CrossMark

Stupid Tutoring Systems, Intelligent Humans

Ryan S. Baker[1]

Abstract The initial vision for intelligent tutoring systems involved powerful, multi-faceted systems that would leverage rich models of students and pedagogies to create complex learning interactions. But the intelligent tutoring systems used at scale today are much simpler. In this article, I present hypotheses on the factors underlying this development, and discuss the potential of educational data mining driving human decision-making as an alternate paradigm for online learning, focusing on intelligence amplification rather than artificial intelligence.

The Initial Vision of Intelligent Tutoring Systems

One of the initial visions for intelligent tutoring systems was a vision of systems that were as perceptive as a human teacher (see discussion in Self 1990) and as thoughtful as an expert tutor (see discussion in Shute 1990), using some of the same pedagogical and tutorial strategies as used by expert human tutors (Merrill et al. 1992; Lepper et al. 1993; McArthur et al. 1990). These systems would explicitly incorporate knowledge about the domain and about pedagogy (see discussion in Wenger 1987), as part of engaging students in complex and effective mixed-initiative learning dialogues (Carbonell 1970). Student models would infer what a student knew (Goldstein 1979), and the student's motivation (Del Soldato and du Boulay 1995), and would use this knowledge in making decisions that improved student outcomes along multiple dimensions.

   An intelligent tutoring system would not just be capable of supporting learning. An intelligent tutoring system would behave as if it genuinely *cared* about the student's success (Self 1998).[1] This is not to say that such a system would actually simulate the

---

[1] A higher standard than many human instructors (Farber and Miller 1981).

✉ Ryan S. Baker
   ryanshaunbaker@gmail.com

[1]   Teachers College, Columbia University, New York, NY, USA

process of caring and being satisfied by a student's success, but the system would behave identically to a system that did, meeting the student's needs in a comprehensive fashion.

And these systems would not just be effective at promoting students' learning, the systems themselves would also learn how to teach (O'Shea 1982; Beck 1997). The systems would study what worked, and when it worked, and they would improve themselves over time (Beck 1997; Beck et al. 2000).

In 2015, after decades of hard work by many world-class scientists, we have wonderful demonstrations of the potentials of this type of technology. We have systems that can provide support on every step in a student's thinking process (VanLehn et al. 2005; Anderson et al. 1995), systems that can talk with students in natural language (Nye et al. 2014; earlier on, see Stevens and Collins 1977); systems that model complex teacher and tutor pedagogical strategies (Heffernan and Koedinger 2002; Khachatryan et al. 2014); systems that recognize and respond to differences in student emotion (D'Mello et al. 2010; Arroyo et al. 2014); and simulated students that enable human students to learn by teaching (Leelawong and Biswas 2008; Matsuda et al. 2010).

And at the same time, we have intelligent tutoring systems being used by tens or hundreds of thousands of students a year, and achieving outcomes that would make the earliest proponents of intelligent tutoring systems proud, with statistically significant positive impacts on student learning, including SQL-Tutor (Mitrovic and Ohlsson 1999), ALEKS (Craig et al. 2013), Cognitive Tutor (Pane et al. 2014) and ASSISTments (Koedinger et al. 2010).

A Disconnect

But there is a disconnect between the vision of what intelligent tutoring systems could be, and what they are; a disconnect between the most impressive examples of what intelligent tutors can do, and what current systems used at scale do. In fact, the most widely used intelligent tutoring systems are in some ways the furthest from the initial vision of researchers like Carbonell and Self.

We can start with the ways that tutors used at scale resemble this initial Vision. *domain* modeling is present in many of the systems used at scale. For example, Cognitive Tutors use production-rule models to represent skill (Anderson et al. 1995), ALEKS uses prerequisite hierarchies to represent the connections between items (Falmagne et al. 2013), and SQL-Tutor uses extensive constraint-based models to represent appropriate performance (Mitrovic and Ohlsson 1999).

So, too, student knowledge modeling is seen in some of the systems used at scale, with Cognitive Tutors using Bayesian Knowledge Tracing (BKT; Corbett & Anderson, 1995), and ALEKS using formulas based on knowledge space theory (Falmagne et al. 2013). But here, despite the decades of work on knowledge modeling, and the intense competition between approaches seen in published papers (see, for instance, Pavlik et al. 2009; Pardos et al. 2011; Papousek et al. 2014), the approaches used in practice are largely fairly simple. For example, many systems in wide use depend on simple heuristics to assess student mastery, such as whether the student gets three right in a row (Heffernan and Heffernan 2014).

And work to adapt to affect, engagement, meta-cognition, and self-regulated learning has yielded projects that have improved student outcomes (Baker et al., 2006;

Arroyo et al., 2007; D'Mello and Graesser 2012), but has largely been investigated in small-scale studies rather than being deployed at large scale. What is most intriguing is that some of these projects have been conducted in the context of systems being deployed at scale; but the research innovations, even apparently successful ones, are not then integrated into the systems deployed at scale.

So, too, despite the initial enthusiasm about systems that can use reinforcement learning to improve themselves (e.g. Beck 1997; Beck et al. 2000), few systems incorporate this capacity. There have been some simulation studies around reinforcement learning and related machine learning approaches (e.g. POMDPs) for intelligent tutors (Chi et al. 2010; Rafferty et al. 2011), but little work to deploy these solutions into approaches used at scale.

As such, we are left with a bit of a puzzle. ITS research has been successful at producing impressive technologies (and there are many beyond the small sample discussed here), and ITS systems are now being used by tens or hundreds of thousands of learners, but the systems being used at scale are generally not representative of the full richness that research systems demonstrate.

New Excitement with MOOCs

This mystery is particularly relevant at this historical moment. Massive Online Open Courses, or MOOCs (McAuley et al. 2010), have emerged into the consciousness of a large proportion of educated people worldwide. MOOCs provide a combination of video lectures, online assignments, and discussion forums (and connectivist MOOCs, or c-MOOCs, provide additional pedagogies as well – Rodriguez 2012). These systems can incorporate intelligent tutor-style assignments (Aleven et al., 2015) but typically provide an experience more focused on didactic lectures and discussion forums than on the types of activities typical to intelligent tutoring systems.

Many of the leading proponents of MOOCs have advertised them as crucibles for innovation, with huge potential to revolutionize education; making high-quality learning materials, once available only to very limited numbers of learners, available to the masses. Much of the rhetoric and ideas around MOOCs matches the earlier enthusiasm around intelligent tutoring systems – MOOCs will leverage the power of big data and reinforcement learning to improve themselves (Raghuveer et al. 2014); MOOCs will adapt to individual learners and provide a truly personalized learning experience (Norvig 2012).

Thus far, most MOOCs have fallen far short of the hype. Intelligent tutoring style or simulation-based assignments have only begun to be embedded into MOOCs (Ostashewski 2013; Diaz et al. 2013; Aleven et al., 2015); collaborative chat activities or social media leveraging activities have only been lightly deployed (Joksimović et al. 2015; also see dance.cs.cmu.edu); and at the time of this writing, most MOOCs are still limited to providing very basic multiple-choice assignments surrounding sometimes low-quality video lectures, and overwhelmingly large discussion forums which many instructors participate in only lightly or not at all. Perhaps MOOCs can be forgiven for not achieving in three years[2] what intelligent tutoring systems still struggle to provide after decades, but the fact remains that the same pattern of development seems to be

---

[2] MOOCs existed as a concept several years before this (e.g. McAuley et al. 2010), but have achieved most of their scale much more recently.

replicating itself: large-scale deployment of solutions that fall far short of the visions and the rhetoric.

As such, this appears to be a valuable time to take stock of how online learning systems used at scale differ from the original vision of intelligent tutoring systems, and what this might mean.

A Different Vision

One potential response to the still relatively simple technology seen in online learning is that these developments take time, and that solutions can simply be slow to make their way from the research laboratory, to the research classroom, to the full diversity of classrooms (Corbett et al. 2001). This perspective very well may be right. There are a number of economic factors that come into play that may slow the progress of innovations into use. Certainly, it is the perspective that a great deal of my own work has adopted. It has been my dream, and continues to be my dream, that intelligent tutoring systems that incorporate detectors of – say – gaming the system, and adapt in real-time when students game the system, will one day be commonplace.

But I find myself wondering, is it possible that this is *not* what the world of educational technology will look like? Is it possible that there's another path towards developing excellent online learning technologies? And is it possible that this world has been developing around us, that this alternate path is already happening, while we continue to work on building more and more sophisticated intelligent tutors?

So let me pose the possibility of a different way that the excellent online learning systems of tomorrow could be developed. Perhaps we do not in fact need intelligent tutoring systems. Perhaps instead what we need, what we are already developing, is **stupid tutoring systems**.[3] Tutors that do not, themselves, behave very intelligently. But tutors that are **designed intelligently**, and that **leverage human intelligence**.

In other words, perhaps what we need is stupid tutoring systems, and intelligent humans.

What would this look like?

Envision that we design a system, with relatively simple interactions with students. A student is posed a mathematics problem. They can answer it, or request a hint. If they ask for a hint, they get a pre-defined set; if they make a wrong answer, they get a message telling them why they are wrong, or perhaps a scaffolding problem that helps them with a key step towards the answer. They keep working on math problems for the current skill, until they can get three in a row right. And the next morning, their teacher can look up which problems they and their classmates got right and wrong.

I am referring, of course, to the ASSISTments system (Heffernan and Heffernan 2014), one of the most widely used (and simplest) intelligent tutoring systems in common usage today.

ASSISTments is not just an example of simple design. It's an example of good design. It's an example of data-driven design. Data-driven design is not a new idea in AIED; it dates back at least to Self's (1990) model of the iterative design of intelligent tutoring systems.

---

[3] Credit where credit is due: I have heard this term used before by Darren Gergle and Piotr Mitros.

But systems like ASSISTments bring iterative design based on experimentation to a new level. There have been innumerable[4] studies testing different aspects of the ASSISTments system (Ostrow and Heffernan 2014), answering questions such as: should we use hints or scaffolds (Razzaq and Heffernan 2006)? should hints be delivered using text or video (Ostrow and Heffernan 2014)? how much do messages on growth mindsets benefit students (Ostrow et al. 2014)? should we display gaming behavior indicators to teachers (Walonoski & Heffernan, 2006)? The design of ASSISTments is based, from the ground up, on data. Data collected through hundreds of A/B tests, quick randomized controlled trials. Data analyzed by humans.

ASSISTments looks a whole lot like this vision I am articulating. And it has scaled. And it has helped students learn (Koedinger et al. 2010).

This idea of intelligence being in humans rather than the tools they use is not a novel idea, of course. The research communities on human computation (Quinn and Bederson 2011), on crowdsourcing (Brabham 2008), and on intelligence amplification (Freund 2013), have had the same idea. The distinction between learning analytics and educational data mining can in some ways be brought back to this difference (Baker and Siemens 2014). The idea that people develop tools, and tools are refined and used by intelligent individuals based on practice, is a well-known and long-standing idea within human-computer interaction (Winograd and Flores 1986). This often takes the form of considering human-tool systems or human-human-tool systems or broader socio-technological systems. Sometimes, system-based views on technology and education can descend into fancy rhetoric and thought-provoking essays (Winograd and Flores 1986; Greeno 1997; Baker 2016), rather than solid empirical evidence. I am not arguing for that as a general research approach. I am a hard-edged passionate believer in data, data, and more data, with as quantitative a lens as possible.

But we seem to be entering an era where data is being more used in the service of design and human decision-making, than automated personalization.

Educational Data Mining: Making Discoveries, Improving Education

A lot of the rhetoric around the emerging field of educational data mining has been that big data will enable us to develop rich student models that can be used in the kinds of automated personalization that we see in intelligent tutoring systems. I am familiar with that rhetoric. I have written a lot of it (Baker & Yacef, 2009; Baker, 2010; Baker and Siemens 2014).

But that has not been the only vision for big data and education. A second vision, present from the start, is that we can use educational data mining to make basic discoveries in the science of learning and enhance theory (Beck and Mostow 2008; Jeong and Biswas 2008; Baker & Yacef, 2009; Baker, 2010).

For example, sophisticated models of affect, engagement, meta-cognition, self-regulated learning, and domain structure have often been looked at as tools to enable automated intervention; systems that can tell when a student is bored (for instance), and adapt to re-engage that student. But instead of building them into an intelligent tutor, we can make them tools for research and analysis by intelligent humans. The findings of these analyses can be used in turn to enhance the design of online learning systems.

---

[4] Well, perhaps not *innumerable.* But a whole lot.

For example, Koedinger and colleagues (2012) used learning factors analysis to re-fit the mappings between knowledge components and specific problem steps in Cognitive Tutor Geometry. Though the skills in this learning system were derived through extensive cognitive modeling, they were still imperfect, and educational data mining methods were able to figure out how. Koedinger and his colleagues found through this research that some problem steps that were thought to involve the same skill involved different cognitive skills; for example, some problems involving computing the area of a circle, thought to involve a single skill, involved backwards reasoning instead of forward reasoning, resulting in an additional knowledge component to learn. In subsequent work, Koedinger et al. (2013) used these findings to re-design a tutor lesson, leading in an experimental study to significantly faster learning.

A similar goal can be achieved with predictive analytics models – models that make a prediction of some longer-term outcome, such as course failure or dropout (Arnold and Pistilli 2012; Barber and Sharkey 2012; Ming and Ming 2012), or failure to graduate (Dekker et al. 2009; Kovačić 2010). Some of these models rely upon fine-grained student behavior (Arnold and Pistilli 2012; Ming and Ming 2012), others rely more upon demographics or other relatively stable student attributes (Kovačić 2010; Barber and Sharkey 2012). While the use of demographic data can be controversial and can be argued to be relatively less actionable (and I argue this point in my MOOC – Baker 2014), learner behaviors often provides direct indicators that are easy to think of interventions for.

As such, EDM has the potential to help us better understand learning and the phenomena that surround it. This can help us in turn to enhance the design of online learning environments. I'll discuss some examples of the use of predictive analytics models for this in the following section.

Learning Analytics: Better Reporting

Several years after the emergence of the educational data mining community, a second community emerged, seemingly working in the same space – the learning analytics community. As is often the case when two communities emerge in the same general area, at first there was confusion as to what the boundaries were between these communities.

It quickly became clear that, despite common interests, there were important differences between learning analytics and educational data mining. George Siemens and I summarize a few of the core differences in (Siemens and Baker 2012; Baker and Siemens 2014). But one of the key differences, at least in terms of the questions this article considers, was a shift from using data mining to support automated intervention, to using it to support reporting.

A system can report on a student's state to several different potential stakeholders. Open learner models and related systems report on a student to the student themselves, and can also provide reports to student peers for comparative purposes (e.g. Bull and Nghiem 2002). Many at-risk prediction systems report on a student to their instructors (e.g. Arnold and Pistilli 2012). Other systems present reports to guidance counselors, parents (Broderick et al. 2010; Hawn, 2015; Bergman, under review), regional curriculum coordinators, and school or university leaders (e.g. Zapata-Rivera and Katz 2014).

One of the best-known examples of the use of reporting to drive change is the case of Course Signals, originally Purdue Course Signals (Arnold and Pistilli 2012). This system takes models that can predict student success, applies the models to make predictions in real time, determines why students are at risk, and provides this information to instructors, along with practice recommendations. For example, the system may suggest that an instructor email a student to discuss their inactivity in the course, and may even recommend specific text for such an email. It is, of course, up to the instructor's discretion whether he or she will follow those recommendations; this is typically seen as an advantage, as an instructor may be aware of situation-specific information unavailable to the system that suggest an alternate course of action is more appropriate in specific cases. Course Signals has been found to lead to significantly higher course and university retention (Arnold and Pistilli 2012).

Other early-warning systems, similar to Course Signals, have sprung up, with a veritable ecosystem of companies (and non-profits, and university projects) offering predictive analytics on student success and early warnings for instructors when a student is at risk, including ZogoTech (Wood and Williams 2013), and the Open Academic Analytics Initiative (Jayaprakash et al. 2014).

The emergence of these themes is also starting to be seen in other types of online learning and AIED technologies. For example, the S3 project gives teacher an ongoing distillation of student activities in a full-room multi-screen collaborative learning activity, giving the teacher the ability to orchestrate and change the activities students were working on based on this information, or to interact with individual student groups in real time (Slotta et al. 2013). In the Virtual Collaborative Research Institute system, real-time information is given to instructors on student participation in collaborative chat and whether students are agreeing or disagreeing, towards helping the instructors to take real-time action to improve the quality of collaborative discussions (Van Leeuwen et al. 2014), specifically targeting groups having problems (van Leeuwen et al. 2015). Learning analytics analysis of student performance in serious games is now starting to be offered to instructors as well (Serrano-Laguna and Fernández-Manjón 2014).

These systems join the long-term efforts of intelligent tutoring systems like Cognitive Tutor, Reasoning Mind, and ASSISTments to provide extensive reports to teachers (Anderson et al. 1995; Feng and Heffernan 2005; Miller et al., 2015). In the case of Reasoning Mind, teachers use these reports in real-time, obtaining information that a student is struggling with a specific concept *right now*, and engaging in proactive remediation (Miller et al., 2015). In the case of ASSISTments, teachers often read the reports of the previous night's homework before class, and re-work their planned lecture based on data about what questions students struggled with (Feng and Heffernan 2005).

Analytics-based reporting for parents are just emerging, from the attempts of charter schools in New York City to provide predictions to parents along with details on the factors creating risk for individual students (Hawn, 2015), to text messages sent to parents that give details on missing assignments and low grades (Bergman et al., under review), to text messages and online reports to parents on what material their students are studying and how they are performing (Broderick et al. 2010).

These types of analytics have become useful at a broader grain-size as well. Data from automated detectors of student engagement is now being made available to

regional coordinators for the Reasoning Mind system to identify specific classrooms where teachers need additional support (Mulqueeny et al., 2015).

These systems address different goals from each other – from trying to prevent course dropout at the college level, to changing content instruction and classroom pedagogy, to identifying trouble spots in regional implementations of learning technologies. But what they share in common is the goal of getting key information to a human being who can use it. Some solutions are more prescriptive – Course Signals recommends specific actions and email text to instructors. Other systems simply give indications of performance and behavior, and let the instructor or parent decide what to do. As a group, these solutions place the ultimate decisions in the hands of a human being.

Advantages of Humans

In the previous two sections, I have discussed how educational data mining and learning analytics methods – particularly automated detection of complex constructs, and predictive analytics – can be put to powerful use in two fashions beyond automated intervention: re-design based on discovery with models analysis, and reporting.

In both these uses of prediction models, the common thread is that AI technology is used to derive important information about an online learning system, but the action taken is not by the system itself; instead action is taken by a human. The learning system is not itself intelligent; the human intelligence that surrounds the system is supported and leveraged. Designers are informed to support re-design and enhancement of a learning system; instructors are informed so that they can support the student right away.

There are several advantages to this approach, relative to a more automated intervention strategy.

First of all, automated interventions can be time-consuming to author. Researchers seldom report how long it took them to develop new interventions, but authoring an entirely new behavior for a pedagogical agent is not cheap. For example, it took this author several months to design and implement the pedagogical agent which responded to gaming the system in Cognitive Tutors (Baker et al., 2006). That agent only worked in a small number of Cognitive Tutor lessons; scaling the intervention would have been less costly than building it in the first place, but it still would have taken considerable effort.

Second, automated interventions can be brittle. No predictive model is perfect (human predictions are not exactly perfect either[5]). An automated system cannot recognize when a model is clearly wrong, due perhaps to unusual circumstances or a change in context. And if an automated intervention is not working, it's difficult for a system to recognize this and change tack. More importantly, if an automated intervention goes badly wrong in an unexpected way (the student starts crying), the system has limited scope to recognize this and take action.

Automated interventions are brittle in a different way as well: students can adapt faster than automated systems. An encouraging message may not be so encouraging the 12th time; a student may figure out how to defeat an intervention

---

[5] "In the 21st century, in order to control traffic jams in the air, there will be more and more flying policemen." – Villemard, 1910.

designed to prevent gaming the system, and find new ways to game (a behavior reported in Murray and VanLehn 2005).

Third, students change over time. Automated interventions therefore need to be re-checked and adapted over time. For example, overall student attitudes towards intelligent tutoring systems appear to have changed a great deal over the last 20 years. Schofield (1995) reported that Pittsburgh high school students were extremely engaged and even came outside of regular class hours to use the Cognitive Tutor, a behavior that does not appear to be common in American classrooms today. However, extremely high engagement along the lines reported by Schofield has been more recently reported among students using Cognitive Tutors in the Philippines (Rodrigo, Baker, & Rossi, 2013). It is not clear that this engagement will persist if intelligent tutoring systems become a regular part of education there.

None of these limitations are insurmountable. If there's sufficient resources, new interactions can be developed; an intelligent tutoring system could conceivably be designed to recognize when an intervention is failing for a specific student; systems can be re-checked and re-designed over time (and this already happens); and systems can be tested to see if students respond as expected to all interventions.

But on the whole, this presents some possible reasons why human-driven changes are playing a larger role than automated intervention. Humans are flexible and intelligent. Humans cannot sift through large amounts of information quickly, which is why they need data mining and reporting to inform them. But once informed, a human can respond effectively.

Going Forward

In this article, I have discussed how the original vision for intelligent tutoring systems – powerful, flexible systems that adapt in a range of ways to the learner – does not seem to entirely match the intelligent tutoring systems we see at scale. We are not seeing as much artificial intelligence as we expected, at least not in terms of how these systems interact with students. Instead, we seem to see much less rich tutoring systems that nonetheless leverage a lot of a different type of intelligence – human intelligence. We are developing what one could flippantly call ***stupid tutoring systems***: tutors that are not, in and of themselves, behaving in an intelligent fashion. But tutors that are ***designed intelligently***, and that ***leverage human intelligence***.

Modern online learning systems used at scale are leveraging human intelligence to improve their design, and they are bringing human beings into the decision-making loop and trying to inform them (and the information they provide is in many cases distilled using sophisticated algorithms).

If we were to adopt this as an alternate paradigm for artificial intelligence in education (AIED) – artificial intelligence as intelligence amplification (Freund 2013), how would the field change? What would be some of the new problems?

First of all, we'd need to face the challenge that human beings vary in quality. We know that different teachers have different impacts on student success (Darling-

---

[0] "In the 21st century, in order to control traffic jams in the air, there will be more and more flying policemen." – Villemard, 1910.

Hammond 2000); we know that there is a range in the results produced by different human tutors (Kulik and Kulik 1991); we know that some designers and programmers are better than others (Brooks 1975). So not all human responses to reporting are going to be equally effective.

And no one, no matter how sharp, gets it right all the time.

So we need to design processes that help human beings figure out what works, and processes to scale what works, and processes to figure out why it works.

We know a decent amount about how to do this for designing learning systems. There are educational data mining methods like learning decomposition explicitly designed to help us figure out which strategies work (Beck and Mostow 2008); and a growing body of literature on A/B testing and automated experimentation in education (Mostow 2008; Ostrow and Heffernan 2014). Beyond this, there's an active if controversial body of research on how to determine which teachers are effective (cf. McCaffrey et al. 2003). Platforms like the Pittsburgh Science of Learning Center LearnLabs (Koedinger et al. 2012a, b) and more recently the efforts to make ASSISTments an open platform for research (Ostrow and Heffernan 2014) are positive trends in this direction. And there's a long tradition of distilling educational research into guidelines for practice and design (Bransford et al. 1999; Koedinger et al. 2012a; Pashler et al. 2007; Clark and Mayer 2003). These guidelines can support scalability, so that solutions we develop in one platform can influence future platforms.

There has been less work to do this for human action on reports. Systems like Course Signals attempt to scaffold effective practice by suggesting actions to instructors (Arnold and Pistilli 2012). But these systems do not allow bottom-up improvement, just improvement by designers. Other online platforms use teacher professional development to share and disseminate effective practices – for example, ASSISTments, Reasoning Mind, ALEKS, and Cognitive Tutor all provide extensive professional development for teachers. Beyond this, there are now online communities and discussion forums for teachers to share strategies (Maull et al. 2011). But these approaches bring the recommended practices outside the system, and as such are not particularly timely. A valuable area of future research may be to use crowd-sourcing to solicit strategies from instructors, and data mining to test their effectiveness. Human-driven intervention strategies found to be effective could then be automatically suggested to instructors, much like Course Signals does. This would effectively create a recommender system for instructors, helping less effective instructors to catch up to their more effective peers.

A second opportunity for research along these lines is how to improve our models to take account of how they are used. We already know that indicators used as the basis for intervention lose their effectiveness as predictors, a well-known phenomenon in economics (Campbell's Law; Campbell 1976). For example, if Bayesian Knowledge Tracing (BKT) is used in a learning system without mastery learning, it can predict post-test scores (Corbett & Anderson, 1995). But if BKT is used to drive mastery learning, it ceases to be able to predict post-test scores (Corbett and Bhatnagar 1997). This is a plausible concern for the use of these models discussed above. Imagine a teacher getting predictive analytics on student failure after every assignment. If the instructor took action after the third assignment, but the system did not take this into account, the system's prediction after the fourth assignment might be overly pessimistic. As such, we need to investigate how to

create second-order models that provide useful analytics information after intervention has already begun.

Relatedly, we may also find that we can identify cases where our models are not working, based on instructor behavior. If an instructor chooses not to intervene in some cases, this may suggest that the instructor is recognizing an outlier or special case that our model cannot recognize. It may be possible to re-train and enhance our models based on this information. Even if a model is optimally accurate for its original training sample, it may become less accurate as the system it is in changes, as the academic culture around it changes, or as the populations it is used with shift. Human action will be most effective if the data and models provided to those humans is of the highest possible quality for the context of use. Building models that are robust to instructor behavior and that change as their context of application changes will become an essential challenge.

To sum up, the ultimate goal of the field of Artificial Intelligence in Education is not to promote artificial intelligence, but to promote education. The leading systems in AIED (in least in terms of degree of usage) seem to represent a different paradigm than the classic paradigm of intelligent tutoring systems. Reframing our research questions and perspectives in the light of this evidence may help us to better understand what we as a community are doing, and how we can be even more successful in doing it.

In the end, our goal is not to create intelligent tutoring systems or stupid tutoring systems, but to create intelligent and successful students.

References

Aleven, V., Sewall, J., Popescu, O., Xhakaj, F., Chand, D., Baker, R., Wang, Y., Siemens, G., Rosé, C., & Gasevic, D. (2015). The Beginning of a beautiful friendship? Intelligent tutoring systems and MOOCs. *Proceedings of the 17th International Conference on Artificial Intelligence in Education*, 525–528.

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

Arnold, K. E., Pistilli, M. D. (2012) Course Signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. New York, NY: ACM, pp. 267–270.

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., & Woolf, B. P. (2007). Repairing disengagement with non-invasive interventions. *Proceedings of the International Conference on Artificial Intelligence in Education*, 195–202.

Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education, 24*(4), 387–426.

Baker, R. S. J. d. (2010). Data mining for education. In B. McGaw, P. Peterson, & E. Baker (Eds.) *International Encyclopedia of Education* 3rd edn, (pp. 112–118). Oxford, UK: Elsevier.

Baker, R. S. (2014). *Big Data And Education*. New York, NY: Teachers College, Columbia University.

Baker, R.S. (2016) Stupid Tutoring Systems, Intelligent humans. Manuscript under review.

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In K. Sawyer (Ed.) *Cambridge Handbook of the Learning Sciences* 2nd edn, (pp. 253–274).

Baker, R. S. J. d., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17.

Baker, R. S. J. d., Corbett, A. T., Koedinger, K. R., Evenson, S. E., Roll, I., Wagner, A. Z., Naim, M., Raspat, J., Baker, D. J., & Beck, J. (2006). Adapting to when students game an intelligent tutoring system. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 392–401.

Barber, R., Sharkey, M. (2012) Course correction: using analytics to predict course success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, New York, NY: ACM, 259–262.

Beck, J.E. (1997) Modeling the Student with Reinforcement Learning. *Proceedings of the Machine learning for User Modeling Workshop at the Sixth International Conference on User Modeling.*

Beck, J.E., Mostow, J. (2008) How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 5091, 353–362.

Beck, J.E., Woolf, B.P., Beal, C.R. (2000) ADVISOR: A machine learning architecture for intelligent tutor construction. *Proceedings of the 7th National Conference on Artificial Intelligence,*New York, NY: ACM, 552–557.

Bergman, P. (under review). Parent-child information frictions and human capital investment: evidence from a field experiment. Manuscript under review. Working paper retrieved online 2/10/2016 from http://www. columbia.edu/~psb2101/BergmanSubmission.pdf.

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence: The International Journal of Research into New Media Technologies*, 14(1), 75–90.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Broderick, Z., O'Connor, C., Mulcahy, C., Heffernan, C., & Heffernan, N. (2010). Increasing parent engagement in student learning using an intelligent tutoring system. *Journal of Interactive Learning Research*, 22(4), 523–550.

Brooks, F. P. (1975). *The mythical man-month*. Boston, MA: Addison-Wesley.

Bull, S., Nghiem, T. (2002) Helping learners to understand themselves with a learner model open to students, peers and instructors. *Proceedings of the Workshop on Individual and Group Modelling Methods that Help Learners Understand Themselves*, 5–13. Biarritz, France.

Campbell, D. T. (1976). *Assessing the impact of planned social change*. Hanover, NH: Dartmouth College.

Carbonell, J. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems*, 11(4), 190–202.

Chi, M., VanLehn, K., Litman, D. (2010) Do micro-level tutorial decisions matter: Applying reinforcement learning to induce pedagogical tutoring tactics. *Proceedings of the 10th International Conference on Intelligent Tutoring Systems*, 184–193.

Clark, R. C., & Mayer, R. E. (2003). *E-learning and the science of instruction*. San Francisco: Jossey-Bass.

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

Corbett, A. T., & Bhatnagar, A. (1997). Student modeling in the ACT programming tutor: adjusting a procedural learning model with declarative knowledge. *User Modeling: Proceedings of the Sixth International Conference, UM97*, 243-254.

Corbett, A. T., Koedinger, K. R., & Hadley, W. (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, NJ: Lawrence Erlbaum Associates, Inc.

Craig, S. D., Hu, X., Graesser, A. C., Bargagliotti, A. E., Sterbinsky, A., Cheney, K. R., & Okwumabua, T. (2013). The impact of a technology-based mathematics after-school program using ALEKS on student's knowledge and behaviors. *Computers & Education*, 68, 495–504.

D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K. (2010) A time for emoting: when affect-sensitivity is and isn't effective at promoting deep learning. *Proceedings of the 10th International Conference on Intelligent Tutoring Systems*, 6094, 245–254.

Darling-Hammond, L. (2000) Teacher Quality and Student Achievement. *Education Policy Analysis Archives*. Available online at http://epaa.asu.edu/ojs/index.php/epaa/article/view/392

Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting student drop out: A case study. *Proceedings of the 2nd International Conference on Educational Data Mining* (EDM'09), 41–50.

Del Soldato, T., & du Boulay, B. (1995). Implementation of motivational tactics in tutoring systems. *Journal of Artificial Intelligence in Education*, 6(4), 337–378.

Diaz, G., Garcia Loro, F., Castro, M., Tawfik, M., Sancristobal, E., Monteso, S. (2013) Remote electronics lab within a MOOC: Design and preliminary results. *Proceedings of Experiment@ International Conference*, Redwood, CA: Deepdyve, 89–93.

D'Mello, S. K., & Graesser, A. C. (2012). AutoTutor and affective AutoTutor: learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems*, 2(4), 1–38.

Falmagne, J., Albert, D., Doble, C., Eppstein, D., & Hu, X. (2013). *Knowledge spaces: Applications in education*. Berlin-Heidelberg: Springer.

Farber, B., & Miller, J. (1981). Teacher burnout: A psycho-educational perspective. *Teachers College Record*, 83(2), 235–243.

Feng, M., & Heffernan, N. T. (2005). Informing teachers live about student learning: reporting in the assistment system. *Technology, Instruction, Cognition and Learning, 3*, 1–14.

Freund, Y. (2013) Artificial intelligence vs intelligence amplification. California Institute for Telecommunications and Information Technology. https://www.youtube.com/watch?v=sEGdszE86bY

Goldstein, I. J. (1979). The genetic graph: a representation for the evolution of procedural knowledge. *International Journal of Man-Machine Studies, 11*(1), 51–77.

Greeno, J. G. (1997). On claims that answer the wrong question. *Educational Researcher, 26*(1), 5–17.

Hawn, A. (2015). The bridge report: bringing learning analytics to low-income, urban schools. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge.* Washington, DC: Association for Computing Machinery, 410–411.

Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education, 24*(4), 470–497.

Heffernan, N.T., Koedinger, K.R. (2002) An intelligent tutoring system incorporating a model of an experienced human tutor. *Proceedings of the 6th International Conference on Intelligent Tutoring Systems, 2363*, 596–608.

Jayaprakash, S. M., Moody, E. W., Lauria, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early alert of academically At-risk students: An open source analytics initiative. *Journal of Learning Analytics, 1*(1), 6–47.

Jeong, H., Biswas, G. (2008) Mining Student Behavior Models in Learning-by-Teaching Environments. *Proceedings of the 1st International Conference on Educational Data Mining,* Worcester, MA: International Educational Data Mining Society, 127–136.

Joksimović, S., Kovanović, V., Jovanović, J., Zouaq, A., Gašević, D., & Hatala, M. (2015). What do cMOOC participants talk about in Social Media? A Topic Analysis of Discourse in a cMOOC. *Proceedings of the 5th International Learning Analytics and Knowledge (LAK) Conference,* New York, NY: ACM, 156–165

Khachatryan, G., Romashov, A., Khachatryan, A., Gaudino, S., Khachatryan, J., Guarian, K., & Yufa, N. (2014). Reasoning mind Genie 2: An Intelligent learning system as a vehicle for international transfer of instructional methods in mathematics. *International Journal of Artificial Intelligence in Education, 24*(3), 333–382.

Koedinger, K. R., McLaughlin, E. A., & Heffernan, N. T. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. *Journal of Educational Computing Research, 43*(4), 489–510.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012a). The knowledge-learning-instruction framework: bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*, 757–798.

Koedinger, K. R., McLaughlin, E. A., & Stamper, J. C. (2012b). Automated student model improvement. *Proceedings of the 5th International Conference on Educational Data Mining,* 17–24.

Koedinger, K.R., Stamper, J., McLaughlin, E., Nixon, T. (2013) Using Data-Driven Discovery of Better Student Models to Improve Student Learning. *Proceedings of the International Conference on Artificial Intelligence and Education, 7926*, 421–430.

Kovačić, Z. (2010). Early Prediction of Student Success: Mining Students Enrolment Data. *Proceedings of Informing Science & IT Education Conference (InSITE),* 647–665.

Kulik, C. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: an updated analysis. *Computers in Human Behavior, 7*, 75–95.

Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: the Betty's brain system. *International Journal of Artificial Intelligence in Education, 18*(3), 181–208.

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J. (1993). Motivational techniques of expert human tutors: lessons for the design of computer-based tutors. *Computers as Cognitive Tools, 1993*, 75–105.

Matsuda, N., Keiser, V., Raizada, R., Tu, A., Stylianides, G., Cohen, W. (2010) Learning by Teaching SimStudent: Technical Accomplishments and an Initial Use with Students. *Proceedings of the 10th International Conference on Intelligent Tutoring Systems, 6094*, 317–326.

Maull, K. E., Saldivar, M. G., & Sumner, T. (2011). Online curriculum planning behavior of teachers. *Proceedings of the 3rd International Conference on Educational Data Mining,* 121–130.

McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction, 7*(3), 197–244.

McAuley, A., Stewart, B., Siemens, G., Cormier, D. (2010) The MOOC model for digital practice. Available online at http://www.davecormier.com/edblog/wp-content/uploads/MOOC_Final.pdf

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica, CA: RAND Corporation.

Merrill, D. C., Reiser, B. J., Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences, 2*(3), 277–305.

Miller, W. L., Baker, R., Labrum, M., Petsche, K., Liu, Y-H., Wagner, A. (2015). Automated detection of proactive remediation by teachers in reasoning mind classrooms. *Proceedings of the 5th International Learning Analytics and Knowledge Conference*, 290–294.

Ming, N.C., Ming, V.L. (2012) Automated Predictive Assessment from Unstructured Student Writing. *Proceedings of the 1st international Conference on Data Analytics.*

Mitrovic, A., & Ohlsson, S. (1999). Evaluation of a constraint-based tutor for a database language. *International Journal of Artificial Intelligence in Education*, 10, 238–256.

Mostow, J. (2008). Experience from a Reading tutor that listens: evaluation purposes, excuses, and methods. In C. K. Kinzer, & L. Verhoeven (Eds.), *Interactive Literacy Education: Facilitating Literacy Environments Through Technology* (pp. 117–148). New York: Lawrence Erlbaum Associates, Taylor & Francis Group.

Mulqueeny, K., Kostyuk, V., Baker, R. S., & Ocumpaugh, J. (2015). Incorporating effective e-learning principles to improve student engagement in middle-school mathematics. *International Journal of STEM Education, 2*(15).

Murray, R.C., VanLehn, K. (2005) Effects of Dissuading Unnecessary Help Requests While Providing Proactive Help. *Proceedings of the International Conference on Artificial Intelligence in Education*, Amsterdam, Netherlands: IOS Press, 887–889

Norvig, P. (2012) Google's Peter Norvig on Online Education. Public lecture given at Stanford University. http://www.youtube.com/watch?v=1ylzPGD3jRY

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, 24(4), 427–469.

O'Shea, T. (1982). A self-improving quadratic tutor. In D. Sleeman, & J. S. Brown (Eds.), *Intelligent Tutoring Systems* (pp. 309–336). New York: Academic Press.

Ostashewski, N. (2013) Building for massive scalability: the production case of an astronomy MOOC. *Proceedings of the World Conference on E-Learning in Corporate*, *Government*, *Healthcare*, *and Higher Education*, Chesapeake, VA: AACE

Ostrow, K. S., & Heffernan, N. T. (2014). Testing the multimedia principled in the real world: A comparison of video vs. In *Text feedback in Authentic Middle School math assignments. Proceedings of the 7th International Conference on Educational Data Mining* (pp. 296–299).

Ostrow, K. S., Schultz, S. E., & Arroyo, I. (2014). Promoting Growth Mindset Within Intelligent Tutoring Systems. In CEUR-WS (1183), Gutierrez-Santos, S., & Santos, O.C. (eds.) *EDM 2014 Extended Proceedings*, 88–93.

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.

Papousek, J., Pelanek, R., & Stanislav, V. (2014). Adaptive practice of facts in domains with varied prior knowledge. *Proceedings of the International Conference on Educational Data Mining (EDM)*, 6–13.

Pardos, Z. A., Baker, R. S. J. D., Gowda, S. M., & Heffernan, N. T. (2011). The sum is greater than the parts: ensembling models of student knowledge in Educational software. *SIGKDD Explorations*, 13(2), 37–44.

Pashler, H., Bain, P. M., Bottqe, B. A., Graesser, A., Koedinger, K., McDaniel, M., Metcalfe, J. (2007). *Organizing instruction and study to improve student learning.* IES practice guide NCER 2007–2004. Washington, DC: US Department of Education, Institute of Education Sciences.

Pavlik, P.I., Cen, H., Koedinger, K.R. (2009) Performance Factors Analysis – A New Alternative to Knowledge Tracing. *Proceedings of the International Conference on Artificial Intelligence in Education*, Amsterdam, The Netherlands: IOS Press, 531–538

Quinn, A.J., Bederson, B.B. (2011) Human computation: a survey and taxonomy of a growing field. *Proceedings of ACM SIGCHI*, 1403–1412, New York, NY: ACM

Rafferty, A. N., Brunskill, E., Griffiths, T.L., Shafto, P. (2011) Faster teaching by POMDP planning. *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, 6738, 280–287.

Raghuveer, V.,Tripathy, B., Singh, T., Khanna, S. (2014) Reinforcement learning approach towards effective content recommendation in mooc environments. *Proceedings of IEEE International Conference on MOOC, Innovation and Technology in Education* (*MITE*), 285–289, Patiala: IEEE

Razzaq, L., & Heffernan, N. T. (2006). Scaffolding vs. In *Hints in the assistment system. Proceedings of the 8th International Conference on Intelligent Tutoring Systems (ITS 2006)* (pp. 635–644).

Rodrigo, M. M. T., Baker, R. S. J. d., Rossi, L. (2013). Student off-task behavior in computer-based learning in the Philippines: comparison to prior research in the USA. *Teachers College Record, 115*(10), 1–27.

Rodriguez, C. O. (2012). MOOCs and the AI-Stanford like courses: two successful and distinct course formats for massive open online courses. *European Journal of Open, Distance and E-Learning, 2012*(2).

Schofield, J. (1995). *Computers and classroom culture*. London: Cambridge University Press.

Self, J. A. (1990). Theoretical foundations of intelligent tutoring systems. *Journal of Artificial Intelligence in Education*, 1(4), 3–14.

Self, J. A. (1998). The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. *International Journal of Artificial Intelligence in Education (IJAIED)*, *10*, 350–364.

Serrano-Laguna, A., & Fernández-Manjón, B. (2014). Applying learning analytics to simplify serious games deployment in the classroom. In *IEEE International Global Engineering Education Conference (EDUCON)*, Istanbul, 872–877.

Shute, V. (1990). *Rose garden promises of intelligent tutoring systems: blossom or thorn? paper presented at the space*. Applications and Research (SOAR) Symposium: Operations.

Siemens, G., Baker, R.S.J.d. (2012) Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, New York, NY: ACM

Slotta, J. D., Tissenbaum, M., & Lui, M. (2013). Orchestrating of complex inquiry: three roles for learning analytics in a smart classroom infrastructure. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge, 270–274*. New York, NY: ACM.

Stevens, A., & Collins, A. (1977). *The goal structure of a Socratic tutor. Technical Report #3518*. Cambridge, MA: Bolt, Beranek, and Newman, Inc..

Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2014). Supporting teachers in guiding collaborating students: effects of learning analytics in CSCL. *Computers & Education*, *79*, 28–39.

van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of cognitive activities during student collaboration: effects of learning analytics. *Computers & Education*, *90*, 80–94.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., & Wintersgill, M. (2005). The. *Andes physics tutoring system: Lessons learned, International Journal of Artificial Intelligence in Education*, *15*(3), 1–47.

Walonoski, J. A., & Heffernan, N. T. (2006). Prevention of off-task gaming behavior in intelligent tutoring systems. *Proceedings of the International Conference on Intelligent Tutoring Systems*. Berlin-Heidelberg: Springer, 722–724.

Wenger, E. (1987). *Artificial Intelligence And Tutoring Systems*. Los Altos, CA: Morgan Kauffman.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition*. Reading, MA: Addison-Wesley.

Wood, D.S., Williams, G. (2013) Data-Driven Factors that Increase Student Course Completion: A Two-Year Study. *Proceedings of the 9th Annual Symposium on Student Retention*.

Zapata-Rivera, D., & Katz, I. (2014). Keeping your audience in mind: applying audience analysis to the design of interactive score reports. *Assessments in Education: Principles, Policy, and Practice*, *21*(4), 442–463.