

An Interview Reflection on “Intelligent Tutoring Goes to School in the Big City”

Kenneth R. Koedinger¹ · Vincent Alevan¹

Published online: 9 December 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Our 1997 article in IJAIED reported on a study that showed that a new algebra curriculum with an embedded intelligent tutoring system (the Algebra Cognitive Tutor) dramatically enhanced high-school students’ learning. The main motivation for the study was to demonstrate that intelligent tutors that have cognitive science research embedded in them could have real impact in schools. This study was one of the first large-scale classroom evaluations of the integrated use of an Intelligent Tutoring System (ITS) in high schools. A core challenge was figuring out how to embed this new technology into a curriculum and into the existing social context of schools. A key element of the study design was to include multiple kinds of assessments, including standardized test items and items measuring complex problem solving and use of representations. The results were powerful: “On average the 470 students in experimental classes outperformed students in comparison classes by 15 % on standardized tests and 100 % on tests targeting the [course] objectives.” We suggested that the study was evidence “that laboratory tutoring systems can be scaled up and made to work, both technically and pedagogically, in real and unforgiving settings like urban high schools.” Since this study, many more classroom studies comparing instruction that includes an ITS against business as usual have been conducted, often showing advantages for the ITS-enhanced curricula. More rigorous randomized field trials are now more commonplace, but the approach of using multiple assessments in large-scale randomized field trials has not caught on. Cognitive task analysis will remain fundamental to the success of ITSs. A key remaining question for ITS is to find out how they can be used most effectively to support open-ended problem solving, either online or offline. Given all the recent excitement around Massive Open Online Courses (MOOCs), it is interesting to note that our field of Artificial Intelligence in Education

✉ Kenneth R. Koedinger
koedinger@cmu.edu

¹ Human Computer Interaction Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213-3891, USA

has been making huge, less recognized, progress with impact on millions of students and with the majority of those students finishing the course!

Keywords Intelligent tutoring systems · School-based field studies · Learning science · Cognitive model · Cognitive tutor

We interviewed Ken Koedinger who is the lead author on the 1997 IJAIED paper *Intelligent Tutoring Goes to School in the Big City* (Koedinger et al. 1997). In this interview, we look back, trace how the paper may have influenced subsequent work, and discuss issues for the future as they relate to this paper.

Q) What was the main motivation for the paper?

A) A key point of that paper was to provide a demonstration of the use of an intelligent tutoring system as part of a complete curriculum and an early attempt to evaluate its impact on student learning outcomes. We described a year-long study that occurred in the 1993–94 school year in three city schools in the US (Pittsburgh, PA) and involved a high proportion of low income and low performing students.

Q) Why was this work, at the time, an important thing to do and investigate?

A) We wanted to demonstrate that intelligent tutors that have core cognitive science research embedded in them could have real impact in schools and could be used practically. We wanted to show that research wasn't just scientists in arm chairs coming up with cool ideas, but that the research could lead to something that could make a difference.

Q) Was this the first classroom study that compared a curriculum with an intelligent tutoring system vs. a business-as-usual control condition?

A) There had been prior field trials including use of the LISP programming tutor with college students starting in 1984 (Anderson et al. 1989). Intelligent tutors for high school math tutors had also been used in smaller-scale experimental comparisons in schools (cf., Schofield et al. 1990). So, it was definitely not the first experimental trial in schools. There were also some other intelligent tutors being trialed in real settings including middle schools (e.g., Lester et al. 1997), colleges (e.g., Mitrovic 1998), medical schools (e.g., Eliot and Woolf 1996), and military training settings (e.g., Lesgold et al. 1988). I think it is safe to say that, at the time, it was the largest such study and the first I know of that involved a substantial integration of an intelligent tutoring system into a full-year K12 curriculum.

Q) What was particularly challenging in undertaking this study?

A) There were lots of challenges, including hardware issues. In those days, the Apple II series computers the schools used were just barely powerful enough to run the tutors. The live production system that drives tutor decisions, which could

easily respond in a blink on a smartphone today, sometimes caused noticeable delays in producing responses to student entries. Also, classrooms were not always prepared for the electricity demands of 30 computers. When the Apple-donated computers arrived at the first school, the room was not ready – new electrical circuits and outlets were needed. But, the school principal wanted to get started right away: “We can’t wait on the remodeling; just put extension cords down the hallways into the computer lab so we can get going.” This statement represents the enthusiasm we had from that school’s administration to try something new. But, it also represents a challenge of the time – hardware limitations – that is greatly reduced today.

Another core challenge was figuring out how to embed this new technology into the existing social context of schools, that is, into the instructional practices teachers were already using. A big theme of the Algebra tutor curriculum for the Pittsburgh Urban Mathematics Project, PUMP, as it was called at the time, was to work with teachers to understand how to integrate technology with classroom instruction. We took the strong position that we would redesign the whole course from the bottom up including replacements for the textbook, new kinds of assessments, use of collaborative learning in the classroom portion of the curriculum, and new teaching approaches. We evolved to the point where we recommended use of non-technology text materials (on paper and unbound) and practices in the regular classroom for 3 days a week and use of the tutor in the computer lab during the other 2 days a week. We did not have quite enough tutor material to reach this goal in this first study, but it was soon to come. Integration of the tutor technology with other teaching practices was an important goal of this study. This goal was a consequence of the struggles I had felt in a classroom trial with ANGLE, my geometry proof tutor (Koedinger and Anderson 1990). In that work, I experienced difficulties in trying to integrate the tutor into a pre-existing textbook curriculum.

Q) So a key driver of this approach was integration into the school, into the classroom, and into a curriculum. I’ll come back to that as we discuss the contemporary relevance of this work. Another interesting aspect of this study was that student learning outcomes were measured both with standardized test items and test items focused on representations and problem solving. The latter types of items were not so typical of standardized tests at the time and maybe not even of standardized tests today. To what extent has that aspect (i.e., using different types of test items, including items from standardized tests, to measure learning outcomes) been influential? Do you think it has been appropriately influential?

A) That’s a great question! I find today’s debate around assessment to be too black and white: Some people are for standardized assessment and some people are against it. What we did is to follow Ann Brown’s suggestion (Brown 1992) to use multiple kinds of assessment, both standardized tests and tests that we designed to best reflect the curriculum goals of achieving a more practical and general understanding of algebra. We continue to debate over standardized assessments versus open-ended assessments. A quite reasonable response to that

debate is to include both kinds of assessments. Unfortunately, this idea has not caught on in practice. In fact, my biggest frustration with what I think is otherwise a great idea, namely, the US Department of Education's major push for more randomized field trials, is that their What Works Clearinghouse (w-w-c.org) stuck to using single standardized tests as their only outcome measure. Using just one standardized assessment measure is not only potentially limited in terms of validity (i.e., multiple choice responses may not well measure the real thinking and performance we want students to achieve), but is also limited in that a single assessment is likely to miss (or underestimate) all of the learning that may be resulting from an intervention.

Q) You have mentioned a practical motivation, that is, to see if cognitive science could have a real influence on education. What do you see as a specific intellectual contribution and maybe you could also comment on whether that was misunderstood or lost along the way?

A) Certainly, there have been a lot of papers about cognitive tutors and model tracing intelligent tutoring systems. This paper contributed to that general literature. But what set it apart was that it addressed the social context of use, discussed how to connect a tutoring system with a curriculum and how to connect it with teachers. Another key part of the goal of that project and an important part of the work's impact was our investigation into interrelationships between isolated symbolic math procedures, real world problem solving, and students' interest in and motivation to learn math. I think it is fair to say that the paper stood out at the time as representing the AI in Education field – that the kind of work all of us had been doing could have significant and measurable impact on student learning outcomes. But, more importantly, it helped to press the field forward intellectually in the sense of setting an example for how others might try to measure the impact of their technologies on student learning outcomes.

Q) Now 18 years later, this paper is the most highly cited paper in the history of the journal. Is that a surprise to you? You may of course plead the fifth!

A) (laughing) Well I guess those stats first came out a number of years ago, so the first time I heard it, I was definitely surprised and pleased! That it remains in that position is a great honor.

Q) At the time, did you feel that this was a big step for the field to do this study? Of course, the outcomes were unknown before it was run.

A) I don't know, I think that, like many young researchers, I was pretty head down in trying to get the job done. I was certainly passionate about the project and had a strong bias to believe it would work. You know, it is interesting to reflect on the naïve enthusiasm that comes from a young researcher. This study was not the most rigorous experimental study ever run. It was a quasi-experiment

that did not involve random assignment of students or even classrooms to condition. Nevertheless, because there were not many evaluations at the time, fewer in real classrooms, and none of this size, this study stood out. But, to get back to that point of naïve enthusiasm, what I was getting at was that it probably had more of a marketing character than what a senior rigorous scientist might produce. It was really pitching a strong argument for the effectiveness of that system, which in retrospect, went beyond the quality of design of that particular study. Nevertheless, it was useful for the field in terms of inspiring people to push forward to try to do these kinds of evaluations. The rigorous randomized field trial evaluations that have since been done (e.g., Ritter et al. 2007a, b; Pane et al. 2014) were completely appropriate (except for the limitation to a single post assessment) and have, for the most part, validated the impact of the approach.

Q) Following up from that last comment, in recent years we have seen multiple meta-reviews in ITS and AIED, interestingly by researchers who aren't always closely associated with our research community. In summarizing the literature, I would argue that by now, many high-quality studies and meta-reviews (Steenbergen-Hu and Cooper 2013, 2014; Ma et al. 2014; VanLehn 2011; Pane et al. 2014; Kulik and Fletcher 2015) confirm the value of intelligent tutoring systems, perhaps with smaller effects than we might hope to see or that we have seen in some earlier studies. That confirmation seems to me one way (though definitely not the only way) in which this line of work has continued through the years. Any thoughts?

A) Well yeah, lots of thoughts, but one in particular is that I think we have become more nuanced about what an evaluation study is and I think we need to become even more nuanced about it. For instance, some studies funded by the Dept. of Education in the US or reported in What Works Clearinghouse (w-w-c.org), are really policy-level experiments (e.g., Pane et al. 2014; Sarkis 2004; Ritter et al. 2007a, b), in that they are not about the scientific investigation of whether intelligent tutors work or whether the key components of intelligent tutors are effective or not. Instead these policy-level experiments are about whether a policy to recommend use of these systems results in better outcomes. As such, classrooms are counted as receiving the treatment as long as they were assigned to it, and, irrespective of whether the tutoring system got used as intended, or even irrespective of whether it got used at all. At some level, this goes back to the social context issue. Namely, we want our systems to not only be effective in principle, but get used in practice (and get used as intended) and adoption is part of the social context. If adoption isn't successful, that should be something we as researchers ought to also be concerned about.

But, at the same time, it is also worth knowing whether systems are effective when used as intended. And it is important to keep the two ideas clearly separated. Some reported null results of Cognitive Tutors have been misinterpreted as suggesting the approach is not effective. If we are being scientific, we shouldn't be interpreting null results at all. But, to the point here, it is quite likely these null results are a consequence of poor implementation.

If I were to give a message to young researchers on this topic, it would be that there are multiple valuable forms of evaluation and the key is to find the right match with an important research question.

Q) Let's go back to the study published in 1997. You mention some limitations of the study that, you see now, resulted from your youthful exuberance. Are there any other limitations worth highlighting especially when considered from a contemporary perspective?

A) I wasn't meaning to say that the flaws in the study were a consequence of youthful exuberance, rather I was trying to say that the kind of marketing cheerleading for the approach was part of the youthful exuberance. I think we did the best we could at the time. It wasn't so much about not knowing better, it was more about the challenge of even getting the cooperation of schools to engage in a quasi-experimental study. Simply convincing some schools and teachers that weren't using the tutor to take all of our multiple post assessments was not easy. But, for sure, a random assignment design is better.

Another limitation of that study is that while it indicated strong potential for the whole course and tutor package, it did not help us understand which features were crucial. Since then we have tried to do many more studies that contrast specific elements of intelligent tutoring systems to understand which features and which underlying principles of learning instruction really matter (cf., Koedinger et al. 2012). A nice example, I'm sure you'll agree, is a study we did together contrasting a version of the Geometry Cognitive Tutor that prompted students to self-explain with one that did not (Alevan and Koedinger 2002).

Finally, that study only began to scratch the surface regarding issues of technology adoption and the social context of its use. There are still many related open questions today.

Q) That connects nicely to the topic of practical impact that I wanted to bring up. It seems to me that this paper, if I may state an opinion rather than a question, has had substantial practical impact throughout the years. Do you agree? What do you see as the practical impact of this study, and in particular, what role did this study play in the dissemination of the cognitive tutor?

A) This study was the initial snowball that started what turned out to be a widespread dissemination of the Cognitive Tutor course. The snowball quickly rolled down hill, getting bigger and bigger with the number of schools using the course nearly doubling every year in those earlier years. From the three schools involved in this study in the 1993–94 school year, we reached 75 schools in the 1998–99 school year (Corbett et al. 2001). In the process, the research team became increasingly overwhelmed and we began charging fees to staff the demand. From this snowball effect emerged the idea that we could create a self-sustaining company and, in 1998, Carnegie Mellon University and a large team of founders formed Carnegie Learning (<http://carnegielearning.com>). Now Cognitive Tutors are found in about 3000 schools and over a half million students

use the courses each year. I suspect that if you add up all the students impacted it is well over 5 million. There has been a lot of publicity in recent years about the tens or hundreds of thousands of students starting MOOCs and some small percentage completing them. Meanwhile our field of AI in Education has been making this huge, less recognized, progress with impact on millions of students and with the majority of those students finishing the course!

We have reflected about factors that led to this widespread dissemination (e.g., Corbett et al. 2001), including both the close integration of the technology into the social context of use and the evaluation results of enhanced student learning outcomes. These were both prime topics of this first paper reporting on Cognitive Tutor Algebra. So I do think it is fair to say that it played a key role in driving, and guiding, the dissemination success.

Q) To follow up, how do you see the role of scientific evidence of educational effectiveness of the system play into the dissemination of educational technologies? Did the study described in the 1997 paper contribute to the dissemination of cognitive tutors? Also, in today's environment, how important of a factor is scientific evidence in the marketplace?

A) Scientific evidence of educational effectiveness is a crucial component, but it is not sufficient. I wish we were in a world where such evidence was necessary for the success of an educational product, but we are not there yet. And, to be sure, such evidence is not sufficient. Even when using multiple assessments as we did, there are real risks that important learning or motivational outcomes are not being measured. And, as I mentioned, most evaluations today only use a single assessment, so they are at even greater risk of missing important outcomes or even negative side-effects. Thus, in addition to evidence of effectiveness, a product should also be evaluated in the context of educational theory – there should be a justification for why the product works that is supported by what we know about how people learn. Finally, the product should be intuitively compelling and desirable to teachers and students.

Although scientific evidence of effectiveness is not yet a necessary factor in the marketplace, it is much more prevalent today than it was in the early 1990s. This paper had real influence on that movement. Furthermore, it was important in also highlighting the other important roles of underlying theory and of adapting technology into the social context.

Q) Continuing on the topic of practical impact and trying to trace the paper's history to today and maybe the future, one observation that can be made about today's e-learning landscape, although there are notable commercial success stories of intelligent tutors like Cognitive Tutor and some of the constraint-based tutors, intelligent tutoring systems seem to be missing in action almost entirely in e-learning and MOOCs. Given the findings in your 1997 paper and subsequent evaluation studies on intelligent tutors (summarized in the meta-reviews mentioned above), this is surprising. Is certain key research still missing that would

facilitate the transitioning of tutor technology into actual educational practice? Are there other things that ITS researchers can do to help facilitate this transition?

A) One of the themes of the original paper keeps re-emerging in my conversations with respect to thinking about the role of technology at the college level. Namely, we have always viewed cognitive tutors not as a substitute for teachers but as a part of the toolkit for teachers that would enhance their ability to do the things they can uniquely do. That view of the role of technology, essentially freeing the teacher to do the magical things that only great teachers can do, is sometimes lost, certainly in caricatures of cognitive tutors as simply being about isolated practice. The idea that ITSs should free the teacher to do more of what she/he can do so well is critically important. In that regard, one of the key open issues is: How much of the more open-ended problem solving and reasoning that we want students to learn can intelligent tutoring systems support, either by aiding students directly during open-ended problem solving or by helping them prepare for such through well-chosen isolated practice of core skills? Another related question is: What's the relationship between that kind of more self-directed open-ended behavior we want students to eventually engage in and what they get from more structured practice within an intelligent tutoring system? These are tough questions that we haven't done enough to address. I do see clear and growing interest in these questions in the AIED community, which is great.

Q) What I'm hearing is that the next frontier for intelligent tutoring systems and a path towards continued impact and utility would be to address more open-ended kinds of competencies and skills.

A) Including more open-ended activities within intelligent tutors is one clear route, and one that should be taken, but it is not the only one. The other route is to more deeply investigate the relationship between what we can do really effectively in intelligent tutors, which is to isolate the things that are hardest for students to learn and give them better learn-by-doing experiences to help them learn. ITSs of the future should focus on core skills that (could) get reused over and over again, but that too few students adequately acquire, such as critical thinking and argumentation skills. The open-ended experience does not have to occur in the tutor. We can design tutoring environments to facilitate students' *preparation* to participate in open-ended activities outside of the tutoring environment. Self-directed student projects and social collaborative activities are powerful and the magic of the teacher's role should be preserved here. But, such activities can be more productive with the right kind of ITS preparation and follow-up. There's great potential in better understanding how tutors can position students to be effective in self-directed collaborative groups and, furthermore, to use self-directed projects to inspire students to seek a tutor to improve a skill they now recognize they need.

Q) So there might be some amount of bouncing back and forth between more traditional intelligent tutoring on structured parts of the task domain and new learning environments in which self-directed, open-ended problem solving takes place?

A) Yes, we need more research that tracks that kind of bouncing back and forth in a way that might, on one hand, facilitate students having used a tutor to be better in a project, or in the other direction, motivate students who have been engaged in a project to work hard to learn something new. For example, a self-directed collaborative project to build a robot or to design an interactive birthday card might reveal for a student that she needs more practice on programming or algebra skills.

Q) The rumor on the street is that you are working on a paper to be published in 2017 under the title Intelligent Tutoring Systems Go to School Worldwide. I was wondering if you might want to give a sneak preview of what's going to be in that paper.

A) (laughing) Go Worldwide!!! There are tons of interesting opportunities in that regard, particularly given how it's possible to get low cost technological devices, like smartphones, into the hands of poor communities and so forth. So, yeah, that's what we'll do in that paper. 2017 is that my deadline? That's a great goal and I think that we need to find ways to get people to work together on big goals like that. I keep emphasizing social context issues but it's presuming that the underlying deep cognitive analysis is being done well and effectively. You can't leave that out. In other words, we need to do everything for these projects to succeed. If we're going to go worldwide, we're going to need teams of folks to work together to get it done. Yeah, let's do it!

Final Thoughts and Reflections

Q) Any disappointments? Thoughts about how AIED could have a bigger impact?

A) One disappointment I have is that research efforts that have advanced our understanding of what works and/or advanced what's possible technologically have been slow to be incorporated in fielded Cognitive Tutor products. The research team at Carnegie Learning, led by Steve Ritter, has been extremely open about engaging in collaboration and supporting research studies. Nevertheless, much of the innovation in the field in general, and even in our contributions, has not seen much uptake. I am not sure why, but I *am* sure that there is more than one reason.

Some technically attractive efforts such as free-form explanation feedback (Popescu et al. 2003), collaborative peer tutoring (Walker et al. 2009), hand-written equation entry (Anthony et al. 2005), and teachable agents (Matsuda et al. 2012), have shown promise, but have not always produced clear demonstrations of improving student learning outcomes above and beyond those achieved by the existing tutor. Some approaches to supporting student metacognition and motivation, such as help-seeking (Roll et al. 2014), sense-making on errors (Mathan and Koedinger 2003), or reducing gaming the system (Baker et al. 2006) have been demonstrated to improve outcomes,

but have not been incorporated, perhaps in part because of the cost of technical development. In some cases where better outcomes in rigorous experimental studies have been clearly achieved, such as learning benefits for adding menu-based self-explanation (e.g., Alevan and Koedinger 2002), adding worked examples (e.g., Salden et al. 2010) or adding more personalized problem scenarios (Walkington 2013), there has been some limited influence on the product. However, it has not always been in line with particulars of the research design (e.g., Carnegie Learning added an upfront example in each unit, but the recommended approximate 50–50 example-problem ratio and use of fading were not incorporated). Some of the most recent and powerful demonstrations have shown how learning is improved through inserting new carefully designed tasks based on data analytic approaches that discover hidden skills that were not addressed in prior instruction (Koedinger and McLaughlin 2010; Koedinger et al. 2013). Neither of those particular results nor the methodology for producing them have been incorporated in industry.

These are projects I know well and are directly relevant to Cognitive Tutors, but there are many other powerful results in our AIED field that are too slowly getting out into use.

While slowly improving, a key issue remains that research is not the primary driver in decision making within companies – particularly when development and marketing departments express competing opinions. We need to change this situation by producing, as Bror Saxberg (2015) says, more “learning engineers” who know how to use scientific principles and analytic methods to produce better learning experiences and learning outcomes.

Q) What influence has this paper had on the wider research community as reflected by papers that have cited it?

A) Scanning the most highly cited papers that cite this one (using Google Scholar on May 4, 2015), one observation is the wide variety of disciplinary contexts and issues in which this paper is discussed. The disciplines span education, psychology, and computer science. The issues include debates on learning theory (Anderson et al. 1996), algebra teaching reform (Kieran 2007), how complex learning (e.g., of algebra) is implemented in the human brain (Anderson 2007), how technology is changing how children learn in school (Roschelle et al. 2000), and use of Bayesian networks to improve student modeling (Conati et al. 2002). In other words, researchers from many fields and topics of interest have found value in this piece of AI in Education research.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Alevan, V. A., & 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.

- Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.
- Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP Tutor. *Cognitive Science*, 13, 467–506.
- Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher*, 25(4), 5–11.
- Anthony, L., Yang, J., & Koedinger, K. R. (2005). *Evaluation of multimodal input for entering mathematical equations on the computer*. ACM Conference on Human Factors in Computing Systems (CHI 2005), Portland, OR, 4 Apr 2005, pp. 1184–1187.
- Baker, R. S. J. d., Corbett, A. T., Koedinger, K. R., Evenson, S., Wagner, A. Z., Naim, M., et al. (2006). Adapting to when students game an intelligent tutoring system. In M. Ikeda, K. D. Ashley, & T. W. Chan (Eds.), *Proceedings of the 8th International Conference on Intelligent Tutoring Systems, ITS 2006* (pp. 392–401). Berlin: Springer.
- Brown, A. L. (1992). Design experiments: theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141–178.
- Conati, C., Gertner, A., & Vanlehn, K. (2002). Using Bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction*, 12(4), 371–417.
- Corbett, A. T., Koedinger, K. R., & Hadley, W. H. (2001). Cognitive tutors: From the research classroom to all classrooms. In P. S. Goodman (Ed.), *Technology enhanced learning: Opportunities for change* (pp. 235–263). Mahwah: Lawrence Erlbaum Associates.
- Eliot, C., & Woolf, B. P. (1996). Iterative development and validation of a simulation-based medical tutor. In *Intelligent tutoring systems* (pp. 540–549). Springer Berlin Heidelberg.
- Kieran, C. (2007). LEARNING AND TEACHING ALGEBRA. *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics*, 1, 707.
- Koedinger, K.R., & Anderson, J.R. (1990). Theoretical and empirical motivations for the design of ANGLE: A new geometry learning environment. Appears in the working notes of the 1990 *AAAI Spring Symposium on Knowledge-Based Environments for Learning and Teaching*, Stanford University, March 27–29.
- 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: Cognitive Science Society.
- 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., 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. ISSN: 0364-0213. doi: [10.1111/j.1551-6709.2012.01245.x](https://doi.org/10.1111/j.1551-6709.2012.01245.x). print / 1551-6709.
- Koedinger, K. R., Stamper, J. C., McLaughlin, E. A., & Nixon, T. (2013). Using data-driven discovery of better student models to improve student learning. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, pp 421–430.
- Kulik, J. A. & Fletcher, J.D. (2015). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. *Review of Educational Research*. Prepublished April 17, 2015. doi: [10.3102/0034654315581420](https://doi.org/10.3102/0034654315581420)
- Lesgold, A.M., Lajoie, S., Bunzo, M., & Eggan, G. (1988). *Sherlock: A coached practice environment for an electronics trouble shooting job*. LRDC Report. Pittsburgh: University of Pittsburgh, Learning Research and Development Center.
- Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: affective impact of animated pedagogical agents. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems* (pp. 359–366).
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: a meta-analysis. *Journal of Educational Psychology*, 106(4), 901. doi:[10.1037/a0037123](https://doi.org/10.1037/a0037123).
- Mathan, S., & Koedinger, K. R. (2003). Recasting the feedback debate: Benefits of tutoring error detection and correction skills. In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *Artificial intelligence in education: Shaping the future of learning through intelligent technologies, Proceedings of AI-ED 2003* (pp. 13–18). Amsterdam: IOS Press [Best Student Paper Award.].
- Matsuda, N., Cohen, W. W., Koedinger, K. R., Keiser, V., Raizada, R., Yarzebinski, E., Watson, S. P., & Stylianides, G. J. (2012). Studying the effect of tutor learning using a teachable agent that asks the student tutor for explanations. In M. Sugimoto, V. Alevan, Y. S. Chee, & B. F. Manjon (Eds.), *Proceedings of the International Conference on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL 2012)* (pp. 25–32). Los Alamitos: IEEE Computer Society.
- Mitrovic, A. (1998). Experiences in implementing constraint-based modeling in SQL-Tutor. In: Goettl, B., Half, H., Redfield, C., Shute, V. (eds.), *Proceedings of ITS'98*, pp. 414–423.

- 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.
- Popescu, O., Alevan, V., & Koedinger, K. R. (2003). A knowledge-based approach to understanding students' explanations. In V. Alevan, et al. (Eds.), *Suppl Proc 11th Int Conf on Artificial Intelligence in Education*, Vol. VI (pp. 345–355). School of Information Technologies, University of Sydney.
- Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007a). Cognitive tutor: applied research in mathematics education. *Psychonomic Bulletin & Review*, 14(2), 249–255.
- Ritter, S., Kulikowich, J., Lei, P., McGuire, C. L., & Morgan, P. (2007b). What evidence matters? A randomized field trial of Cognitive Tutor Algebra I. In T. Hirashima, U. Hoppe, & S. S. Young (Eds.), *Supporting learning flow through integrative technologies* (Vol. 162, pp. 13–20). Amsterdam: Ios Press.
- Roll, I., Baker, R. S. J. d., Alevan, 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. doi:10.1080/10508406.2014.88397.
- Roschelle, J. M., Pea, R. D., Hoadley, C. M., Gordin, D. N., & Means, B. M. (2000). Changing how and what children learn in school with computer-based technologies. *The Future of Children*, 10, 76–101.
- Salden, R. J. C. M., Koedinger, K. R., Renkl, A., Alevan, V., & McLaren, B. M. (2010). Accounting for beneficial effects of worked examples in tutored problem solving. *Educational Psychology Review*. doi:10.1007/s10648-010-9143-6.
- Sarkis, H. (2004). *Cognitive tutor algebra I program evaluation: Miami–Dade County public schools*. Lighthouse Point: The Reliability Group. Available at www.carnegielearning.com/web_docs/sarkis_2004.pdf.
- Saxberg, B. (2015). Why we need learning engineers. *Chronicle of Higher Education*. Retrieved from <http://chronicle.com/article/Why-We-Need-Learning-Engineers/229391/>
- Schofield, J. W., Evans-Rhodes, D., & Huber, B. R. (1990). Artificial intelligence in the classroom: the impact of a computer-based tutor on teachers and students. *Social Science Computer Review*, 8(1), 24–41.
- Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, 105(4), 970–987. doi:10.1037/a0032447.
- Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, 106(2), 331–347. doi:10.1037/a0034752.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.
- Walker, E., Rummel, N., & Koedinger, K. R. (2009). Integrating collaboration and intelligent tutoring data in evaluation of a reciprocal peer tutoring environment. *Research and Practice in Technology Enhanced Learning*, 4(3), 221–251. ng Sciences, Inc. ISSN 1819–0146.
- Walkington, C. (2013). Using learning technologies to personalize instruction to student interests: the impact of relevant contexts on performance and learning outcomes. *Journal of Educational Psychology*, 105(4), 932–945.