University Press

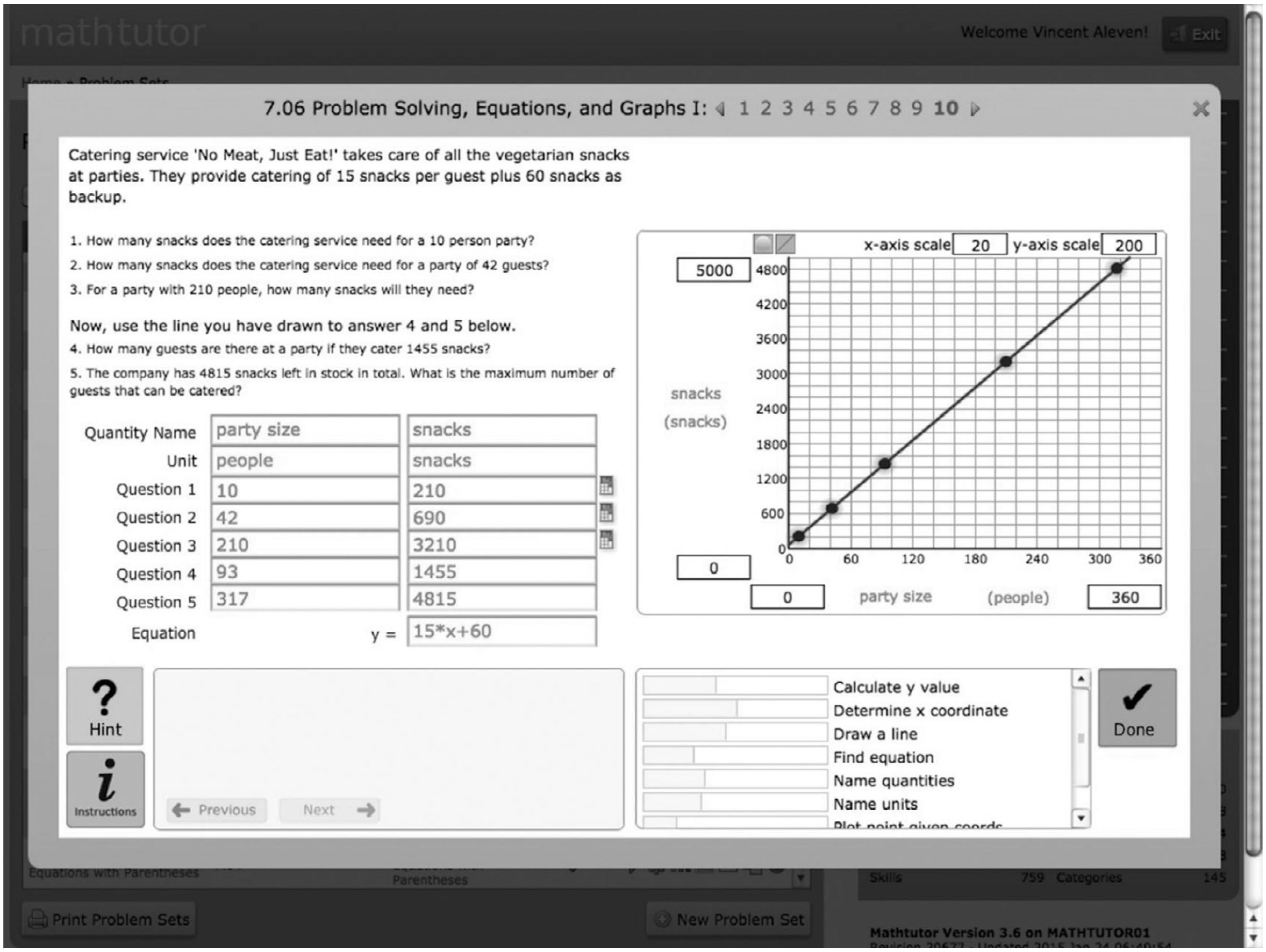

Figure 37.1 Solved problem in Mathtutor, a web-based system in the Cognitive Tutor family. Students enter numbers and expressions in the cells and plot points in the graph with each entry receiving feedback.

procedural knowledge (Anderson, Corbett, Koedinger, & Pelletier, 1995) and its acquisition (Li, Matsuda, Cohen, & Koedinger, 2015). Cognitive Tutors have also long served as a vehicle for in vivo experiments to identify features of effective AI-supported learning by doing (Koedinger & Aleven, 2007). A set of dedicated authoring tools, with which non-programmers can create Cognitive Tutors (e.g., educational researchers, learning engineers, instructional designers, and instructors), has substantially facilitated this line of research (Aleven, McLaren, Sewall et al., 2016).

Contrary to some impressions, teachers matter (see Figure 37.3) and always have (e.g., see Koedinger & Aleven, 2016; Koedinger, Anderson, Hadley, & Mark, 1997). Recent advances have made this link even stronger. From the very beginning, teachers were involved in the design of Cognitive Tutor software as an integrated part of curricula for middle-school and high-school mathematics (Koedinger & Aleven, 2016; Koedinger, Anderson, Hadley, & Mark, 1997; Koedinger & Corbett, 2006). In these *blended* curricula, teachers are constantly involved. They guide students as they use Cognitive Tutors (40 percent of the time) and lead, orchestrate, and oversee classroom activities the remaining 60 percent of the time. When Cognitive Tutors started to be used in classrooms, it soon became apparent that they could profoundly change the classroom culture, as well as the way teachers interact with their students (Schofield, 1995). On balance, this change was a very positive one, freeing teachers to engage students in more personalized interactions (see Figure 37.3, top left). At the same time, it was challenging for teachers to monitor a whole class of students working with tutoring software and to identify the students who most need extra help.

A recent project by Holstein, McLaren, and Aleven (2019) showed that newly-designed teacher smart glasses, which displayed real-time analytics from the tutoring software, can help address this challenge (Holstein, McLaren, & Aleven, 2018; see Figure 37.3, bottom left). When teachers used the smart glasses, it resulted in them spending more time with students who had more to learn, based on a pre-test. As a result, students learned more, in particular those who had more to learn. The tool thus served as an equalizing force. The study suggests that even though a teacher has limited time for each student, a little

https://doi.org/10.1017/9781108894333.046 Published online by Cambridge University Press

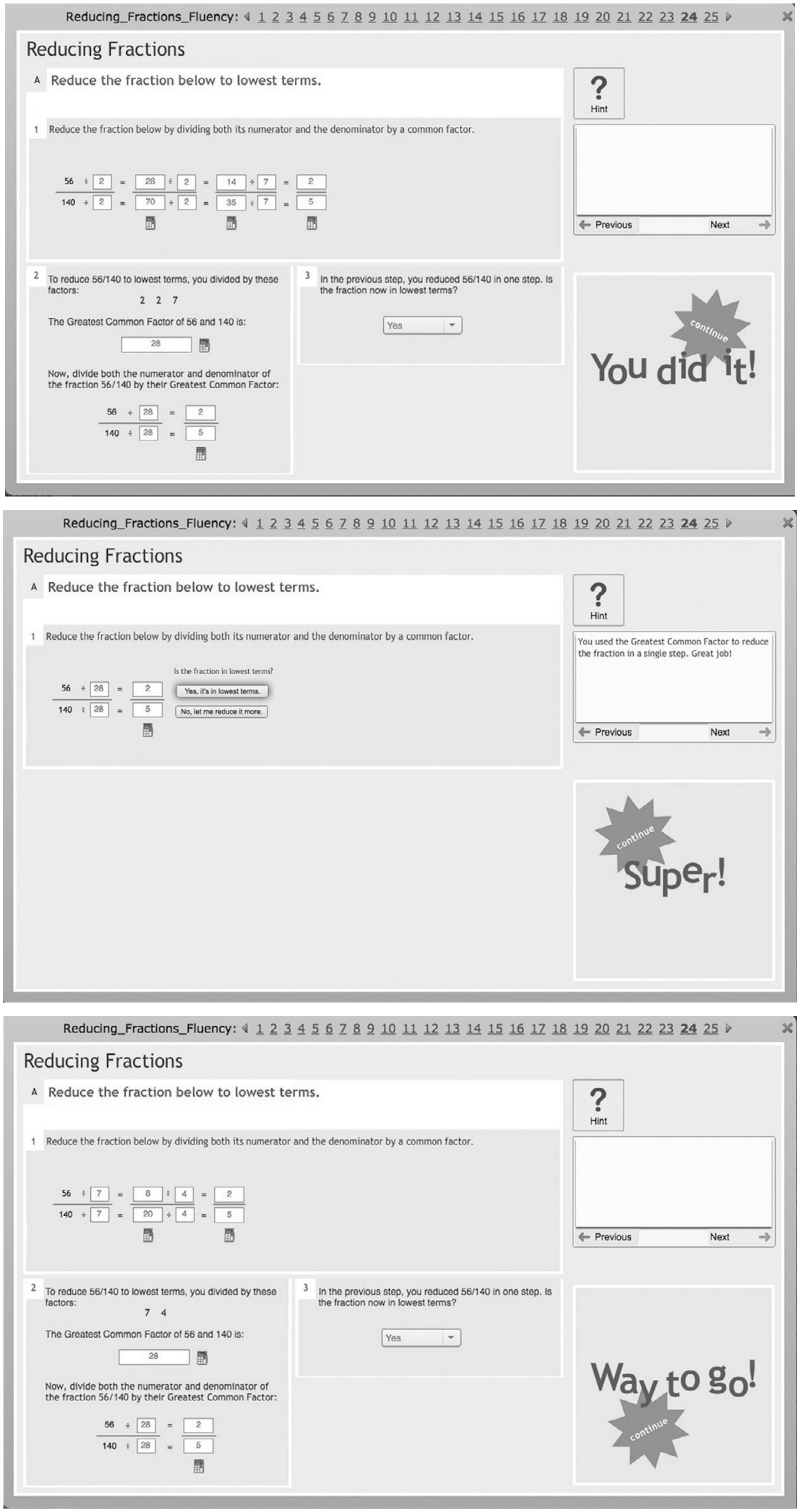

Figure 37.2 Three different solutions to the same fraction reduction problem in a Cognitive Tutor for fractions learning (Doroudi, Holstein, Aleven, & Brunskill, 2015).

Figure 37.3 Views of a Cognitive Tutor classroom, illustrating the active role of teachers and the role of (often spontaneous) collaboration that occurs among students working individually

bit of teacher help goes a long way. Well-designed technology – teachers were involved in the design of the smart glasses from start to finish – can help make sure this help is timely and targeted at students who really need it, while not ignoring other students. Building on this research, recent work is developing richer, systematic conceptions of how AI, teachers, and students can augment each other (Holstein, Aleven, & Rummel, 2020).

COGNITIVE TUTORS: ADVANCING LEARNING-BY-DOING PRINCIPLES

We pursue a synthesis/integration of prior Cognitive Tutor principles (Anderson, Corbett, Koedinger, & Pelletier, 1995) and principles of effective learning-by-doing, especially deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993), characterized by "carefully tailored practice tasks," "feedback that provides knowledge of results," and "high levels of repetition"(Kellogg & Whiteford, 2009, p. 254). In general, we find overlapping concepts addressed in usefully different ways that yield opportunities for greater clarification and better guidance for learning engineers. For example, the Cognitive Tutor principle to design instruction based on a production rule model of the task domain is highly similar to the deliberate practice notion of well-tailored tasks, as elaborated in the Cognitive Tutors Support Deliberate Practice and More section.

In this section, we briefly overview the original Cognitive Tutor principles and some additions based on recent research. We then discuss how Cognitive Tutors address the principles of deliberate practice, how these connections have been deepened in recent work on Cognitive Tutors, and how looking at deliberate practice from the viewpoint of Cognitive Tutors affords elaborations and added nuance to the principles of deliberate practice. As a nutshell summary, we suggest that effective learning requires **doing**, and more specifically, that effective learning requires *repeated* practice on *well-tailored tasks* in **varied** contexts with **explanatory** *feedback* and **as-needed instruction**. The italicized elements (i.e., repeated practice, well-tailored tasks, and feedback) come from the literature on deliberate practice and correspond with key elements of the Cognitive Tutor principles derived from the ACT-R theory of cognition (Anderson, Corbett, Koedinger, & Pelletier, 1995) and consistent with other theories of learning units of knowledge, such as constraints in constraint-based modeling (Ohlsson, 1994). The bold elements (i.e., varied contexts, explanatory feedback, and as-needed instruction) reflect elaborations that are drawn from the broader empirical research literature, particularly variation and explanatory feedback, or from Cognitive Tutor experience, particularly as-needed instruction.

Prior History of Cognitive Tutor Principles

There have been repeated attempts to articulate and refine general principles from research contributing to the design of Cognitive Tutors and from research within Cognitive Tutors. Anderson, Corbett, Koedinger, and Pelletier (1995) provide one of the early attempts to articulate general principles in the design of Cognitive Tutors and provide evidence, for example, for design based on a detailed analysis of the cognitive processes that are needed for a student to become an expert.

Other papers have elaborated or extended the eight principles presented in Anderson, Corbett, Koedinger, and Pelletier (1995), including Koedinger (2002) and Koedinger and Aleven (2007). One of these eight that has received particular attention in recent years is to "represent student competence as a production set" (Anderson, Corbett, Koedinger, & Pelletier, 1995, p. 179). It is worth highlighting that this *principle* is really a *learning engineering method* rather than a learning science domain-general principle. This method entails cognitive task analysis (Baker, Corbett, & Koedinger, 2007; Clark, Feldon, van Merriënboer, Yates, & Early, 2007; Lovett, 1998; Martin, Mitrovic, Mathan, & Koedinger, 2011) of the particulars of the cognitive structure underlying the demands of the subject-matter domain, ideally represented in production rules and declarative chunks. The recent attention to it has sometimes been more conceptual than literal. Following the spirit of this principle, some efforts have pursued more precise and empirically supported expressions of the

https://doi.org/10.1017/9781108894333.046 Published online by Cambridge University Press

components of desired student knowledge without pursuing, as a practical matter, a full production system implementation. We describe both kinds of efforts while drawing a connection with the deliberate practice notion of well-tailored tasks. Through this connection and others, the chapter provides a first-time explicit linkage between cognitive tutor and deliberate practice principles.

Cognitive Tutors Support Deliberate Practice and More

Learning-by-doing (and its various terminological/conceptual siblings: deliberate practice, active learning, testing effect, retrieval practice) is critical to most substantial learning, and Cognitive Tutors implement its key instructional features. As Kellogg and Whiteford (2009, p. 254) indicate, three key instructional features of deliberate practice are "carefully tailored practice tasks," "feedback that provides knowledge of results," and "high levels of repetition." All three have been a core part of Cognitive Tutors, though sometimes expressed in different terms, and have been explored in different practical and research contexts. At the same time, recent research on Cognitive Tutors leads us to elaborate on and add nuance to these principles.

Cognitive Tutors provide *feedback* as a form of adaptive *step-level guidance* and a variety of nuanced questions about the timing of feedback have been explored in experimental studies. While immediate correctness feedback after each problem-solving step has been demonstrated to be effective and highly efficient (Corbett & Anderson, 2001), some situations have been identified where more delayed feedback produces better learning outcomes (Mathan & Koedinger, 2005; Schooler & Anderson, 1990).

Cognitive Tutors provide high levels of *repetition* and three forms of *task tailoring*. First, *personalized mastery learning* provides *tailored repetition* as it selects tasks tailored to individual students' needs, and repeats until mastery is achieved. Second, adaptive step-level guidance provides another form of task tailoring as it adapts feedback and *as-needed instruction* (in student-requested hints) to the particular errors or problem-solving strategies individual students pursue.

A third form of task tailoring is newer in the Cognitive Tutor literature and is worth particular emphasis. This tailoring happens at a longer time scale, in the design loop (Aleven, McLaughlin, Glenn, & Koedinger, 2017) between instances of running a given course. It involves the creation of tasks that are tailored to the learning needs of all students as they are driven by the structure of the task environment. A line of research has developed more efficient and objective *quantitative* cognitive task analysis methods (e.g., Koedinger & McLaughlin, 2016; Liu & Koedinger, 2017) that supplement existing qualitative cognitive task analysis methods, like structured interviews (Clark, Feldon, van Merriënboer, Yates, & Early, 2007) and think aloud protocols (Ericsson & Simon, 1985). In this approach, analysis of log data from a tutoring system (which captures detailed records of the student–tutor interactions) yields a refined, more empirically-accurate KC model, often with domain-specific discoveries of *hidden skills* – that is, non-obvious but critical cognitive processes that are challenging for novices and are often acquired implicitly and thus not recognized by experts. As such, these hidden skills are typically not well addressed in existing instruction. When hidden skills are discovered, multiple instructional design moves can follow, corresponding to all three types of task tailoring. Koedinger, McLaughlin, and Stamper (2012) and Liu, Koedinger, and McLaughlin (2014) use hidden skill discoveries to redesign step-level guidance and personalized mastery learning and experimentally demonstrate improved student learning outcomes. Other studies use hidden skill discoveries to design new tasks (Huang, Aleven, McLaughlin, & Koedinger, 2020; Koedinger, Booth, & Klahr, 2013; Koedinger & McLaughlin, 2010).

A number of close-the-loop studies confirm the value of redesign based on hidden skill discoveries as a key form of design loop adaptivity. For example, in algebra symbolization, a tutor version that provided focused support to students for composing multi-operator expressions from single-operator ones (a hidden skill detected in data) was found to be more effective than the original tutor, which did not support this hidden skill (Koedinger & McLaughlin, 2010). In geometry area problems with composite shapes, a tutor version that provided explicit support to students for decomposing a composite shape into constituent parts helped students learn better than the original tutor (Koedinger, Booth, & Klahr, 2013). In geometry area problems, backward application of an area formula was found to be harder, in offline data analysis, than forward application when the inverted area formula involved a challenging operator such as a square root. A redesigned tutor that gave practice in backwards application of circle area (which has the challenging operator) but not of other area formulas (which do not involve a challenging operator) yielded greater student learning (Liu & Koedinger, 2017). These experiments led us to an increased appreciation for the power of *design loop adaptation* (Aleven, McLaughlin, Glenn, & Koedinger, 2017). Design loop adaptivity is a form of adaptation *to all students* and thus, likely has broader impact than the more frequently pursued forms of technology-enhanced within-course adaptations to individual students.

Misconceptions about hidden skill discoveries and KC decomposition include that they are always seeking finer-grained distinctions, are divorced from conceptual understanding, and do not address the big picture of learning. To be sure, hidden skill discoveries are to be distinguished from other non-cognitive forms of isolated instruction that overemphasize facts and normative observable (rather than empirically derived latent) components of knowledge (Fries, Son, Givvin, & Stigler, 2020). Instead, hidden skills are often deeper, more integrative knowledge components (Koedinger, Booth, & Klahr, 2013) that are crucial to connecting fundamental knowledge components. An

https://doi.org/10.1017/9781108894333.046 Published online by Cambridge University Press

important nuance here is that, although recommendations to practice the target job task imply whole-task practice, achieving desirable transfer-appropriate processing of hidden KCs may better be achieved with focused part-tasks. While potential part-tasks may bear little surface resemblance to the whole job task, they can produce more transfer to it than practice on similar tasks (Koedinger & McLaughlin, 2010; Yannier, Hudson, & Koedinger, 2020).

The learning benefits of *varied* examples and practice tasks is a well-explored topic (e.g., Carvalho & Goldstone, 2019; Paas & van Merriënboer, 1994; Rohrer & Taylor, 2007). Cognitive Tutor curriculum design achieves variation in the design and selection of tasks within each tutor unit on a particular topic, for example, area of geometric shapes. A typical approach is to introduce tasks corresponding with the knowledge components in that topic area such that, for example, a unit may start with a couple of tasks to find a rectangle area, then move to a couple of tasks to find a triangle area, etc. These tasks vary in terms of the underlying knowledge components being addressed (e.g., rectangle versus triangle area) and also in terms of the problem givens, including different numbers and problem contexts (e.g., find the area of a parking lot or a football field).

Another form of variation occurs as problems increase in complexity, for example, ones where an irregular shape, like the front of a house that needs to be painted, needs to be found by determining the right combination of formulas, like adding the triangular roof to the rectangular base and subtracting out the square windows. These more complex tasks have been a natural hunting ground for hidden skill discoveries that have led to effective well-tailored task design (Koedinger, Booth, & Klahr, 2013). The incremental addition of complexity in unit design is consistent with the cognitive tutor principle of "Minimize working memory load that is extraneous to learning" (#5 in Anderson, Corbett, Koedinger, & Pelletier, 1995). It also leads to increased interleaved practice as practice of different knowledge components is naturally intermixed, both across and within problem-solving tasks. Interleaved practice as a particular way to provide varied contexts is a principle that has been explored explicitly in the context of Cognitive Tutors (Patel, 2017; Patel, Liu, and Koedinger 2016) and has been theoretically confirmed through computational learning experiments (Li, Cohen, & Koedinger, 2013; MacLellan, 2017).

The topic of instructional *feedback* has been explored in so many studies there are now at least 23 meta-analyses (Hattie, 2008, p. 173). This literature contrasts at least three types of feedback: correctness feedback, the knowledge of results feedback emphasized in deliberate practice, and explanatory feedback. The first two are always present in Cognitive Tutors and the third is commonly provided, as per recommendations for hint writing (e.g., Aleven, McLaren, Roll, & Koedinger, 2016), but it is not enforced. Knowledge of results feedback is implemented in Cognitive Tutors at every step in a problem-solving task by guaranteeing students either come to a correct response on their own, perhaps with some high-level hinting, or that step is given as an example, through a so-called bottom-out hint that provides the correct response (cf., Shih, Koedinger, & Scheines, 2008). To be sure, when students are assisted, the personalized cognitive mastery algorithm will ensure that there is a future repetition of the target competence whereby students can demonstrate their mastery without assistance. *Explanatory* feedback goes a step further by providing an explanation for why a correct response is appropriate or an incorrect response is inappropriate (c.f., van der Kleij, Feskens, & Eggen, 2015). Such explanations are typical in high-level hint messages in Cognitive Tutors and in some feedback messages.

As learning-by-doing approaches, both the deliberate practice and Cognitive Tutor literatures do not put much emphasis on declarative instruction. It is reasonable that up-front teacher-led instruction is an assumed element both in the expertise development contexts that have been the main focus of deliberate practice research and in the classroom contexts that have been the main focus of Cognitive Tutor research. To be sure, instruction is addressed in both approaches through feedback during practice. In addition, Cognitive Tutors provide *as-needed instruction* in the form of next-step hints. Cognitive Tutor research on hints has been nicely summarized by Aleven, McLaren, Roll, and Koedinger (2016, p. 210) as "help helps, but only so much." That is, evidence from a variety of study types showed that explanatory on-demand, next-step hints tend to have a modest effect on student learning, although they noted that no studies were ideally designed to rigorously test that proposition. A data-mining study by Roll, Baker, Aleven, and Koedinger (2014) raises an interesting question about the timing of as-needed instruction in the context of deliberate practice. They found that on-demand next-step help, with a rationale, is not effective when students have low mastery of the knowledge component involved (compared to students' just trying the step), suggesting a need to make tutor hints more sensitive to student skill mastery.

In addition to work on explanations offered by the tutoring system, a range of studies with Cognitive Tutors have found advantages of supporting students in simple, computer-supported forms of *self-explaining* their problem-solving steps, often with feedback on correctness. For example, studies showed advantages of explaining steps in geometry problems by referencing a geometry principle that justifies the step (Aleven & Koedinger, 2002), explaining *how* geometry principles apply by highlighting relevant geometry objects in an interactive diagram (Butcher & Aleven, 2013), and explaining relations (either in verbal or diagrammatic form) between graphical and symbolic representations in both fractions learning (Rau, Aleven, & Rummel, 2015), and equation solving (Nagashima, Bartel, Silla, Vest, Alibali, & Aleven, 2020). These studies further support benefits of as-needed instruction (we view computer-supported self-explanation as such) in the context of deliberate practice. It would be

https://doi.org/10.1017/9781108894333.046 Published online by Cambridge University Press

interesting to know if Roll, Baker, Aleven, and Koedinger's (2014) finding – that on-demand explanatory hints are helpful primarily when a student has a medium level of skill mastery – might extend to self-explanation.

IMPLICATIONS FOR THEORY AND PRACTICE

Implications for Cognitive Theory

Most theorizing around multimedia learning focuses on purported domain-general principles of learning and instruction. Research on Cognitive Tutors (and related interactive online learning environments) has led us to the importance of both developing domain-specific theories of learning (cf., Aleven & Koedinger, 2013) and developing theories of how domain-specific characteristics drive the selection and application of domain-general principles (cf., Koedinger, Corbett, & Perfetti, 2012). Developing *well-tailored tasks* requires a domain-specific theory of the latent knowledge demands of the domain. In addition to the data-driven design loop methods we have successfully employed, we also see progress in using computational architectures of domain-general learning to automatically generate domain-specific theories. SimStudent (Li, Matsuda, Cohen, and Koedinger, 2015) and Apprentice Learner (MacLellan, Harpstead, Patel, & Koedinger, 2016; Weitekamp, Harpstead, & Koedinger, 2020) have been used in this way. These computational learning theories take deliberate practice interactions as input – that is, they learn by doing from (a subset of the) cognitive tutor hints (bottom-out hints), feedback (on correctness), and repetition – and produce a domain-specific theory as output – that is, they incrementally generate a production system that makes novice-like student errors along the way and eventually can solve domain tasks like an expert.

These computational learning theories have generated predictions about domain-specific learning challenges. For example, when tutored on equation solving, SimStudent acquired more than one production rule for the solving tasks of the general form $ax = b$, particularly needing a separate production for the specific $-x = b$ form where the –1 coefficient is not explicitly present (Li, Matsuda, Cohen, and Koedinger, 2015). Indeed, students find this specific form about twice as hard as the general one, and the automatically generated domain-specific theory predicts that students will be better off with well-tailored tasks specific to this $-x = b$ form. Such confirmed theory-derived predictions are scientific evidence for the learning theory and are of practical importance as two productions rather than one implies that designers should develop two different sets of well-tailored tasks rather than one.

These computational learning theories have also replicated human learning findings. For example, like human learners (Patel, Liu, & Koedinger, 2016), these systems learn better from the *varied* contexts produced by interleaving fraction addition and multiplication tasks than from less varied contexts produced by blocked sequencing of these tasks (Li, Cohen, and Koedinger, 2013; MacLellan, 2017). The role of variation through interleaved practice intermixing is a theoretically crucial part of production rule learning. Initial blocked practice of a knowledge component (e.g., two rectangle area tasks in a row) aids the learning of the then-parts of productions – these produce the explicit performed steps of the procedure as it may be referred to intuitively, for example, by math education researchers. However, the production rule notion of procedural knowledge is importantly different in that the if-part of the production rule is a critical part of what needs to be learned and often the most difficult. Interleaved practice aids if-part learning, particularly in providing situations whereby the brain's learning mechanisms can refine if-parts that are currently overly general (e.g., if a fraction arithmetic problem is given, convert to a common denominator). The tutor's error feedback in such situations (e.g., a student converts on a fraction multiplication problem) provides the brain data to refine the if-part and add a new condition (e.g., the operator not multiplication). The experiments with computational models of learning have both precisely defined this mechanism and theoretically demonstrated its effectiveness.

Implications for Instructional Design

The work on Cognitive Tutors has implications for instructional design that pertain to a wide range of educational technologies and even to supporting learning more generally. This work exemplifies a broader approach to learning engineering, grounded in learning sciences theory (e.g., Koedinger, Corbett, & Perfetti, 2012). Learning engineering is an emerging interdisciplinary field that employs an iterative design process for improving learning, driven by learning science theory and by data from learning analytics, field-based design research, and rapid large-scale experimentation.

In a learning engineering approach, instructional design involves not just *applying* learning science (e.g., principles), but actually *doing* learning science as part of the process of designing instruction, and *extending* learning science as a consequence. That is to say, learning engineers create effective instruction through data-driven iterations that identify hidden skills and create novel tasks and scaffolds (i.e., design-loop adaptivity – Aleven, McLaughlin, Glenn, & Koedinger, 2017). It is becoming clear that data-driven methods can often detect such skills, even if the system's initial design was informed by careful, in-depth (though small *N*) cognitive task analysis (Tofel-Grehl & Feldon, 2013).

Designing effective scaffolds for hidden skills (or other skills, for that matter) is a matter not only of *applying* multimedia or other instructional design principles, but can require adapting learning principles to given knowledge goals (e.g., facts versus skills versus principles). It may even involve reconciling competing principles that

https://doi.org/10.1017/9781108894333.046 Published online by Cambridge University Press

seemingly lead to opposing design recommendations. For example, studying worked examples prior to problem-solving has been shown to be effective when the targeted knowledge is complex and principled (Kalyuga, Chandler, Tuovinen, & Sweller, 2001), but practice without much studying of examples works better for simpler forms of knowledge (e.g., facts or vocabulary), as demonstrated in a line of studies on the testing effect (Roediger & Karpicke, 2006a, 2006b). Similarly, while self-explanation (e.g., supported by prompts and feedback) can often enhance conceptual learning when the targeted knowledge is complex and principled, it may not have an effect with simpler forms of the knowledge-to-be-learned (e.g., choice of article for learning English as a second language (Wylie, Sheng, Mitamura, & Koedinger, 2011).

These examples illustrate that learning science and learning engineering form a two-way street, mutually influencing each other. Early experiences with this approach strongly suggest we need more instruction and learning environment design consciously targeting hidden skills. In addition to the examples of designing new scaffolds for hidden skills, this approach was successful in improving an online statistics course (Lovett, Meyer, & Thille, 2008). We also need to apply a learning engineering approach to consciously targeting, in our designs, *when*, *where*, and *what* learning. By this notion we mean that attention is paid in instructional design to helping students learn the (often perceptual) cues that distinguish conditions under which problem-solving actions are valid or strategic, including approaches such as adaptive, technology-based implementations of perceptual learning (Kellman & Krasne, 2018) or representation learning (Rau, 2017).

LIMITATIONS AND FUTURE WORK

Returning to our nutshell summary – effective learning requires *repeated* practice on *well-tailored tasks* in **varied** contexts with **explanatory** *feedback* and **as-needed instruction** – we reflect on limitations of existing research. One class of limitations is within these six cognitive tutor features and another is surrounding them. All six could use more research to better assess ecological and external validity and identify potential boundary conditions.

Repetition is important but not enough is known about how much repetition is needed, with Schnackenberg, Sullivan, Leader, and Jones (1998) being a nice exception. The earlier section on deliberate practice points to a number of Cognitive Tutor experiments demonstrating data-driven and domain theory-based design of *well-tailored tasks* yielding better learning. These efforts have been in mathematics and pursued by the authors and their collaborators. The external validity, as well as the general practicality, of this approach to well-tailored task design would be enhanced by explorations in other content areas and by other teams of researchers. With respect to *feedback*, one effort in particular has suggested a content-treatment boundary condition related to content wherein certain kinds of errors are particularly instructive (Mathan & Koedinger, 2005). This approach gives feedback relative to an intelligent novice model of desired performance in which some errors are considered reasonable and thus not subject to immediate feedback, rather than relative to an expert model where all errors receive immediate feedback. While these results are promising and theoretically compelling (i.e., error self-correction provides an extra level of sense-making to aid learning and recall), we are not aware of any studies trying to replicate them in other content domains.

The benefits of *varied* contexts have been demonstrated particularly in contrasts between interleaved and blocked practice to two topics, with experiments in the lab (e.g., Rohrer, 2012) and embedded in school settings (e.g., Patel, Liu, & Koedinger, 2016). At the same time, lab experiments indicate a theoretically compelling boundary condition on benefits of interleaving versus blocking. Carvalho and Goldstone (2014) predict and demonstrate a content-treatment interaction such that interleaving two topics/categories yields better learning for some content, but blocking yields better learning for other content. The boundary is determined by whether there is high similarity (low variation) in the stimuli across examples in the two topic areas (e.g., apart from the operator symbol, fraction addition and multiplication problems are otherwise the same) or high variation (e.g., examples of both cats and dogs are different from each in many ways). For high similarity content, an interleaving treatment produces better learning whereas for a high variability content, a blocking treatment produces better learning. In low variation content, interleaving produces variation in the response (e.g., the steps for adding fractions are much different from the steps for multiplying) and consecutive examples with similar stimuli (e.g., $\frac{1}{5} + \frac{2}{5} = \frac{3}{5}$ and $\frac{1}{5} * \frac{2}{5} = 2/25$) help a learner identify the few critically different features of the stimuli (e.g., the operator). In contrast, for content where the variation is already high, blocking helps the learner identify the few critically similar features of two consecutive examples (e.g., two cats may look quite different but both purr and meow). While this theoretically-predicted content-treatment interaction has been demonstrated in lab experiments with high internal validity and it may provide some justification for why so much current instruction is blocked rather than interleaved, there are too few school-embedded experiments that test the predicted benefits of blocking for high variability content.

Although *explanatory* feedback is typically provided in Cognitive Tutors through high-level hints and sometimes in error feedback messages, explanatory information is not systematically employed. Further, it has not been experimentally explored in systematic studies within Cognitive Tutors. One particularly useful study would be to experimentally contrast knowledge of results feedback only with explanatory feedback in addition to knowledge of results feedback. Meta-analysis results (van der

https://doi.org/10.1017/9781108894333.046 Published online by Cambridge University Press

Kleij, Feskens, & Eggen, 2015) suggest, from cross-study comparisons, a likely added benefit for explanatory feedback, however, the direct experimental evaluations of explanatory feedback do appear to have a control condition with knowledge of results feedback, but contrast with corrective feedback as a control (Moreno, 2004; Moreno & Mayer, 2005). Another nuanced limitation is that explanatory feedback is not given in Cognitive Tutors when students are correct on their own. Mitrovic, Ohlsson, and Barrow (2013) provide compelling evidence for learning efficiency benefits of providing such explanations as positive feedback, particularly when students are first successful on the related knowledge component. We also see a need for further research investigating the effectiveness of on-demand next-step explanatory hints in the context of learning by doing. Particularly welcome would be controlled experiments that test the value of explanatory hints versus bottom-out hints, where the latter give the answer only. Further nuance could be added by studies that test the value of explanatory hints versus prompting for self-explanations (under the same circumstances as hints are given). As well, it would be interesting to confirm experimentally that on-demand, explanatory next-step hints are most effective when the student has a medium level of mastery for the given knowledge component, as well as to test experimentally whether that finding generalizes from explanatory hints to prompted self-explanation.

As we have indicated, Cognitive Tutors not only provide feedback, but also *as-needed instruction* in the form of next-step hints. One open question is whether or when is such instruction sufficient and up-front instruction redundant. For example, Yannier, Hudson, and Koedinger (2020) found quite effective elementary science learning as a consequence of a predict–observe–explain inquiry cycle without any up-front instruction. It may be that skipping up-front instruction and starting directly with cognitive tutoring (i.e., in a formative assessment or deliberate practice loop) provides for equally effective but more efficient instruction for some content areas. Consistent with the KLI Framework, we hypothesize that skipping up-front instruction may be particularly relevant for learning goals that are facts and skills and perhaps less effective for high-level principles.

REFERENCES

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

Aleven, V., & Koedinger, K. R. (2013). Knowledge component approaches to learner modeling. In R. Sottilare, A. Graesser, X. Hu, & H. Holden (eds.), *Design Recommendations for Adaptive Intelligent Tutoring Systems* (*Learner Modeling*, 1, pp. 165–182). Orlando, FL: US Army Research Laboratory.

Aleven, V., McLaren, B. M., Roll, I., & Koedinger, K. R. (2016). Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, *26*(1), 205–223.

Aleven, V., McLaren, B. M., Sewall, J., van Velsen, M., Popescu, O., Demi, S., Ringenberg, M., & Koedinger, K. R. (2016). Example-tracing tutors: Intelligent tutor development for non-programmers. *International Journal of Artificial Intelligence in Education*, *26*(1), 224–269.

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In R. E. Mayer, & P. Alexander (eds.), *Handbook of Research on Learning and Instruction* (2nd ed., pp. 522–560). New York: Routledge.

Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP tutor. *Cognitive Science*, *13*(4), 467–505.

Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of Learning Sciences*, *4*, 167–207.

Baker, R. S. J. D., Corbett, A. T., & Koedinger, K. R. (2007). The difficulty factors approach to the design of lessons in intelligent tutor curricula. *International Journal of Artificial Intelligence in Education*, *17*(4), 341–369.

Butcher, K., & Aleven, V. (2013). Using student interactions to foster rule-diagram mapping during problem solving in an intelligent tutoring system. *Journal of Educational Psychology*, *105*(4), 988–1009.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*, 481–495.

Carvalho, P. F., & Goldstone, R. L. (2019). When does interleaving practice improve learning?. In J. Dunlosky, & K. Rawson (eds.), *The Cambridge Handbook of Cognition and Education* (pp. 411–436). Cambridge: Cambridge University Press.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243.

Clark, R. E., Feldon, D., van Merriënboer, J., Yates, K., & Early, S. (2007). Cognitive task analysis. In J. M. Spector, M. D. Merrill, J. J. G. van Merrie¨nboer, & M. P. Driscoll (eds.), *Handbook of Research on Educational Communications and Technology* (3rd ed., pp. 577–593). Mahwah, NJ: Lawrence Erlbaum Associates.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, *4*, 253–278.

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In *CHI '01: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 245–252). New York: ACM Press.

Corbett, A. T., McLaughlin, M., & Scarpinatto, K. C. (2000). Modeling student knowledge: Cognitive Tutors in high school and college. *User Modeling and User-Adapted Interaction*, *10*, 81–108.

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, *116*(39), 19251–19257.

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, *332*(603), 862–864.

Doroudi, S., Holstein, K., Aleven, V., & Brunskill, E. (2015). Towards understanding how to leverage sense-making, induction/refinement and fluency to improve robust learning. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (eds.), *Proceedings of the 8th International Conference on Educational Data Mining, EDM 2015* (pp. 376–379). Worcester, MA: International Educational Data Mining Society.

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.

Ericsson, K. A. & Simon, H. A. (1985). Protocol analysis. In T. A. van Dijk (ed.), *Handbook of Discourse Analysis* (Vol. 2; pp. 259–268). London: Academic Press.

Fries, L., Son, J. Y., Givvin, K. B. & Stigler, J. W. (2020). Practicing connections: A framework to guide instructional design for developing understanding in complex domains. *Educational Psychology Review*, *33*, 739–762.

Hattie, J. (2008). *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*. New York: Routledge.

Holstein, K., Aleven, V., & Rummel, N. (2020). A conceptual framework for human–AI hybrid adaptivity in education. In I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (eds.), *Proceedings, 21th International Conference on Artificial Intelligence in Education, AIED 2020* (pp. 240–254). Cham: Springer.

Holstein, K., McLaren, B. M., & Aleven, V. (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. In C. P. Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. du Boulay (eds.), *Proceedings, 19th International Conference on Artificial Intelligence in Education, AIED 2018* (Part 1, pp. 154–168). Cham: Springer.

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. *Journal of Learning Analytics*, *6*(2), 27–52.

Huang, Y., Aleven, V., McLaughlin, E., & Koedinger, K. (2020). A general multi-method approach to design-loop adaptivity in intelligent tutoring systems. *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II, 12164*, 124–129.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, *93*, 579–588.

Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist*, *44*, 250–266.

Kellman, P. J., & Krasne, S. (2018). Accelerating expertise: Perceptual and adaptive learning technology in medical learning. *Medical Teacher*, *40*(8), 797–802.

Koedinger, K. R. (2002). Toward evidence for instructional design principles: Examples from Cognitive Tutor Math 6. Invited paper. In D. Mewborn, P. Sztajn, D. Y. White, H. G. Wiegel, R. L. Bryant, & K. Nooney (eds.), *Proceedings of the 24th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 21–49). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Koedinger, K. R. & Aleven, V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review*, *19*(3), 239–264.

Koedinger, K. R., & Aleven, V. (2016). An interview reflection on "intelligent tutoring goes to school in the big city." *International Journal of Artificial Intelligence in Education*, *26*(1), 13–24.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8* (1), 30–43.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, *342*, 935–937.

Koedinger, K. R., & Corbett, A. T. (2006). Cognitive Tutors: Technology bringing learning sciences to the classroom. In R. K. Sawyer (ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 61–78). New York: Cambridge University Press.

Koedinger, K. R., Corbett, A. C., & Perfetti, C. (2012). The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, *36*(5), 757–798.

Koedinger, K. R. & McLaughlin, E. A. (2010). Seeing language learning inside the math: Cognitive analysis yields transfer. In S. Ohlsson, & R. Catrambone (eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 471–476). Austin, TX: Cognitive Science Society.

Koedinger, K. R., & McLaughlin, E. A. (2016). Closing the loop with quantitative cognitive task analysis. In T. Barnes, M. Chi, and M. Feng (eds.), *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 412–417). Raleigh, NC: International Conference on Educational Data Mining (EDM).

Koedinger, K. R., McLaughlin, E. A., & Stamper, J. C. (2012). Automated student model improvement. In K. Yacef, O. Zaïane, H. Hershkovitz, M. Yudelson, & J. Stamper (eds.), *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 17–24). Greece: Chania.

Li, N., Cohen, W. W., & Koedinger, K. R. (2013). Problem order implications for learning. *International Journal of Artificial Intelligence in Education*, *23*(1–4), 71–93.

Li, N., Matsuda, N., Cohen, W. W., and Koedinger, K. R. (2015). Integrating representation learning and skill learning in a human-like intelligent agent. *Artificial Intelligence*, *219*, 67–91.

Liu, R., & Koedinger, K. R. (2017). Closing the loop: Automated data-driven cognitive model discoveries lead to improved instruction and learning gains. *Journal of Educational Data Mining*, *9*(1), 25–41.

Liu, R., Koedinger, K. R., & McLaughlin, E. A. (2014). Interpreting model discovery and testing generalization to a new dataset. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 107–113). Worcester, MA: International Conference on Educational Data Mining.

Lovett, M. C. (1998). Cognitive task analysis in service of intelligent tutoring system design: A case study in statistics. In *International Conference on Intelligent Tutoring Systems* (pp. 234–243). Berlin: Springer.

Lovett, M. C., Meyer, O., & Thille, C. (2008). JIME-The open learning initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning. *Journal of Interactive Media in Education*, *2008*(1), Art. 13.

MacLellan, C. J. (2017). Computational *M*odels of *H*uman *L*earning: Applications for *T*utor *D*evelopment, *B*ehavior *P*rediction, and *T*heory *T*esting [Doctoral Dissertation]. Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA.

https://doi.org/10.1017/9781108894333.046 Published online by Cambridge University Press

MacLellan, C. J., Harpstead, E., Patel, R., & Koedinger, K. (2016). The apprentice learner architecture: Closing the loop between learning theory and educational data. In *Proceedings of the 9th International Conference in Educational Data Mining* (pp. 151–158). Worcester, MA: International Educational Data Mining Society.

Martin, B., Mitrovic, T., Mathan, S., & Koedinger, K.R. (2011). Evaluating and improving adaptive educational systems with learning curves. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research (UMUAI)*, *21*(3), 249–283.

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

Mayer, R. E. (2020). *Multimedia Learning* (3rd ed.), New York: Cambridge University Press.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger III, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*(2), 399.

Mitrovic, A., Ohlsson, S., & Barrow, D. K. (2013). The effect of positive feedback in a constraint-based intelligent tutoring system. *Computers & Education*, *60*(1), 264–272.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, *32*, 99–113.

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, *97*, 117–128.

Nagashima, T., Bartel, A. N., Silla, E., Vest, N., Alibali, M. W., & Aleven, V. (2020). Enhancing conceptual knowledge in early algebra through scaffolding diagrammatic self-explanation. In *Proceedings of International Conference of the Learning Sciences, ICLS 2020* (Part 1, pp. 35–42). Nashville, TN: International Society of the Learning Sciences.

Ohlsson, S. (1994). Constraint based student modeling. In J. E. Greer, & G. McCalla (eds.), *Student Modelling: The Key to Individualized Knowledge-Based Instruction*. NATO ASI Series (Series F: Computer and Systems Sciences) (vol. 125, pp. 167–189). Berlin: Springer.

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.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. *Educational Evaluation and Policy Analysis*, *36*(2), 127–144.

Patel, R. (2017). Addressing Interference in Fraction Learning: What Difficulties are Desirable? [Doctoral Dissertation]. Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA.

Patel, R., Liu, R., & Koedinger, K. (2016). When to block versus interleave practice? Evidence against teaching fraction addition before fraction multiplication. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2069–2074). Austin, TX: Cognitive Science Society.

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, *29*(4), 717–761.

Rau, M. A., Aleven, V., & Rummel, N. (2015). Successful learning with multiple graphical representations and self-explanation prompts. *Journal of Educational Psychology*, *107*(1), 30–46.

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, *14*(2), 249–255.

Roediger, H. L., and Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.

Roediger, H. L., and Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*, 355–367.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics practice problems boosts learning. *Instructional Science*, *35*, 481–498.

Roll, I., Baker, R. S. J. D., Aleven, V., & Koedinger, K. R. (2014) On the benefits of seeking (and avoiding) help in online problem- solving environments. *Journal of the Learning Sciences*, *23* (4), 537–560.

Schnackenberg, H. L., Sullivan, H. J., Leader, L. F., & Jones, E. E. K. (1998). Learner preferences and achievement under differing amounts of learner practice. *Educational Technology Research and Development*, *46*, 5–16.

Schofield, J. W. (1995). *Computers and Classroom Culture*. New York: Cambridge University Press.

Schooler, L. J., & Anderson, J. R. (1990). The disruptive potential of immediate feedback. In M. Piattelli-Palmarini (ed.), *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 702–708). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Shih, B., Koedinger, K. R., & Scheines, R. (2008). A response time model for bottom-out hints as worked examples. In *Proceedings of the First International Conference on Educational Data Mining*, 2008, Montreal, QC, pp. 117–126.

Tofel-Grehl, C., & Feldon. D. F. (2013). Cognitive task analysis-based training: A meta-analysis of studies. *Journal of Cognitive Engineering and Decision Making*, *7*(2), 293–304.

van der Kleij, F. M., Feskens, C. W. R., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, *85*(4), 475–511.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221.

VanLehn, K. (2016). Regulative loops, step loops and task loops. *International Journal of Artificial Intelligence in Education*, *26* (1), 107–112.

Weitekamp, D., Harpstead, E., & Koedinger, K. R. (2020). An interaction design for machine teaching to develop AI tutors. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, April 2020, Honolulu, HI, pp. 1–11.

Wylie, R., Sheng, M., Mitamura, T., & Koedinger, K. R. (2011). Effects of adaptive prompted self-explanation on robust learning of second language grammar. In *International Conference on Artificial Intelligence in Education* (pp. 588–590). Berlin: Springer.

Yannier, N., Hudson, S. E., & Koedinger, K. R. (2020). Active learning is about more than hands-on: A mixed-reality AI system to support STEM education. *International Journal of Artificial Intelligence in Education*, *30*(1), 74–96.