High-Leverage Opportunities for Learning Engineering

Ryan S. Baker Ulrich Boser

Allison Shelley for American Education: Images of Teachers and Students in Action

Penn Center for Learning Analytics
University of Pennsylvania

Contents

Executive Summary

There is increasing interest in developing the discipline and practice of learning engineering to improve student outcomes. Learning engineering combines scientific knowledge and theory on learning, and applies a rigorous combination of theory, data, and analysis to develop and improve educational systems and methodologies to produce enduring, high-quality learning. Learning Engineering is most frequently applied in educational technologies, which are increasingly used for learners of all ages. In terms of academic disciplines, learning engineering brings together a combination of computer science, data science, cognitive psychology, behavioral science, education and instructional design. Though the idea of learning engineering was first introduced by Herbert Simon in the 1960s, the uptake of learning engineering by the broader field of education has been slow, both among researchers and educational technology developers. However, today, the use of learning engineering has increased considerably, supported by a leap forward in the instrumentation of learning environments, the advent and low cost of cloud computing, and advances in data science methods.

Prior efforts have demonstrated that a data-intensive learning engineering approach has the potential to revolutionize education. In ASSISTments, for instance, Worcester Polytechnic Institute's Neil Heffernan has shown that crowdsourcing hints from teachers has a clear and robust positive effect on student outcomes (Patikorn & Heffernan, 2020). Similarly, OpenStax rolled out embedded retrieval practice to help students review material, weaving new and old concepts into the review process, leading to substantial gains on the final exam (Butler et al., 2014). Many platforms have rolled out infrastructures that offer rich support for personalization, such as Newsela, which offers the same reading assignment at different reading levels. At a more macro level, predictive analytics platforms such as Civitas have used predictions of which students are at risk of dropping out of college (and why) to support interventions that lead to increases in student on-time graduation (Milliron et al. 2014). However, the benefit of this approach has not yet extended to the full range of student learning experiences, nor to a broad range of learning software developers and educational organizations.

In this report, we discuss the potential of learning engineering to bring the theory and practice of learning forward, including both key areas where learning engineering can bring benefits, and key steps towards making those benefits a reality. We also discuss some of the challenges that have slowed uptake of learning engineering, and how they can be addressed. To summarize, this report identifies the top areas within the science and engineering of learning that with bigger and better data, improvements in algorithmic approaches, and more support for experimental research, could lead to major improvements in educational outcomes.

Ten Key Areas of Opportunity for Learning Engineering

Better Learning Engineering

Opportunity No. 1
Enhance R&D Infrastructure in Widely Deployed Platforms

A wide variety of online learning platforms are used by K-12 and college students in the U.S. every day, but only a small number of these platforms are engaged in ongoing processes of iterative improvement to benefit learners or learning science research. Despite often having an interest in R&D, most platforms make very limited use of their data and do not share their data with the broader research community. Most platforms do not have capacity for automated experimentation. Developing infrastructure and tools so that the millions of educational practitioners and researchers in the U.S. can conduct research and study educational improvement with the thousands of scaled learning systems, and mechanisms to translate those findings into improved educational technologies, is a key opportunity for learning engineering.

Opportunity No. 2
Bring Learning Engineering to Domain-Based Education Research

Adaptive learning systems -- and learning systems and curricula in general -- depend on a high-quality model of the content and learning needs in a given domain, including both the structure of the domain -- which skills are prerequisite to other skills -- and the misconceptions and conceptual misunderstandings that students struggle with. Support for tools to discover these models more efficiently from existing data, and support for sharing these models publicly, would reduce the substantial amount of duplicated effort seen today. Similarly, more should be done to identify the skills and knowledge that are most relevant to future learning, to make learning more efficient for students who need to catch up.

Opportunity No. 3
Build Components to Create Next Generation Learning Technologies Faster

Currently, developing a new learning platform with advanced adaptivity takes years of effort, limiting entrants to the field and leading to a proliferation of lower-quality platforms. There is substantial duplication of effort across learning platforms. The average quality of future learning systems could be substantially increased by creating reusable components for generally-applicable development tasks such as student modeling, modeling complex competencies, mindset interventions, and the educational applications of natural language processing.

Support Human Processes

Opportunity No. 4
Enhance Human-Computer Systems

Computers and humans have different strengths -- by using each for what they are best at, we can obtain better results than either alone. The learning engineering challenge is to help them work better together -- using computers for routine and repetitive parts of instruction, empowering teachers and tutors with more complete information from the computer, and developing technology that knows when to loop in a tutor or teacher when the learner isn't making progress. Currently, development of learning systems often focuses on the interaction between the student and the technology, but prototype examples of design demonstrate what is possible when more attention is paid to designing for teachers as part of a more complex and comprehensive learning system.

Opportunity No. 5
Better Engineer Learning System Implementation in Schools

Many learning systems and curricula work well under favorable conditions -- motivated teachers and supportive administration, with significant involvement from the developers in teacher professional development and ongoing support as the system is being used. However, these same learning systems and curricula often fail when extended to a broader range of classrooms, where teachers may be unfamiliar or uncomfortable with new methods and technologies, and may attempt to assimilate new technologies back into traditional teaching practices. Learning engineering can play a role in determining which practices around the use of learning technology are both effective and scalable, monitoring through data whether these practices are occurring, and designing automated or semi-automated methods to help teachers use the right practice at the right time. For example, a pop-up message in a dashboard might encourage a teacher to speak with one student who has been struggling for the last thirty minutes and to provide a peer tutor to a different student who might be uncomfortable with direct teacher feedback.

Opportunity No. 6
Improve Recommendation, Assignment, and Advising Systems

Despite the increasing use of dropout prediction and course failure prediction technologies in schools and universities, course selection and registration processes at many institutions remain passive, leaving it to students to identify and select courses with minimal support, despite a substantial body of research in this area and large datasets from prior students. This results in many students taking "excess" credits in college or community college that don't count towards degree requirements and use up financial aid. Learning engineering could be used to develop advising systems that proactively analyze student trajectories and make recommendations (to advisors or to the students themselves) that increase the likelihood of successful graduation. Relatedly, predictive models are not built into K-12 school assignment, leading to students being assigned to schools that they are dissatisfied with and/or perform poorly at. Much better results are possible if prediction algorithms are used to help students select schools that are appropriate for them, and help match students to schools where they are likely to succeed.

Better Learning Technologies

Opportunity No. 7
Optimize for Robust Learning and Long-Term Achievement

Most of the research in learning engineering focuses on rapid innovation cycles that improve short-term learning gains and on skills that are simple to assess. While this work is important, learning engineering also needs to push towards longitudinal work that verifies that developers are selecting designs and algorithms that benefit students over a longer period of time, and that students are not simply being trained for specific skills but are being prepared to learn in new situations.

Opportunity No. 8
Support Learning 21st Century Skills and Collaboration

Much of the learning technology currently in use focuses on relatively narrow academic skills, but more complex skills such as communication, critical thinking, and collaboration -- often referred to as “21st-century skills”-- will be key to career and life success in the coming decades. Using new technologies and data collection, in combination with analytics, learning engineering can focus on the development of reliable and valid measures of these hard to measure constructs, and produce learning experiences that support their development. For instance, natural language processing could be used to measure collaboration in a MOOC, or machine learning could be used to track “grit” within an online platform.

Opportunity No. 9
Improved Support for Student Engagement

Student affect and behavioral disengagement are amenable to measurement, and are associated with differences in learner outcomes over a decade later. However, though the technology exists to measure engagement and affect, there has been considerably less effort to use these measurements to improve engagement and affect. Though a small number of pilot studies have been effective at improving engagement and learning, these approaches have not scaled. Learning engineering needs to be used to figure out which engagement/affective interventions (both teacher-driven and automated) are effective for which students, in which situations. In parallel, work is needed to figure out how to design these interventions in ways that teachers, school leaders, parents, and students are comfortable with.

Opportuinty No. 10
Design Algorithms and Learning Systems for Diversity and Equity

Recent evidence suggests that many research findings on learning technologies and machine-learned models which are obtained on convenience samples do not generalize to broader and more diverse groups of learners. As a field, learning engineering needs to incorporate equity as a foundational principle: to understand which differences between learners matter, and how these differences impact the effectiveness of learning technologies. A key part of this is collecting much more complete data on learner diversity, and checking models and findings in terms of it. Race, ethnicity, second-language status, gender, neurodiversity, disability status, urbanicity, and military-connected status can all impact effectiveness, but are often not collected or analyzed.

Ten Key Areas of Opportunity for Learning Engineering

Recommendations Table

Top Ten Opportunity	Example
Better Learning Engineering	
Opportunity No. 1 **Enhance R&D Infrastructure in Widely Deployed Platforms**	Make high-quality data available for a broader range of platforms Develop an ecosystem where researchers can more easily build on each others' findings and research code Develop general-purpose software components for identifying how effective content is Extend experimentation infrastructure to a broader range of learning platforms, along with good tools for authoring content for studies Extend experimentation testing infrastructure to study the effectiveness of combined interventions Develop general-purpose software components for reinforcement learning Embed measures of student affect, self-regulated learning, and engagement into learning engineering platforms Support the development of easier processes and technologies for IRB and privacy compliance for learning platforms
Opportunity No. 2 **Bring Learning Engineering to Domain-Based Educational Research**	Create a network to incentivize and scaffold widespread sharing and collaboration on domain knowledge gaps Fund support for hybrid AI/human methods for knowledge graph discovery Support infrastructure for discovering and remedying student misconceptions

Ten Key Areas of Opportunity for Learning Engineering

Top Ten Opportunity	Example
Opportunity No. 3 **Build Components to Create Next-Generation Learning Technologies Faster**	Create production-grade components for student modeling that can be integrated into different learning systems and used at scale Support research on the data needs and practical limitations of modern student modeling algorithms Create reusable components for interventions such as mindset interventions Develop production-grade toolkits to facilitate modeling complex student behavior Develop toolkits for natural language processing in education
Support Human Processes	
Opportunity No. 4 **Enhance Human-Computer Systems**	Increase the richness of data given to teachers, while maintaining usability and comprehensibility Provide teachers with real-time recommendations about when to provide additional support to students, and what kind of support to provide Support research on integration of computer tutoring and human tutoring
Opportunity No. 5 **Better Engineer Learning System Implementation in Schools**	Improve integration of data between classroom practices, students' learning experiences, and teacher professional development to study which practices around the use of learning technology are effective and scalable. Develop a taxonomy of teacher practices around the use of learning technology, and use it to study which practices and professional development is effective and scalable. Develop automated and semi-automated methods to encourage teachers to use the right practice at the right time
Opportunity No. 6 **Improve Recommendation, Assignment, and Advising Systems**	Develop advising and recommendation systems that support better advising practices Design explainable AI methods for re-purposing prediction models into easy-to-understand recommendations for advisors and students Fund infrastructure that enables experimentation around prediction and recommendation and connects it with outcome data

Ten Key Areas of Opportunity for Learning Engineering

Top Ten Opportunity	Example
Better Learning Technologies	
Opportunity No. 7 **Optimize for Robust Learning and Long-Term Achievement**	Increase awareness of existing cognitive science findings around robust learning Incentivize and plan for longer-term follow-up for A/B studies
Opportunity No. 8 **Support Learning 21st-Century Skills and Collaboration**	Develop data science challenges to drive competition to create reliable and valid measures of 21st-century skills, including collaboration, using new technologies and data collection methods Develop data science challenges to drive competition to create learning systems that scaffold collaboration, and support the development of 21st-century skills
Opportunity No. 9 **Improved Support for Student Engagement**	Examine which engagement/affective interventions (both teacher-driven and automated) are effective for which students, in which situations Create a competition where engagement/affective interventions are combined and compared, in a sample large enough to also study individual differences Develop better understanding of teacher and student preferences and comfort for engagement/affective interventions
Opportunity No. 10 **Design Algorithms and Learning Systems for Diversity and Equity**	Require that projects collect more complete data on learner identity and characteristics Require that projects check models and findings for algorithmic bias and differential impact Encourage participatory and inclusive design, involving members of the communities impacted

Our Approach

In 2020, a community came together to discuss the potential of the emerging field of learning engineering and key challenges and opportunities for the field. This report represents the findings and discussions from this community convening.

Due to the COVID-19 pandemic, and the impossibility of traveling to hash out ideas together in person, the organizing team chose an unusual structure. This structure lacked face-to-face discussion and large group discussion, but enabled the organizers to solicit the opinions of individuals worldwide, at times of their convenience. It also allowed for multiple rounds of soliciting different opinions (allowing us to realize, in many cases, that a key perspective was missing, and then solicit it).

In this asynchronous virtual convening, a set of questions was posed to a group of around 30 researchers, learning system developers, policy-makers, and thought-leaders. Some of these stakeholders were supported in assembling into small groups (taking both interests and time zones into account) and met virtually to propose ideas on the future of the field. Other stakeholders met one-on-one with the organizers of the report, or offered written comments. The organizers followed up with clarifying questions on some of the most intriguing ideas, and put together a report summarizing the findings. The result, this report, attempts to represent the perspective of many stakeholders, while bringing it into a single voice and set of coherent recommendations.

What is Learning Engineering?

Teachers, policy makers, researchers, and parents have wanted to know which practices are best for understanding and supporting learning since before there was an established science of learning. However, even as the science of learning has emerged over the last decades (Hoadley, 2018), there is still a considerable gap between theories and findings on learning, and the curricula and approaches that are used in the real-world (Uncapher & Cheng, 2019).

There is increasing interest in building the discipline of learning engineering, combining scientific knowledge on learning with rigorous practice and data to develop and improve educational systems and methodologies to produce enduring, high-quality and efficient learning. Put differently, learning engineering is a field of study at the intersection of computer science and learning science, and the field aims to harnesses the power of big data and learning analytics to gain insights on how students learn, to improve learning within various educational systems, and to inform evidence-based decision making around the design of learning activities.

Learning engineering is a new and emerging field that builds off of work from across various domains in the education, technology and science space. Such fields include instructional design, curriculum development and evaluation, learning analytics, computer science, and learning science.

As Dede, Richards, and Saxberg (2018) note, learning engineering it not just about better designs -- it is also about efficiency:

> "...it is very valuable to find more efficient ways to reach the same levels of mastery–engineering is about effectiveness and efficiency, within the constraints of real-world delivery, to free up resources to do even more to help learners."

Key Accomplishments of Learning Engineering

While the field of learning engineering is still emerging and being codified, recent work and previous work that could be classified as learning engineering have already begun to contribute significantly to our understanding of how to best design, measure and implement learning systems. Learning engineering's blend of theoretical and algorithmic innovation, combined with a focus on developing and testing systems in a real-world setting with the goal of making a real impact on learning in the classroom and related high-stakes objectives, has led to substantial advances in learning science and technologies.

One such contribution is the development and implementation of theoretical frameworks that provide systematic and structured guides on how to best implement research findings in an applied setting. These frameworks are designed to help guide instruction and/or instructional design, providing instructional principles to practitioners and researchers that are based upon empirical evidence. Perhaps the most widely-used framework is the Knowledge Learning Instruction Framework (KLI; Koedinger, Corbett, Perfetti, 2012), which provides researchers and practitioners a systematic rigorous framework for examining the interaction between changes in instructional events and students' transfer, retention, and preparation for future learning. Work within the KLI framework has investigated how practices interact, and which combination of practices are most appropriate (Koedinger, Booth, & Klahr, 2013), as well as applying KLI principles to the study of new environments such as massive online open courses (Koedinger et al., 2016), extending beyond the original team that developed KLI (e.g. Bergamin & Hirt, 2018; Borracci et al., 2020).

Learning engineering (and its antecedent work) has also led to the development of paradigms for learning system and content development, which provide an overall approach to learning design. For instance, cognitive task analysis breaks down tasks into components to better understand and identify what skills, knowledge, and actions are needed to complete the task at an acceptable performance level (Lovett, 1998). Constraint-Based Modeling is used to support students in learning material where there may be multiple correct answers by representing the features of a correct answer rather than the process of finding it, bridging naturally into giving students feedback on the errors that they make during problem solving (Mitrovic, 2010). Knowledge graphs/spaces provide an ontology or model of the knowledge and skills needed for a specific task or domain knowledge (Doignon & Falmagne, 2012), and are used to select what content a student should work on next. Instruction modeling is the practice of engineering automated replications of the instructional practices of expert teachers (Khachatryan, 2020).

A Practical Improvement: Improved Feedback

One of the major successes of learning engineering and the learning platforms it supports has been the provision of useful feedback to learners. Providing timely, accurate feedback is crucial for student success (Razzaq et al., 2020), but is often prohibitively time consuming for teachers. By automating feedback and using learning engineering to improve it iteratively, it is possible to scalably give students feedback that is adaptive to a student's current state of knowledge in a domain and topic, at the time which is best for their learning (McBroom et al., 2018; Pardo et al., 2019).

Multiple approaches to automated feedback have found success in learning environments. Technology can provide teachers with recommendations on how to guide students based on student performance (e.g., Ingebrand & Connor, 2016). Automated student feedback and adaptive scaffolding have led to learning gains for students (e.g., Kim et al., 2018; Kroeze et al., 2019; Zhu et al., 2020). Student-facing dashboards are one location where such feedback can be given. For example, one study explored the impact of a dashboard that provided real-time feedback of their digital engagement with course material, giving alerts when engagement was detected as low (Khan & Pardo, 2016).

There are a number of dimensions along which the design of feedback can vary, and the best approach to feedback -- the best combination of design choices -- may vary by context (Koedinger, Booth, & Klahr, 2013). Numerous factors have been studied to determine what forms of feedback are optimal for different learners and situations. For example, one study determined that the nuanced revising behaviors of writers (i.e., patterns of deletion, addition, and keystrokes) can inform adaptive feedback to promote better writing output (Roscoe et al., 2015). Another study found that responses to feedback varied significantly across grade levels but that superficial changes to feedback messages were not impactful (Howell et al., 2018). Feedback systems that account for students' affective states can enhance engagement and, ultimately, learning by providing personalized prompts and activity recommendations (Grawemeyer et al., 2017).

Educational Data Mining and Learning Analytics: Opportunities for learning engineering

A final major area of contribution from learning engineering is the growth of the educational data mining/learning analytics community. Although seen by some as a separate field from learning engineering, there is a large overlap in the broader motivations as well as individual scientists and practitioners between learning engineering and learning analytics. The key role of data and analytics in recent advances in learning engineering suggests that, at minimum, if they are separable fields, they function best together closely, like peanut butter and jelly.

While Learning Engineering has focused on well-defined knowledge domains, Learning Analytics has been focused on broader processes of educational attainment and student risk of not persisting in their educational goals. A hallmark of this work has been the analysis of the log data from educational technology systems and using that data to make inferences about students. Educational Data Mining has focused more specifically around algorithmic development that contributes to advances in the learning sciences and to automated adaptation; added together, these areas have major promise to contribute to our educational understanding and practices.

Data and models of it have played a major role both in refining learning systems, and in creating algorithms that can be used to underpin personalization and adaptivity. Traditionally learning and student progress -- and a system's degree of success in supporting these -- was measured using delayed, distal, and loosely-aligned information such as grades and standardized test scores. The move towards measuring learning and other forms of progress using log data has allowed the development of measures which are immediate, proximal, and tightly-aligned to the learning experience. The potential of log data is also being recognized by psychometricians in the analysis of standardized assessment results (i.e. Bergner & Von Davier, 2019)

Simply developing the ability to measure learning as it changed (i.e. Corbett & Anderson, 1995), was a step that enabled mastery learning, the underpinning of many modern adaptive learning systems. Going beyond that to measuring complex learning and performance in real-time (Gobert et al., 2013; Rowe et al., 2017) enabled learning systems such as Inq-ITS (Li et al., 2018) to provide feedback and support on complex skills such as scientific inquiry. Going further still, recent experimental systems measure and attempt to support students in learning to self-regulate their strategy (Roll et al., 2011; Duffy & Azevedo, 2015) and affect (Karumbaiah et al., 2017; DeFalco et al., 2018). Better measurement has also supported efforts to iteratively engineer learning systems (Aleven et al., 2017; Huang et al., 2020), for instance by identifying where skills are mis-specified (Corbett & Anderson, 1995), or by systematically searching for less-effective learning content (Baker, Gowda, & Salamin 2018).

In the end, it's clear why a data-intensive learning engineering approach can be so powerful. It shifts the very nature of instruction, from intuitive and uncertain, to precise and iterative. Recent interventions have shown an impact on outcomes. While almost all of these interventions are relatively small, they show the potential for the field. One study showed how automated content selection improved outcomes (Wilson & Nichols, 2015); others found that providing support for engagement led to better learning (Baker et al., 2006; Karumbaiah et al., 2017); and work has shown how providing recommendations to instructors can help them select material for their students to work on (Zou et al., 2019).

Other approaches can show what does not work. Using UpGrade from Carnegie Learning, the Playpower Labs team found that adding a specific set of "gamification" features actually reduced learner engagement by 15 percent. Similarly, data has helped identify key context variables within the COVID-19 pandemic. Data from the Zearn platform, for instance, was used by external researchers to study the impacts of COVID-19. The researchers found that COVID-19 caused student progress to skew dramatically by parent wealth, with learning decreasing by approximately half for learners in low-income areas, whereas learning dipped temporarily but quickly returned to normal for students in high-income areas (Chetty et al., 2020).

Top High-Level Opportunities

Better Learning Engineering: Enhance R&D Infrastructure in Widely Deployed Platforms

In the following sections, we provide ten high-level recommendations as to where the high-leverage opportunities are for learning engineering, emerging from our discussions with stakeholders in the AVCs and additional meetings. We group these recommendations into three broad categories: **enhancing learning engineering, supporting human processes, developing better learning technologies.**

One major step towards increasing the positive impact of learning engineering is to improve the tools available for conducting learning engineering. If learning engineering research can be conducted faster, with higher quality, at less effort, the benefits of learning engineering can correspondingly become available faster and to a larger number of research and development teams. There has already been considerable investment in research and development infrastructure for learning engineering, in no small part due to the efforts of Schmidt Futures, and the impacts are seen in the rapid deployment of studies through the ASSISTments platform (Ostrow & Heffernan, 2016) compared to the previous relatively intensive LearnLab model (Koedinger, Corbett, & Perfetti, 2012), which in turn was much faster and easier for researchers than developing all infrastructure and arranging each research study one by one. Developing infrastructure and tools so that the millions of educational practitioners and researchers in the U.S. can use better, faster methods to conduct research and study educational improvement with the thousands of scaled learning systems is a key opportunity for learning engineering. One could argue that all of the other opportunities for learning engineering that this document will discuss will be supported by improving the infrastructure for learning engineering. There are several opportunities in this area.

First, there are opportunities around increasing the scope of educational data available and the tools for collaboration among researchers. One of the key developments facilitating the uptake of educational data mining (EDM) methods has been the wide availability of high-quality data -- however, these benefits have been spread unevenly, with the majority of research conducted on data from a small number of platforms, ASSISTments and MATHia/Cognitive Tutor in particular as well as a number of MOOCs.

Simply making data available for a broader range of platforms would amplify the positive impact on the field. Going further, the creation of an ecosystem where researchers can more easily build on each others' findings and research code would speed work compared to today, where even when researchers share their code, it is often difficult to get it to run correctly (Boettiger, 2015). While both LearnSphere (Liu et al., 2017) and the MORF platform (Gardner et al., 2018) have attempted to create ecosystems that function in this fashion, neither platform's ecosystem has achieved widespread use, due both to limitations in these platforms that require further engineering (such as constraints on what kinds of research is possible in these platforms), and a lack of incentives for researchers to share their code in this fashion.

Part of the benefit of large datasets is that they can help find the complex answers that are needed to resolve long-standing "dilemmas" in education, such as the assistance dilemma (Koedinger & Aleven, 2007). In this instance, large-scale data with variation in what assistance students are offered could help researchers better understand what support is helpful to which students and under what conditions. Does the assistance needed differ for students who generally have very low prior knowledge (cf. Roll et al., 2014)? Does it differ when students have challenges with procedures versus conceptual misunderstandings (cf. VanLehn, 1996)? When is a worked example better than a hint (cf. McLaren et al., 2008)? Do students from different cultural backgrounds respond differently and benefit differently from the same seemingly-cognitive learning support (cf. Ogan et al., 2015)? Can we develop generative and broadly-applicable theory, that works across learning contexts? Large and diverse data sets can help.

In addition, these datasets can also catalyze "benchmark" challenges that researchers and technologists compete on, and incentivize advancements in both fundamental and domain-specific aspects of education. Data competitions around automated essay scoring and future performance prediction have attracted a large number of competitors and produced scientific contributions and technological advancements (Stamper & Pardos, 2016; Taghipour & Ng, 2016); many areas appear ripe for such competitions. For instance, images around math handwriting seems ripe for a competition, given developments in OCR and the amount of handwriting in math. Similarly, the use of voice recognition to identify students struggling to learn to read could benefit from the availability of a benchmark dataset. In this instance, the field needs audio files of younger readers as well as "ground truth" data on future reading issues.

EDM methods can also help to drive progressive refinement and improvement of learning technologies. For example, several online learning platforms now automatically distill information from their student models to determine if specific skills are ill-specified (Agarwal et al., 2018) or if specific learning content is more effective or less effective than the average (Baker et al, 2018). However, these types of approaches are, again, seen only in a small number of learning platforms. Creating infrastructure that can be widely adopted -- such as packages that take data formatted in a standard fashion and generate automated reports -- may speed the adoption of this type of practice.

Second, A/B testing and other forms of rapid automated or semi-automated experimentation (see review in Motz et al., 2018) make it possible to quickly ask questions about learning. Currently, this type of technology is only used within a small number of platforms and studies (see review in Salvi et al., 2017), although its scope is expanding over time. Extending this infrastructure to a broader range of learning platforms, along with good tools for authoring content for studies, would facilitate research on a variety of questions. The same experimental infrastructure that supports scientific discovery may also hold benefits for refining and progressively improving a learning system.

This infrastructure may help us to understand not only whether a specific intervention works, but which interventions work in combination. Over the last decades, there has been a great deal of work investigating single interventions, but considerably fewer studies of whether two interventions thought to impact students in the same fashion (or in different fashions) have additive effect or are in fact counterproductive when combined (Koedinger, Booth, & Klahr, 2013) -- a question that needs considerable data to answer. In these cases, strategies such as reinforcement learning (Zhou et al., 2020) may help a learning system decide what intervention to offer which student, in which situation; creating software packages that implement these algorithms and can be plugged into a variety of learning platforms will speed the process of improving learning platforms.

This sort of experimentation should not be limited to researchers. Teachers can improve their teaching practice by conducting more systematic investigations of their own teaching, a practice seen for decades in other countries (Stigler & Hiebert, 1999). Partnering teachers with learning engineers in these endeavors will benefit both groups. To this end, the Learning Agency Lab has pulled together a cadre of teachers who are using RCTs and other high-quality research approaches to better understand the science of learning. The effort leverages the advent of robust data systems and easy-to-conduct RCTs, and provides teachers with research tools and training in research skills, which offer the potential for improving student outcomes and uncovering new interventions.

More broadly, infrastructure should encourage more open science. This means more sharing of data as well as more sharing of tools. This could take the form of a crowdsourced platform where data could be stored, processed, and analyzed, similar to emerging platforms in other fields focused on nurturing innovation One idea would be the creation of a "data-processing" platform for learning engineering where data can be stored, processed, and analyzed. Similar platforms are emerging in other fields (i.e., Galaxy used for biomedical research; Blankenberg et al., 2014) to support increased open innovation.

This infrastructure can be both bottom up and top down. For instance, many researchers will develop their own ideas of what to test to further their own work, relying on their intuition for important research questions. But at the same time, the field should use the infrastructure to explicitly test some of the ideas laid out in survey papers. Similarly, there are longstanding debates over issues such as the timeliness of feedback that are ripe for further testing within such infrastructures.

Beyond this, support for embedding a greater number of measures -- measures of student affect, self-regulated learning, and engagement, among other constructs -- into these platforms would help us understand the full impact of an intervention. For example, if an A/B test only considers student correctness or the time taken to master a skill, it may miss important impacts on student engagement (cf. Lomas et al., 2013).

As several attendees of the virtual convenings noted, this type of infrastructure presents challenges. The first challenge is funding, and this paper recommends additional funding to create "research as a service" opportunities, incentivizing companies to leverage their data for research. Practically speaking, this means two things. Funding organizations should support the development of research infrastructure. Instead of funding individual researchers who must create their own tools for experimentation, supporting the creation of open access tools will enable a broader range of educators and researchers to perform experiments on large populations, accelerating the rate of discovery.

When it comes to funding, there should also be greater support for private companies opening up their data. Currently a number of firms argue that they can't offer research as a service because there's not yet a market. Additionally, data is often considered proprietary and therefore not made available to researchers. But private industries could be incentivized to open their data and systems to researchers. Increasing data openness would not only add to the science of learning's body of knowledge, but also benefit the platform's continued improvement. The success of such an effort will also depend on data being high enough quality to support learning engineering. Encouraging better data standards will support this effort, especially if it involves identifying and scaling best practices in data capture (i.e. what types of variables have been useful in past research efforts), rather than adopting "lowest common denominator" data schemas that work in all platforms but discard the most useful data.

Data standardization efforts, such as xAPI, IMS Caliper, and the Pittsburgh Science of Learning Center DataShop formats, may remove some of the barriers to learning engineering. However, there have been significant challenges using data collected in the practice of educational research and development, even when using these standards, primarily caused by the lack of consideration of the ultimate uses of the data collected. Support to extend these standards for use in a broader range of learning sciences and learning engineering applications could serve to improve the quality of data and reduce the data engineering efforts currently needed. For instance, these standards could be extended with fuller representation of the context of learner behavior and more in-depth representation of inferences made by the system about learners.

Beyond funding and technical capacity development, there are also key challenges around ethics, legal compliance, and student privacy. Currently, for many developers, the easiest way to comprehensively protect learner privacy -- or at least themselves -- is to avoid sharing any data at all, or to share extremely limited data, such as global summaries or highly redacted data sets. Often, measures taken to avoid holding personally identifying information inhibit the feasibility of longitudinal follow-up or avoiding algorithmic bias. If data limitations -- based on very reasonable privacy concerns -- make it so that learning scientists and learning engineers can only ask certain questions, then largely those are the questions that they will ask.

Learning engineering can be part of the solution to this problem, providing frameworks and research around best practices for data security and privacy within educational technologies: methods for automatically redacting forum post data (Bosch et al., 2020), obfuscation and blurring methods that retain but reduce the specificity of demographic information or introduce a small amount of error into data to prevent confident reidentification (Bakken et al., 2004), platforms that allow trusted organizations to hold personally identifying information for longitudinal follow-up, and platforms that allow trusted organizations to hold personally identifying information for longitudinal follow-up, and platforms that allow analysis using full data but do not allow querying or output in terms of specific data values (Gardner et al., 2018). Support for these technologies will enable a broader range of research, helping to achieve many of the other recommendations in this report.

In terms of compliance, the ASSISTments platform has done important work to streamline Institutional Review Board (IRB) processes at Worcester Polytechnic Institute (WPI) and create standard procedures for the WPI IRB to work with other IRBs. Resources should be created so that this work can be replicated across platforms -- having to obtain approval for each study from a different IRB with different "house rules" is a major delaying and complicating factor for learning engineering research. Compliance issues become even more challenging when dealing with cross-border research (i.e. different expectations for human subjects protections and privacy between the USA and the EU). Ideally, processes should be designed so that there is both standardization (limited bureaucratic effort to bring in new research or school partners) and flexibility (ability to easily accommodate different rules and practices, ideally with a limited number of button-clicks within the platform).

Better Learning Engineering: Bring Learning Engineering to Domain-Based Education Research

Adaptive learning systems -- and learning systems and curricula in general -- depend on a high-quality model of the content and learning needs in a given domain. Ideally, such a model represents both the structure of the domain -- which skills are prerequisite to other skills -- and the concepts that a disproportionate number of students struggle with (including misconceptions and preconceptions about them).

However, as attendees of the asynchronous virtual convenings noted, the current state of domain modeling in learning engineering is highly uneven between domains. There has been considerable work on modeling mathematical domains, with some areas of mathematics having full knowledge graphs developed separately by multiple companies or curricular providers. In this case, there are probably opportunities for the field by creating incentives for these knowledge graphs to be shared more broadly -- there is considerable inefficiency in having multiple organizations (possibly even funded by the same funders) spend hundreds of thousands of dollars to create highly similar knowledge graphs.

By contrast, there has been considerably less focus on domain modeling in other domains. In science, language learning, and reading, models of student learning progressions focus on the development of broader concepts (e.g. Schwarz et al., 2009; Berland & McNeill, 2010; Bailey & Heritage, 2014; Van Rijn et al., 2014) and there are Bayesian Network models that compose meta-skills but do not model prerequisites formally (Martin & VanLehn, 1995). In other domains, there has been even less work on domain modeling (indeed, existing ways of representing domain structure in adaptive learning systems may not even be appropriate in some domains, such as history or social studies).

However, recently approaches have been proposed that may be able to capture prerequisites for knowledge graphs more generally, from data sources such as university course catalogs and student interaction with MOOCs (as well as the content of MOOC videos themselves) (Liang et al., 2017; Pan et al., 2017; Chen et al., 2018). Providing funding for these approaches (particularly if developers keep humans in the loop to identify spurious correlations) has the potential to speed the development and deployment of knowledge graphs to a broader range of domains.

To speed efforts in this area, we recommend the creation of a network of R&D teams -- both in industry and academia -- who receive funding for their work to create knowledge graphs, under the agreement that they will open-source and share the knowledge graphs and findings they produce. This network could include machine learning researchers working on automated approaches to distill knowledge graphs, and a central hub whose work is to integrate across all of the work being conducted into single, crowdsourced, shared knowledge graphs. These shared knowledge graphs would represent commonalities (including translation across different ontologies) as well as the different ways knowledge can be represented based on how skills and concepts are taught.

In addition, the infrastructure improvements discussed above in section “Enhance R&D Infrastructure in Widely Deployed Tools” can be used to support discoveries about student misconceptions that impact their learning (Elmadani et al., 2012), and pedagogical strategies for helping students learn these difficult skills and concepts (cf. Lomas et al., 2016).

Infrastructural improvements of this nature can also provide fine-grained data on what approaches work for helping students learn very specific content. To cite one example, a group of researchers led by Erin Ottmar used research infrastructure within ASSISTments to test the spacing of multiplication signs. The team found using an RCT that students learn more if spacing is consistent like “22 − 4 + 3 × 5 = __” rather than “22 − 4 + 3×5 = __” (Harrison et al., 2020). The work indicates that even slight variations in presentation and the design of content can make a difference for student learning.

Better Learning Engineering: Build Components to Create Next-Generation Learning Technologies Faster

Another key area of enabling research and development for learning engineering is in creating software components that make it possible to build next-generation learning technologies more quickly. Currently, developing a new learning platform with advanced adaptivity takes years of effort, limiting entrants to the field and leading to a proliferation of lower-quality platforms. There are a range of potential tools and components that could be engineered for general-purpose use.

At the moment, while several adaptive learning systems support various forms of adaptivity, their designs are largely one-off -- the same functionality has been repeatedly created several times for different learning systems. Occasionally, the same infrastructure will be used within a company or a team for multiple curricula (i.e. Cognitive Tutors for Algebra, Geometry, Middle School Mathematics; Algebra Nation content for Algebra and Geometry; ALEKS content for Mathematics and Chemistry), but different teams build adaptive functionality from scratch.

Ideally, an effort to create general components for adaptivity would be modular in nature, so a new developer could integrate specific components into their own architecture rather than needing to adopt an entire architecture as-is. Similarly, if there is an intervention with benefits thought to be at least somewhat general across learning domains and populations (such as a values affirmation intervention -- Borman, 2017), it could be placed into a component so it can be reused across systems and does not need to be re-implemented.

Perhaps the largest opportunity for creating reusable components is in the area of student modeling. There has been a great deal of research in the last two decades into how to model a range of aspects of the student. However, little of this research has found its way into actual use in scaled learning systems. Currently, not even Bayesian Knowledge Tracing (BKT; Corbett & Anderson, 1995) -- the most widely-used adaptive learning algorithm in the United States -- has good "plug and play" infrastructure for building models, continually re-fitting them, and deploying the algorithm into a running system for its most common use, mastery learning.

Better toolkits -- implementation-quality instead of research-quality -- are needed. Some algorithms, like BKT and ELO (Klinkenberg et al., 2011), are essentially ready for scaled use, and should be the first priority for development. Beyond that, there are more recently developed algorithms that offer benefits such as better fit to data (Khajah et al., 2016; Zhang et al., 2017), the ability to represent partial overlap between skills (Pavlik et al., 2009), and consideration of memory and forgetting (Mozer & Lindsey, 2016; Settles & Meeder, 2016). However, work is still ongoing to understand and address these algorithms' limitations for real-world use (Yeung & Yeung, 2018). Thus, developing implementation-quality toolkits will also involve research into practical questions such as how much data is needed for these algorithms to function effectively, work already conducted for older algorithms such as BKT (e.g. Slater & Baker, 2018).

Going further, concerns were raised by multiple participants in the asynchronous virtual convenings that current student knowledge modeling typically captures relatively straightforward knowledge -- such as specific algebraic skills or factual knowledge -- rather than deeper conceptual understanding or complex generalizable skills. While a small number of projects have attempted to model and infer conceptual understanding (i.e. Kim et al., 2016; Rowe et al., 2017; Almeda et al., 2019) or inquiry skill (i.e. Gobert et al., 2013), these efforts largely represent "one-off" research projects and there is a lack of production-grade toolkits that can be quickly leveraged by practitioners.

One additional area where reusable software components would speed progress is in natural language processing. Specifically, much more should be done to develop toolkits, and funders should support the creation of toolkits that support a greater variety of activities. Natural language processing can be used for a wide variety of educational applications, from automated feedback on essays (Roscoe et al., 2013), to identifying evidence of 21st century skills within on-line discussion forums (Weinberger & Fischer, 2006; Rosé et al., 2008), to the creation of automated conversational agents (Ventura et al., 2018). Systems exist today which use each of these types of functionality, but the engineering process is highly intensive. Text classification still relies mostly on general-purpose tools rather than tools tailored to educational domains; tools exist for the creation of automated conversational agents, but either require extensive expertise (e.g. Cai et al., 2015) or offer only a limited range of functionality

(Wolfe et al., 2013).

Experts within the AVCs identified three broad areas of natural language processing software components that would be particularly valuable to learning engineering. First, there needs to be development work to make it easier to integrate existing algorithms smoothly and seamlessly into educational technologies via APIs, including algorithms for measuring readability, text cohesion and sophistication, and algorithms for sentiment analysis. Second, tools for integrating speech to text software into educational technologies would make it possible to automatically translate students' spoken responses into text that NLP tools can process. Third, there needs to be more work to develop linguistic analysis resources (such as key word lists and topic models) for specific target domains, such as mathematics. These steps would considerably facilitate the integration of natural language processing into a broader range of learning systems.

Support Human Processes: Enhance Human-Computer Systems

One of the biggest opportunities for learning engineering is in enhancing the systems that emerge when humans (teachers, guidance counselors, school leaders) and learning technologies work together to better support students. Practices such as proactive remediation -- where a teacher obtains information from a learning platform on a specific student's progress and reaches out to them to offer assistance (Miller et al., 2015) -- and re-design of classroom activities based on the previous night's homework data -- create opportunities to combine what computers are good at (rapid measurement and simple inference at scale) and what humans are good at (understanding why a problem is occurring and adjusting teaching accordingly) (Baker, 2016).

Ultimately the learning engineering goal and challenge is to develop ways for human and computer tutors to work in concert. There are several human-computer systems that provide opportunities for enhancement, through learning engineering, and several potential avenues for enhancement.

Perhaps the greatest immediate opportunity is in enhancing the data provided to classroom teachers, through dashboards. Dashboards have become a key part of learning technologies (Bodily & Verbert, 2017), and they have facilitated the development of new pedagogical strategies such as proactive remediation (Miller et al., 2015). Different dashboards provide a range of different types of information, from dropout prediction/SIS dashboards with data on attendance and assessments (Singh, 2018) to fine-grained learning technology dashboards that provide data on in-the-moment student performance (Feng & Heffernan, 2006). In some cases, these two types of dashboard are being integrated -- for example MATHia LiveLab predicts if a student will fail to reach mastery and provides these predictions to teachers (Fancsali et al., 2020).

Thus far, most dashboards do not provide more in-depth data on student cognition or affect, although exceptions exist, such as Inq-ITS's dashboard presenting data on student inquiry skill and supporting teachers in providing scaffolds to struggling students (Adair & Dickler, 2020). Here, the challenge and opportunity-- a human-computer interaction challenge as much as a learning engineering challenge -- is to increase the richness of data given to teachers, while maintaining usability and comprehensibility.

Ethnographic research has suggested that teachers do not just want raw data -- they want real-time recommendations about when to provide additional support to students, and what kind of support to provide (Holstein et al., 2019). In tandem, it will be essential to design dashboards -- or other ways of informing teachers (i.e. Alavi & Dillenbourg, 2012; Holstein et al., 2018) -- that support and encourage effective pedagogies for classroom data use. In other words, data dashboards are not just about communicating information, they are about supporting/changing practices, and will succeed to the extent that they support teachers' (or students', or school leaders') goals. In discussions, John Whitmer noted that one of the key goals for future research will be to identify whether providing a dashboard has an impact on improving student outcomes, an under-studied area (but see Xhakaj, Aleven, & McLaren, 2017).

An additional avenue for enhancing human computer-systems is through enhancing the integration of computer tutoring experiences and human tutoring experiences. Many online learning platforms today offer access to human tutors as a complement to their otherwise digital offerings, from home credit recovery platforms such as Edgenuity (Eddy, 2013) to blended learning systems such as Carnegie Learning (Fancsali et al., 2018) and Reasoning Mind (Khachatryan, 2014). An industry of companies, such as Tutor.com, has grown to offer these services. However, there has been relatively limited formal research on this practice and how to use learning engineering to enhance it. In one of the few examples of research, Carnegie Learning analyzed the factors that predicted that a student would seek human tutoring through a linked platform (Fancsali et al., 2018). Considerably more research is needed to support learning engineering efforts in this area. Specific questions to address include:

Can we understand what leads to tutors being more or less effective in this blended context?

Is the earlier research on what makes human tutors effective relevant in this context?

When should students seek help from the computer, and when should they seek help from a human tutor?

At the moment, human tutoring embedded into computer tutors is on-demand, depends on student meta-cognition, is used very unevenly (Fancsali et al., 2018), and requires tutors to get up to speed very quickly based only on information directly provided by the student. Eventually, through learning engineering, we may be able to develop a more sophisticated blend of approaches -- using computers for routine and repetitive parts of instruction, empowering teachers and tutors with more complete information from the computer, and developing technology that chooses to loop in a tutor or teacher when the learner isn't making progress.

Support Human Processes: Better Engineer Learning System Implementation in Schools

Many learning systems and curricula work well under favorable conditions -- motivated teachers and supportive administration, with significant involvement from the developers in support for teacher professional development, as well as ongoing support during the use of the system. However, these same learning systems and curricula often fail when extended to a broader range of classrooms, where teachers may be unfamiliar or uncomfortable with new methods and technologies, and may attempt to assimilate new technologies back into traditional teaching practices. These challenges in implementation are ultimately challenges for learning engineering. Can we design learning technologies that are easier for teachers to incorporate into their practice, while maintaining the benefits and advantages of these technologies?

The example of mastery learning is an example of a success in learning engineering implementation. Mastery learning, the practice of advancing students between topics only when they demonstrate they know their current topic, proved both effective and difficult to scale in traditional classrooms (Guskey & Gates, 1986; Kulik et al., 1990). Adaptive learning systems such as Cognitive Tutor/MATHia made mastery learning a key part of their approach, and successfully scaled in American classrooms (Koedinger & Corbett, 2006). However, even in these cases, some teachers work around the mastery learning aspects of the system's design, overriding the system to advance students who have not fully learned the current topic (Ritter et al., 2016). These decisions lead students to struggle more and make more errors (Ritter et al., 2016). Understanding why teachers make these decisions will be key to developing learning systems -- and strategies that implement them -- that work in the real world, at scale. We cannot change teacher decisions, or improve our systems to better meet teacher goals, without understanding those decisions.

In these cases, both the design of learning systems and the design of professional support and training become areas where learning engineering is key. As Harvard's Chris Dede noted in the virtual asynchronous convening, "We need to prepare a range of professionals (such as instructional designers, teachers, purchasing agents, managers, regulators) to work in new ways. This will be its own at-scale learning project: we need double or even triple-vision as we develop interventions, to recognize not only that the learner's mind has to be handled in new ways – but also that all the other professionals involved need training, with sufficient practice and feedback, to change their practices. This is another kind of "learning engineering" opportunity that is still in its infancy, yet will be crucial for at-scale success."

This sparks another recommendation. Many teachers adopt practices that are less effective than the practices designers intend -- but some teachers may adopt practices around learning technologies that work better than what the designers intended (e.g. Schofield, 1995). Improved integration of data between classroom practices and students' learning experiences can be used to study which practices around the use of learning technology are effective and scalable, and what contexts/situations these practices work best in. The data can then be used to analyze and detect whether best practices are being used, and develop automated and semi-automated methods to encourage teachers to use the right practice at the right time. For example, a pop-up message in a dashboard might encourage a teacher to speak with a student who has been struggling for the last thirty minutes. Moreso, by also integrating data on teachers' professional development experiences, learning engineers can study which professional development experiences lead to changes in practice which experiences lead to changes in practice which lead in turn to better outcomes for learners.

Hence, creating data systems where data on teachers' professional development, their classroom practices, and students' learning experiences are connected will act as a key enabling factor for research on improving implementation. Once this infrastructure is in place, support for work to develop a taxonomy of teacher practices in classrooms using learning technology will create a framework that can be used across learning platforms to study which practices benefit learners, and how to engineer systems and professional development that produce those practices. These efforts should acknowledge that there may not always be a single set of practices that are optimal in all situations -- the effectiveness of a given strategy may be impacted by local conditions and the attributes of both students and teachers. What's more, a data-infused approach could help determine if some teachers do a better job than their peers in helping a student master a given concept, a longstanding question in education. New-generation approaches will leverage far more information and thus go beyond narrow questions around very aggregate value-add (cf. Rubin et al., 2004) and instead identify specific instructional approaches that help students learn specific material. It may even become possible to identify how the best teachers customize and adapt learning environments for their classrooms, feeding back into the design of both professional development and adaptive support.

Improve Recommendation, Assignment, and Advising Systems

Over the last decade, applications that incorporate models that can predict if a student will fail a course or drop out have become a common part of K-12 and higher education (Bowers et al., 2012; Milliron, et al., 2014). Today, these models are used to provide reports to school leaders and advisors (Milliron et al., 2014; Singh, 2018), or to drive simple automated interventions (Whitehill et al., 2015), and have been successful at improving student outcomes in a variety of contexts (Arnold & Pistilli, 2012; Milliron, et al., 2014; Whitehill et al., 2015).

However, predictive models are generally not yet built into advising or recommender systems used to support students in selecting courses. For example, course selection and registration processes at many institutions involve limited advising, leaving it to students to identify and select courses with minimal support. This leads to many students taking "excess" credits in college or community college that don't count towards degree requirements and use up financial aid (Zeidenberg, 2015).

There is an opportunity to use learning engineering to develop advising systems that proactively analyze student trajectories and make recommendations (to advisors or to the students themselves) that increase the likelihood that the student achieves their goals, during high school (graduation and enrollment into college), college (completion of degree or transfer to 4-year college), and in the workforce (success at obtaining job and at job performance). There are systems in wide use that make course recommendations to students (Bramucci & Gaston, 2012), and there are already models that can make this type of longitudinal prediction (see, for instance, San Pedro et al., 2013; Makhlouf & Mine, 2020), and there are models for how to deliver recommendations of this nature (e.g. Castleman & Meyer, 2020), but only a few examples of this integration (e.g. Jiang, Pardos, & Wei, 2019).

The key challenge is to take these models developed with one purpose -- prediction -- and re-purpose them for a different use, recommendation. Then, learning engineering is needed to make the recommendation and proposed intervention maximally effective at achieving its goals -- going beyond just improving the algorithms to re-engineering the practices of counselors and advisors, shaping their practices with the technology. A key part of this will be conducting iterative design, building on relevant research literatures such as the extensive work on nudge interventions (Hansen & Jespersen, 2013; Damgaard & Nielsen, 2018), to develop recommendations that students and instructors follow, and that achieve their desired goals of improved outcomes.

One concern raised in the AVC sessions regarding this type of technology is that some of the algorithms currently being used for advising and recommendation do not provide details on why recommendations are made, making it difficult for practitioners to understand and trust the recommendations, and raising concerns of unknown algorithmic bias. There are also concerns that current approaches offer recommendations for students "on the bubble" of success and failure while leaving students at very high-risk unsupported.

Recommendation and advising systems can be advanced through increased support for research on re-purposing prediction models in this space for use in recommendation. There are two key steps to this. First, research on how to distill human-interpretable and actionable recommendations out of complex prediction models. Adapting explainable AI methods to the problem of actionability -- so that models are not just explainable, but explainable in ways that enable action -- will require research projects that bring together machine learning researchers, human-computer interaction researchers, and educational researchers.

Second, modern recommender systems in other domains can improve their own performance by studying whether their recommendations are followed, and what the results are; this can be achieved for this problem by building a laboratory at a specific institution such as a community college (or several such institutions) that brings together an infrastructure that enables experimentation around prediction and recommendation and connects it with outcome data. In taking these steps, it will be essential to support and fund solutions that are inspectable, understandable, trustworthy, and beneficial to the full range of learners.

A related problem is matching students to schools in large school districts or local educational agencies. Many school districts still use cumbersome, complicated multi-stage enrollment processes where many schools leave places unfilled and many students end up in schools that they do not prefer and end up leaving. Work over the last decade, in several cities, has shown that even relatively simple matching algorithms can lead to much better matching outcomes (Pathak, 2017). Extending this work with the use of sophisticated AI-driven recommender systems has the potential to guide students to make better choices about which schools they list as their preferences (both more realistic and more likely to lead to personal and career success), as well as guiding schools to choose students more optimally.

Better Learning Technologies: Optimize for Robust Learning and Long-Term Achievement

It is important to design learning experiences that support students in developing robust learning -- learning that is retained over time, transfers to new situations, and prepares students to learn in the future (Koedinger, Corbett, & Perfetti, 2012). Learning design has often emphasized short-term learning outcomes, as they can be easier to measure. As Caitlin Mills noted in an asynchronous virtual convening,

“The things that may be most easy to immediately measure (number of problems done or such) may not be the things we wish to optimize.”

There is a risk that this problem will be amplified by learning engineering, despite its significant benefits overall. The common practice of improving a product using rapid innovation cycles risks focusing on measures that can be applied at that rapid time-scale. Similarly, the practice of using learning system data to assess the effects of an innovation risks focusing energy on improvements that are easy to measure. It is easy to measure immediate performance improvement on an exact well-defined skill -- there are now hundreds of examples of this type of work. It is significantly harder to quickly measure transfer across skills or preparation for future learning -- although many examples still exist (Koedinger, Corbett, & Perfetti, 2012; Kodinger, Booth, & Klahr, 2013). Taking this step is nonetheless essential to guarantee that learning engineering produces learning that is active and useful to learners.

As Caitlin Mills noted in our AVC sessions, the field has poor understanding of which interventions' effects persist over time -- and the dynamics of different interventions across multiple time-scales (i.e. dosage, appropriate repetition, half-life). Measuring long-term impacts requires researchers to plan ahead and maintain continuity of follow-up on students. Relatively few researchers even look at retention of knowledge over the span of a few weeks or months. Over a period of several years, a span of time where students move to new schools, learning systems may change their designs in significant ways, and research team composition is likely to change, it becomes even harder. Unlike research on much coarser-grained interventions, such as charter schools (Sass et al., 2016), we are only aware of one example where students who used an adaptive learning system were followed up over the span of several years (San Pedro et al., 2013, 2015; Almeda & Baker, 2020).

Given that many pedagogies and learning strategies work in the short term and for the exact material studied, but can lead to poor recall and transfer (Donovan & Radosevich, 1999; Ben-Zeev & Star, 2001; Rawson et al., 2013), this limitation in current practice carries risks of worsening outcomes for students, rather than improving them. We need to try to understand what the longer-term impacts of learning engineering decisions made today are. This concern can be addressed in multiple ways.

First, the field should be made more aware of designs and approaches that are already known to lead to worse outcomes in the long-term, such as cramming (Rawson et al., 2013), massed practice (Donovan & Radosevich, 1999), and the lack of interleaving of related skills (Ben-Zeev & Star, 2001). Many learning systems, particularly in mathematics, still use massed practice of skills taught in a block. Extensive research in cognitive psychology suggests that this practice may be less effective for learning.

Second, the field needs to go beyond conducting short-term studies on interventions and designs. Short-term A/B tests are convenient to run, but may favor approaches whose benefits do not sustain. As such, we recommend additional work in this space. Specifically, explicit plans should be put in place to follow up promising studies, systems, and pedagogies, to see if the apparent benefits sustain over a longer term. This should be applied to a range of types of intervention, from learning interventions to persistence/retention interventions. A range of possible benefits may be possible, from greater educational attainment, to career and even health benefits.

This type of follow-up research is easier if it is planned for in advance, by saving key follow-up information (in ways that respect students' choices of whether to be followed longitudinally) and deciding in advance how and when follow-up will occur. This trend is already occurring and simply needs further encouragement, perhaps by setting aside funding for follow-up at the start of a project. Elizabeth Albro at the US Department of Education, Institute of Education Sciences, noted in an AVC conversation that "we have supported long-term follow ups for several IES-funded projects. We also encourage PIs to plan for long-term follow up at the outset of their projects – including attending to language in their consent letters to ensure that long-term data collection is an option." She further suggested that researchers should "use state longitudinal data systems to measure long-term impact", an opportunity to facilitate the sometimes challenging task of tracking longitudinal impacts.

Similarly, there should be greater emphasis on preparation for future learning. What areas of mastery, or ways of learning a topic, improve a student's ability to learn the next topic faster or more effectively (e.g. Bransford & Schwartz, 1999; Chi & VanLehn, 2007)? Much of the work seen in this area so far is specific to a single learning domain, such as a specific mathematical skill being key to the development of other, later skills (Booth & Newton, 2012). For example, work by NWEA and EDC involving a data set of over 200,000 students found that student performance in four 6th-grade mathematical domains were each independently predictive of achievement in 8th grade algebra (Almeda et al., 2020). For example, a one standard deviation increase in Real and Complex Number Systems was related to a third of a standard deviation improvement in math overall, two years later. This applies beyond just mathematics. Emerging work suggests, for instance, that scientific inquiry skills learned in physics can be applied in biology as well (Sao Pedro et al., 2014). By finding which learning activities are most essential for future progress, we can focus instructional and design effort where it has the highest potential impact. More should be done to look at educational impact and accelerated future learning, further downstream from the intervention, even months or years later.

Better Learning Technologies: Support Learning 21st-Century Skills and Collaboration

Much of the learning technology currently in use focuses on relatively narrow academic skills, but more complex skills such as collaboration, communication, and critical thinking -- often referred to as "21st-century skills" -- will be key to career and life success in the coming decades (Dede, 2010). These skills are often hard to measure as they do not have a purely right or wrong answer to easily classify. Using new technologies, new data collection tools, analytics and psychometrics, learning engineering can focus on the development of reliable and valid measures of these hard to measure constructs, and produce learning experiences that support their development.

For example, game-based assessments and simulations appear to have promise for measuring a range of 21st-century skills, from inquiry skills (Gobert et al., 2013; Sparks & Deane, 2015), to cognitive flexibility and conscientiousness (Shute et al., 2015), to collaborative problem-solving (Chopade et al., 2018; San Pedro et al., 2019). Intelligent tutors have also proven to be useful environments for studying self-regulated skills such as help-seeking and strategies for improving these skills (Aleven et al., 2006; Aleven et al., 2016). One of the largest challenges to developing these types of measurements is obtaining reliable and agreed-upon human judgments of 21st-century skills, that can be used to leverage machine learning or to inform evidence-centered design approaches to developing these measures. One path to collecting this data may be to improve tools for visualizing and annotating student log data (Rodrigo et al., 2012; Gobert et al., 2013; Rowe et al., 2019), to support discussion and refinement of coding schemes, comparison between human coders and analysis.of their differences, and data-driven discussion around measurement design.

Top High-Leverage Opportunities

An area of particular importance is 21st-century skills around collaborative learning and collaborative performance. Collaborative work is an integral part of our society both at the academic level and in the workforce. Learning engineering is uniquely positioned to help practitioners, employers, and students better understand collaboration through advanced technology and methodologies. While there has been initial work on collaboration and patterns of collaboration (Lahti et al., 2004), and development of frameworks for using evidence-centered design to assess collaboration (Nouri et al., 2017; Andrews-Todd & Kerr, 2019), this work is still in its beginnings.

Collaboration is an important strategy for learning but current learning tools and systems for collaboration are less advanced than tools and systems for individual learning. Learning engineering can begin to shed more light on best practices for evaluating collaborative work, teams, communication and other skills directly related to 21st century skills. This challenge becomes more tractable as learning shifts increasingly online -- collaboration taking place completely in person is difficult to measure without complex multimodal approaches (Laru & Järvelä, 2008; Noel et al., 2018) or sophisticated equipment (Martinez-Maldonado, et al., 2013) that are difficult to deploy in real classrooms. By contrast, collaboration taking place fully online, whether synchronous (Diziol et al., 2010) or asynchronous (Calvo et al., 2010) can be considerably easier to measure. Discussion forum data, for instance, is quite easy to analyze, leading to research that integrates across grain-sizes, from textual cohesion to social networks (Joksimovic et al., 2015). Even ZOOM recordings, while not collected with data analysis in mind, provide direct images of participants' faces and a view of the document being shared, which will be easier to work with than cameras deployed in classrooms where students are moving around as they work together.

Learning engineering can help develop tools that scaffold collaboration, and better assess collaborative skill, to help learners learn to collaborate, and learn more effectively while collaborating. As an example, researchers could further examine the viability of sociometers (Choudhury & Pentland, 2002), wearable devices that measure social interactions. Such tools could better measure 21st-century skills like collaboration and social engagement by examining which students engage with others and how it shapes performance and learning (Martinez-Maldonado et al., 2012; Evans et al., 2016). In general, data will enable us to study which learner behaviors and strategies lead to effective collaboration and learning outcomes through collaboration -- given the unique nature of collaborative learning, social relationships and the behaviors that support their development may play a key role (i.e. Kreijns, 2004; Gasevic et al., 2013).

Learning engineering has a role to play in creating learning experiences that can measure and support the development of 21st-century skills. This work will involve the creation of better measures and better methods for developing measures. Evidence-centered design and educational data mining have each been successful at measuring specific 21st-century skills (e.g. Shute & Torres, 2012; Gobert et al., 2013; Kantar et al., 2018); however, there is still insufficient work to understand when each method is best and how to use these methods together (but see Mislevy et al., 2012; Rupp et al., 2012).

Support for formalizing methods for measuring 21st-century skills, including collaboration, may expand the use of these methods, particularly if it can articulate and systematize how evidence-centered design and educational data mining should be used together. In addition, work to enhance students' 21st-century skills, including collaboration, has not sufficiently looked into the long-term retention of what is learned, and the translation of those skills to new contexts (a more general problem; see previous recommendation).

These research areas are currently moving forward, and these goals are on track to be achieved -- but not quickly. Hence, the major challenge here is to speed research on key goals such as developing better measures of 21st century competencies and better methods for developing them. More should be done in this area, from additional research funding to greater focus among researchers.

Although considerable effort goes into this problem today, different research and development teams are working on different aspects of this problem. As such, there is limited scope for the type of competition that often accelerate progress. Attempts to bring together large numbers of researchers to discuss these problems has led to committee solutions that do not seem to kick-start the field (e.g. Graesser et al., 2018; Krumm et al., 2016). Instead, the field may benefit from explicit competition, such as seen in the ongoing competition to develop better measures of student knowledge or in other domains such as natural language processing and image processing.

Existing competitions in educational data have been too brief in duration for this type of challenge. We recommend instead establishing challenges like the Loebner Prize that attach funding to demonstrating specific types of functionality in measurement or skill development. For instance, a prize could be given to the first team to produce an automated measurement of collaboration associated with better workplace outcomes in cross-culture workplaces, or the first team to produce an intervention based upon automated measurement of conscientiousness that led to higher conscientiousness in real-world tasks.

Better Learning Technologies: Improved Support for Student Engagement

There is growing acknowledgment that there is more to learning than just what is learned. Student engagement can make a big difference, both to immediate learning (Craig et al., 2004; Cocea et al., 2009), and to longer-term interest and participation in a subject (San Pedro et al., 2015; Almeda & Baker, 2020). Developing technologies that take student engagement and affect into account has therefore become an important goal for many in the learning engineering field. Engagement and affect have been measured from both sensors and from logs of student interactions with learning systems (Calvo & D'Mello, 2010; Baker & Rossi, 2013; Baker & Ocumpaugh, 2014). However, though the technology exists to measure engagement and affect, the technology is not yet in place to reliably use these measurements to improve engagement and affect. Though a small number of approaches have been effective at improving engagement and learning, these technologies have not scaled.

Investments in infrastructure in this area, recommended in our discussions with John Whitmer, may assist in the scaling of this type of technology. Currently, three approaches have been used to collect data on engagement and affect for developing automated measurements: classroom observations, video data, and self-report. The classroom observation path to developing automated measurements has been used in over a dozen systems, has a widely-used Android app for data collection (Baker, Ocumpaugh, & Andres, 2020), and even the financial costs have been systematically studied (Hollands & Bakir, 2015). However, it is not feasible in remote learning contexts. Developing a standard self-report instrument for engagement and affect -- realized as a software plug-in -- and validating it across learner populations will increase the feasibility of collecting large-scale remote data which can be used to develop detectors that recognize student engagement and affect from interaction data. For video, it may be possible to develop a single suite of engagement/affect detectors validated to work across populations, much as has been done for basic emotions by several commercial vendors. The largest challenge to doing this will be the collection of a large-scale and diverse corpus of video learning data, annotated in terms of key individual differences such as age, gender, race/ethnicity, and type of camera/webcam.

Moving forward, learning engineering has the opportunity to examine which engagement/affective interventions (both teacher-driven and automated) are effective for which students, in which situations. A range of possible types of interventions have been developed -- from conversational agents (D'Mello et al., 2009), to visualizations of student engagement (Arroyo et al., 2007; Xia et al., 2020), to messages in between learning activities (DeFalco et al., 2018).

However, relatively little work has studied how student individual differences impact the effectiveness of these interventions (but see D'Mello et al., 2009; Arroyo et al., 2013), and insufficient work has compared different intervention types to each other or attempted to integrate multiple interventions. Learning engineering can help to answer these questions. This limitation in the research so far can be addressed through creating a competition where different interventions are compared and integrated, in a large sample of students where individual difference measures are also collected.

In parallel, work is needed to figure out how to design these interventions in ways that teachers, school leaders, parents, and students are comfortable with. Many interventions that are successful at reducing disengagement or improving affect are not acceptable to students or teachers (e.g. D'Mello et al., 2009), reducing their potential to scale. Greater support for work to understand stakeholder needs and desires, and to design in accordance with these needs (i.e. Holstein et al., 2019) increase the potential for uptake and scaling.

The role of parents is particularly important. As a long-standing body of research shows, parents, home-life and other factors external to the classroom have a considerable impact and can exacerbate achievement gaps (Hara & Burke, 1998). A number of programs show that engaging parents can make substantial impacts (Berkowitz et al., 2015; Mayer, Kalil, Oreopoulos, & Gallegos, 2015). One of the challenges in empowering parents to support their children is the numerous gaps that currently exist in the communication between school administrations, students, and parents. Simple interventions like providing parents login information for school learning management systems can lead to improvements in student achievement (Bergman, 2020), as can providing parents automated text nudges (Bergman et al., 2019). Hence, it is becoming clear that opportunities can be created if learning engineering focuses on parents as a lever.

Better Learning Technologies: Design Algorithms and Learning Systems for Diversity and Equity

We conclude this set of topics with the most important recommendation for the ultimate success of learning engineering as a field, one that interacts in key ways with all of the other recommendations that precede it. Promoting equity in education and closing achievement gaps is a long-standing goal for educators and researchers, but has remained elusive (Hanushek et al., 2019). One approach to equitable instruction is making learning more individualized. As several convening participants noted, it is essential to get beyond one-size-fits-all interventions and create interventions that are sensitive to differences between learners and promote equity. Learning engineering is well-suited to help educators identify these needs and provide for them, in theory producing a more equitable learning experience through technology-enhanced innovation (Aguilar, 2018).

However, it is not a given that learning engineering will steer instruction and assessment towards equity. Researchers and developers must be mindful to avoid algorithmic biases in analytics and recommendations (Gardner et al., 2019; Holstein & Doroudi, 2019), which can lead to models and interventions being less effective for specific (ofter historically-underserved) groups of learners. Research has suggested that models fit on convenience samples can be less effective for specific groups of learners (Ocumpaugh et al., 2014). Building models that are verified to function correctly for all of the groups of learners using it remains a challenge for the field, although tools such as The Generalizer (Tipton, 2014) can help identify schools to sample to achieve a representative population.

However, there is still limited understanding of which differences between learners matter in specific situations, and how these differences impact the effectiveness of learning technologies. As demand increases within school districts for evidence of equity as well as evidence of broader effectiveness (Rauf, 2020), it will become essential for learning engineers to fill this gap. This limitation can be addressed fairly quickly if research funders require that projects work with diverse and representative populations of learners, collect more complete data on learner diversity, and check models and findings for algorithmic bias using these variables. Race, ethnicity, studying in a second-language, gender, neurodiversity, disability status, urbanicity, and military-connected status can all impact algorithm effectiveness (Baker & Hawn, 2021). However, data on these variables is currently seldom even collected (Paquette et al., 2020), a key first step that needs to be taken for the field to move forward on increasing equity.

Similarly, the designs that work in one learning context may not work as well in other contexts, due to differences in school culture, students' prior learning experiences and curricula, and differences in the national culture and background of learners. For example, learning science findings derived from non-diverse populations sometimes do not apply to other /learners (Karumbaiah et al., 2019), and learning science findings obtained in one national culture sometimes do not apply within other cultures (Ogan et al., 2015). Improving the degree to which learning experiences are culturally relevant and build on material that students are already familiar with can also have significant benefits (e.g. Pinkard, 2001; Lipka et al., 2005). To develop designs that function well where they are applied, there is a need for participatory and inclusive design, involving members of the communities impacted (cf. Tuhiwai-Smith, 2013).

A final concern for equity through learning engineering is that not all students have equal experience with or access to technology. Remote learning or opportunities to acquire additional support are limited by a student's access. For example, despite MOOCs being thought of as an educational equalizer, they are not equally available to all learners (Park et al, 2019) and outcomes are often poorer for learners from less privileged socioeconomic backgrounds (Kizilcec et al., 2017).

The recent pandemic has shown that technology access can also be much lower in under-funded school districts than elsewhere (Wolfman-Arent, 2020), and for undergraduates coming from lower-income backgrounds (Jaggars et al., 2021). Learning engineering clearly cannot solve all inequities that lead to differences in access to technologies or how it is used. But the field needs to consider how effective results will be in practice, given these constraints. Realistic estimations of effectiveness will encourage transparency around who will benefit from learning engineering advances, while also shining a spotlight on the inequities that exist for students and the need to address them.

Conclusions & Next Steps

In this report, we have outlined ten key areas of opportunity for research in learning engineering (see executive summary for a summary of these areas), and within those areas propose 33 potential lines of funding (summarized in the recommendations matrix). There are overlaps between many of these potential lines of funding -- for example, better algorithms for equity can and should be pursued in projects focused on other topics, and enhanced R&D infrastructure will support all other research areas in this document. Encouraging researchers to consider multiple of these opportunities in a single project will help to expand coverage and bring the field forward. Over the coming months, we intend to seed thinking around these areas and their possible integrations through a series of thought pieces. However, even with integration between ideas, it is unlikely that any single program can provide support for all of these opportunities. Coordination between funders can ensure that the full set of opportunities presented here are addressed.

As this report demonstrates, learning engineering has the potential to have huge impacts across a variety of areas. A considerable number of successful examples of learning engineering exist. However, scaling and dissemination remain challenges.

In terms of scaling, too many of the most sophisticated technological approaches and pedagogical approaches remain in research classrooms -- either as wholly academic projects, or as demonstrations and pilots by platforms that have broader use. The move towards making learning platforms into platforms for research has led to a proliferation of papers on how to engineer learning better -- but many of those innovations have not scaled, even in the systems being studied. This situation underpins the recommendation around better engineering of implementation, a key step towards scale.

In terms of dissemination, many of the findings of learning engineering could apply in new learning platforms and in non-technological learning situations, but remains applied in a single platform. Even when shared, most learning engineering findings are disseminated in academic journals or at academic conferences. While this is effective at engaging other scientists and sharing ideas as well as promoting collaboration, these mediums are not optimal for putting work into practice at scale.

Teachers, parents, policy makers, and even many learning system developers are unlikely to read (or, often, have access to) academic journals and conferences and thus are often unaware of new results and findings that directly impact their classrooms and learners. Until work is more widely disseminated there will remain a disconnect between the large volumes of high-quality R&D work being generated both in academia and industry and educational practice. Thus, in addition to recommending funding for the opportunities within this document, we also recommend continued efforts to enhance connections between research and practice.

Learning engineering has great promise for enhancing learning experiences, enriching learning, and supporting better long-term achievements by learners. Considerable strides have already been made; we are at the beginning of the field's journey towards transforming education.

About the Authors

Ryan Baker is the Associate Professor in the Graduate School of Education at University of Pennsylvania. He and his colleagues have created detectors in educational software to analyze students' emotional and meta-cognitive behaviors during online classes to study and improve student engagement and retention.

Ulrich Boser is the founder of The Learning Agency and a senior fellow at the Center for American Progress. He is the author of Learn Better, which examines the new science of learning and was listed as Amazon's "best science book of the year."

Acknowledgements

The authors would like to thank Tom Kalil and Kumar Garg, who provided invaluable help and support. Many colleagues gave thoughtful suggestions and feedback including David Porcaro, Bror Saxberg, Gouri Gupta, Ben Motz, John Whitmer, and Pior Mitros.

The authors also wish to acknowledge the support of Erica Snow, Kayla Meyers, and Rina Madhani. The paper was written with support from Schmidt Futures, a philanthropic effort founded by Eric and Wendy Schmidt.

References

Adair, A., & Dickler, R. (2020). Supporting teachers supporting students: Iterative development of TIPS in a teacher dashboard. In M. Gresalfi, & I. S. Horn (Eds.), The interdisciplinarity of the learning sciences, 14th International Conference of the Learning Sciences (ICLS) 2020 (pp. 1769-1770). International Society of the Learning Sciences. https://repository.isls.org/bitstream/1/6434/1/1769-1770.pdf

Agarwal, D., Babel, N., Baker, R. (2018) Contextual derivation of stable BKT parameters for analysing content efficacy. In K. E. Boyer, & M. Yudelson (Eds.), Proceedings of the 11th International Conference on Educational Data Mining (pp. 596-601).

Aguilar, S. J. (2018). Learning analytics: At the nexus of big data, digital innovation, and social justice in education. TechTrends, 62(1), 37-45. https://doi.org/10.1007/s11528-017-0226-9
Alavi, H. S., & Dillenbourg, P. (2012). An ambient awareness tool for supporting supervised collaborative problem solving. IEEE Transactions on Learning Technologies, 5(3), 264-274. https://doi.org/10.1109/TLT.2012.7

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. International Journal of Artificial Intelligence in Education, 16(2), 101-128.

Aleven, V., McLaughlin, E. A., Glenn, R. A., Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In R. E. Mayer & P. Alexander (Eds.), Handbook of research on learning and instruction, (2nd ed., pp. 522–560). Routledge.

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help helps, but only so much: Research on help seeking with intelligent tutoring systems. International Journal of Artificial Intelligence in Education, 26(1), 205-223. https://doi.org/ 10.1007/s40593-015-0089-1

Almeda, M., & Baker, R. S. (2020). Predicting student participation in STEM careers: The role of affect and engagement during middle school. Journal of Educational Data Mining, 12 (2), 33-47. https://doi.org/10.5281/zenodo.4008054

Almeda, M.V., Rowe, E., Asbell-Clarke, J., Scruggs, R., Baker, R., Bardar, E., & Gasca, S. (2019) Modeling implicit computational thinking in Zoombinis Mudball Wall puzzle gameplay. Paper presentation, Technology, Mind, and Society Conference 2019, Washington, D.C., United States. https://www.upenn.edu/learninganalytics/ryanbaker/TMS2019Mudball.pdf

Almeda, M., Silander, M. & Cox, C. (2020). Which focal math skills predict 8th grade Algebra Achievement? Manuscript in preparation.
Andrews-Todd, J., & Kerr, D. (2019). Application of ontologies for assessing collaborative problem solving skills. International Journal of Testing, 19 (2), 172-187. https://doi.org/10.1080/15305058.2019.1573823

References

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 (LAK '12), 267-270. Association for Computing Machinery. https://doi.org/10.1145/2330601.2330666

Arroyo, I., Burleson, W., Tai, M., Muldner, K., & Woolf, B. P. (2013). Gender differences in the use and benefit of advanced learning technologies for mathematics. Journal of Educational Psychology, 105 (4), 957. https://doi.org/10.1037/a0032748

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. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), Artificial intelligence in education: Building technology-rich learning contexts that work (pp. 195-202). IOS Press. http://all.cs.umass.edu/pubs/2007/arroyo_fjdmfbmw_AIED2007.pdf

Bailey, A. L., & Heritage, M. (2014). The role of language learning progressions in improved instruction and assessment of English language learners. TESOL Quarterly, 48 (3), 480-506. https://doi.org/10.1002/tesq.176

Baker, R. S. (2016). Stupid tutoring systems, Intelligent humans. International Journal of Artificial Intelligence in Education, 26(2), 600-614. https://doi.org/10.1007/s40593-016-0105-0

Baker, R. S. J. d, Corbett, A. T., Koedinger, K. R., Evenson, S., Roll, I., Wagner, A. Z., Naim, M., Raspat, J., Baker, D. J., Beck, J. E. (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 (ITS '06) (pp. 392-401). Association for Computing Machinery. https://doi.org/10.1007/11774303_39

Baker, R. S., Gowda, S. M., & Salamin, E. (2018). Modeling the learning that takes place between online assessments. In J. C. Yang (Ed.), Proceedings of the 26th International Conference on Computers in Education, (pp. 21-28). Asian-Pacific Society for Computers in Education. https://www.upenn.edu/learninganalytics/ryanbaker/ICCE2018_Paper-Alef-Baker-v9.pdf

Baker, R. S., & Hawn, A. (2021). Algorithmic Bias in Education. Retrieved 3/1/2021 from https://edarxiv.org/pbmvz/

Baker, R.S.J.d., & Ocumpaugh, J. (2014). Interaction-based affect detection in educational software. In R. Calvo, S. D'Mello, J. Gratch, & A. Kappas (Eds.), The Oxford Handbook of Affective Computing (pp. 233-245). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199942237.013.009

Baker, R.S., Ocumpaugh, J.L., Andres, J.M.A.L. (2020) BROMP Quantitative Field Observations: A Review. In R. Feldman (Ed.) Learning Science: Theory, Research, and Practice, pp. 127-156. New York, NY: McGraw-Hill.

References

Baker, R. S. J. d., & Rossi, L. M. (2013). Assessing the disengaged behavior of learners. In Sottilare, R., Graesser, A., Hu, X., & Holden, H. (Eds.) Design recommendations for intelligent tutoring systems – Volume 1: Learner modeling (pp. 155-166). U.S. Army Research Laboratory. http://radix.www.upenn.edu/learninganalytics/ryanbaker/BakerRossi2013.pdf

Bakken, D. E., Parameswaran, R., Blough, D. M., Franz, A. A., & Palmer, T. J. (2004). Data obfuscation: Anonymity and desensitization of usable data sets. IEEE Security & Privacy, 2 (6), 34-41. https://doi.org/10.1109/MSP.2004.97

Ben-Zeev, T., & Star, J. R. (2001). Spurious correlations in mathematical thinking. Cognition and Instruction, 19(3), 253-275. https://doi.org/10.1207/S1532690XCI1903_1

Bergamin, P., & Hirt, F. S. (2018). Who's in charge?—Dealing with the self-regulation dilemma in digital learning environments. In K. North, R. Maier, & O. Haas (Eds.), Knowledge management in digital change (pp. 227-245). Springer. https://doi.org/10.1007/978-3-319-73546-7_14

Bergman, P., (2020). Nudging Technology Use: Descriptive and Experimental Evidence from School Information Systems. Education Finance and Policy, 15(4), 623-647. https://doi.org/10.1162/edfp_a_00291.

Bergman, P., & Chan., E. (2019). Leveraging Parents through Low-Cost Technology: The Impact of High-Frequency Information on Student Achievement. The Journal of Human Resources. 10.3368/jhr.56.1.1118-9837R1.

Bergner, Y., & von Davier, A. A. (2019). Process Data in NAEP: Past, Present, and Future. Journal of Educational and Behavioral Statistics, 44(6), 706–732. https://doi.org/10.3102/1076998618784700

Berkowitz, T., Schaeffer, M. W., Maloney, E. A., Peterson, L., Gregor, C., Levine, S. C., & Beilock, S. L. (2015). Math at home adds up to achievement in school. Science, 350(6257), 196-198. https://doi.org/10.1126/science.aac7427

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. Science Education, 94(5), 765-793. https://doi.org/10.1002/sce.20402

Blankenberg, D., Von Kuster, G., Bouvier, E., Baker, D., Afgan, E., Stoler, N., ... & Nekrutenko, A. (2014). Dissemination of scientific software with Galaxy ToolShed. Genome Biology, 15 (2), 1-3.

Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. IEEE Transactions on Learning Technologies, 10(4), 405-418. https://doi.org/10.1109/TLT.2017.2740172

References

Boettiger, C. (2015). An introduction to docker for reproduci
ble research, with examples from the R environment. ACM SIGOPS Operating Systems Review, 49(1), 71–79. https://doi.org/10.1145/2723872.2723882

Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman?. Contemporary Educational Psychology, 37(4), 247-253.

Borman, G. D. (2017). Advancing values affirmation as a scalable strategy for mitigating identity threats and narrowing national achievement gaps. Proceedings of the National Academy of Sciences, 114(29), 7486-7488. https://doi.org/10.1073/pnas.1708813114

Borracci, G., Gauthier, E., Jennings, J., Sale, K., & Muldner, K. (2020). The effect of assistance on learning and affect in an algebra tutor. Journal of Educational Computing Research, 57(8), 2032-2052. https://doi.org/10.1177/0735633118822103

Bosch, N., Crues, R. W., Paquette, L., & Shaik, N. (2020). "Hello, [REDACTED]": Protecting student privacy in analyses of online discussion forums. In A. N. Rafferty, J. Whitehill, C. Romero, & V. Cavalli-Sforza (Eds.), Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020) (pp. 39-49). International Educational Data Mining Society. https://pnigel.com/papers/bosch-ilearn-edm20-camera.pdf

Bowers, A. J., Sprott, R., & Taff, S. (2012). Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity, and specificity. The High School Journal, 96(2), 77-100. https://doi.org/10.1353/hsj.2013.0000

Bramucci, R., & Gaston, J. (2012). Sherpa: increasing student success with a recommendation engine. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 82-83)

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. Review of research in education, 24(1), 61-100.

Butler, A. C., Marsh, E. J., Slavinsky, J. P., & Baraniuk, R. G. (2014). Integrating cognitive science and technology improves learning in a STEM classroom. Educational Psychology Review, 26(2), 331-340.

Cai, Z., Graesser, A. C., & Hu, X. (2015). ASAT: AutoTutor script authoring tool. In R. Sottilare, A. C. Graesser, X. Hu, & K. W. Brawner (Eds.). Design recommendations for intelligent tutoring systems – Volume 3: Authoring tools (pp. 199-210). U.S. Army Research Laboratory.

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on Affective Computing, 1(1), 18-37. https://doi.org/10.1109/T-AFFC.2010.1

References

Calvo, R. A., O'Rourke, S. T., Jones, J., Yacef, K., & Reimann, P. (2010). Collaborative writing support tools on the cloud. IEEE Transactions on Learning Technologies, 4(1), 88-97. https://doi.org/10.1109/TLT.2010.43

Castleman, B. L., & Meyer, K. E. (2020). Can text message nudges improve academic outcomes in college? Evidence from a West Virginia Initiative. The Review of Higher Education, 43(4), 1125-1165.
Chen, P., Lu, Y., Zheng, V. W., Chen, X., & Yang, B. (2018). KnowEdu: A system to construct knowledge graph for education. IEEE Access, 6, 31553-31563. https://doi.org/10.1109/ACCESS.2018.2839607

Chetty, R., Friedman, J.N., Hendren, N., Stepner, M., et al. (2020) The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data. White paper. Cambridge, MA: Opportunity Insights. https://opportunityinsights.org/wp-content/uploads/2020/05/tracker_paper.pdf

Chi, M., & Vanlehn, K. (2007). Accelerated future learning via explicit instruction of a problem solving strategy. Frontiers in Artificial Intelligence and Applications, 158, 409-416.

Chopade, P., Khan, S. M., Stoeffler, K., Edward, D., Rosen, Y., & von Davier, A. (2018). Framework for effective teamwork assessment in collaborative learning and problem solving. In C. P. Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. du Boulay (Eds.), Proceedings of the 19th International Conference on Artificial Intelligence in Education (AIED 2018) (pp. 48-59). IOS Press. http://ceur-ws.org/Vol-2153/paper6.pdf

Choudhury, T., & Pentland, A. (2002). The sociometer: A wearable device for understanding human networks. In CSCW'02 Workshop: Ad hoc Communications and Collaboration in Ubiquitous Computing Environments.

Cocea, M., Hershkovitz, A., & Baker, R. S. J. d. (2009). The impact of off-task and gaming behaviors on learning: Immediate or aggregate? In V. Dimitrova, R. Mizoguchi, B. D. Boulay, & A. Graesser (Eds.), Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED 2009) (pp. 507-514). IOS Press. https://doi.org/10.3233/978-1-60750-028-5-507

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. https://doi.org/10.1007/BF01099821

Craig, S. D., Graesser, A. C., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. Journal of Educational Media, 29(3), 241-250. https://doi.org/10.1080/1358165042000283101

D'Mello, S., Craig, S., Fike, K., & Graesser, A. (2009). Responding to learners' cognitive-affective states with supportive and shakeup dialogues. In J. A. Jacko (Ed.), Human-computer interaction: Ambient, ubiquitous, and intelligent interaction (HCI International 2009) (pp. 595-604). Springer. https://doi.org/10.1007/978-3-642-02580-8_65

References

Damgaard, M. T., & Nielsen, H. S. (2018). Nudging in education. Economics of Education Review, 64, 313-342.

Dede, C. (2010). Comparing frameworks for “21st century skills”. In J. A. Bellanca & R. Brandt (Eds.), 21st century skills: Rethinking how students learn (pp. 51-75). Solution Tree Press. https://www.dpsgs.org/pdf/Comparing_Frameworks_for.pdf

Dede, C., Richards, J., & Saxberg, B. (Eds.). (2018). Learning engineering for online Education: theoretical contexts and design-based examples. Routledge. https://doi.org/10.4324/9781351186193

DeFalco, J.A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., Baker, R. S., & Lester, J.C. (2018). Detecting and addressing frustration in a serious game for military training. International Journal of Artificial Intelligence and Education, 28 (2), 152-193. https://doi.org/10.1007/s40593-017-0152-1

Diziol, D., Walker, E., Rummel, N., & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. Educational Psychology Review, 22(1), 89-102. https://doi.org/10.1007/s10648-009-9116-9

Doignon, J. P., & Falmagne, J. C. (2012). Knowledge Spaces. Springer.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don’t. Journal of Applied Psychology, 84(5), 795-805. https://doi.org/10.1037/0021-9010.84.5.795

Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. Computers in Human Behavior, 52, 338-348. https://doi.org/10.1016/j.chb.2015.05.041

Eddy, C. (2013). The effectiveness of Edgenuity when used for credit recovery (Order No. 3606415) [Doctoral dissertation, Texas A&M University]. ProQuest LLC.

Elmadani, M., Mathews, M., & Mitrovic, A. (2012). Data-driven misconception discovery in constraint-based intelligent tutoring systems. In G. Biswas, L. Wong, T. Hirashima, & W. Chen (Eds.), Proceedings of the 20th International Conference on Computers in Education (ICCE 2012) (pp. 1-8). Asia-Pacific Society for Computers in Education. https://core.ac.uk/reader/35468560

Evans, A. C., Wobbrock, J. O., & Davis, K. (2016, February). Modeling collaboration patterns on an interactive tabletop in a classroom setting. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (pp. 860-871)

Fancsali, S. E., Yudelson, M. V., Berman, S. R., & Ritter, S. (2018). Intelligent instructional hand offs. In K. E. Boyer & M. Yudelson (Eds.), Proceedings of the 11th International Conference on Educational Data Mining (pp. 198-207). International Educational Data Mining Society.

References

Fancsali, S. E., Holstein, K., Sandbothe, M., Ritter, S., McLaren, B. M., & Aleven, V. (2020). Towards practical detection of unproductive struggle. In I. I. Bitterncourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), Artificial Intelligence in Education 21st International Conference (AIED 2020) (pp. 92-97). Springer. https://doi.org/10.1007/978-3-030-52240-7_17

Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the assistment system. Technology Instruction Cognition and Learning, 3(1/2), 63. https://sci2s.ugr.es/keel/pdf/specific/articulo/TICL_preview.pdf

Gardner, J., Andres-Bray, M., Brooks, C., & Baker, R. S. (2018). MORF: A framework for predictive modeling and replication at scale with privacy-restricted MOOC data. In Proceedings of the 3rd Workshop on Open Science in Big Data (OSBD 2018). IEEE. https://doi.org/10.1109/BigData.2018.8621874

Gardner, J., Brooks, C., & Baker, R. (2019). Evaluating the fairness of predictive student models through slicing analysis. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK '19) (pp. 225-234). Association for Computing Machinery. https://doi.org/10.1145/3303772.3303791

Gašević, D., Zouaq, A., & Janzen, R. (2013). "Choose your classmates, your GPA is at stake!" The association of cross-class social ties and academic performance. American Behavioral Scientist, 57(10), 1460-1479.

Gobert, J. D., Sao Pedro, M., Raziuddin, J., Baker, R. S. (2013) From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. Journal of the Learning Sciences, 22 (4), 521-563. https://doi.org/10.1080/10508406.2013.837391

Graesser, A. C., Foltz, P. W., Rosen, Y., Shaffer, D. W., Forsyth, C., & Germany, M. L. (2018). Challenges of assessing collaborative problem solving. In Assessment and teaching of 21st century skills (pp. 75-91). Berlin: Springer.

Grawemeyer, B., Mavrikis, M., Holmes, W., Gutiérrez-Santos, S., Wiedmann, M., & Rummel, N. (2017). Affective learning: Improving engagement and enhancing learning with affect-aware feedback. User Modeling and User-Adapted Interaction, 27(1), 119-158. https://doi.org/10.1007/s11257-017-9188-z

Guskey, T. R., & Gates, S. L. (1986). Synthesis of research on the effects of mastery learning in elementary and secondary classrooms. Educational Leadership, 43(8), 73-80. https://uknowledge.uky.edu/cgi/viewcontent.cgi?article=1017&context=edp_facpub

Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. European Journal of Risk Regulation, 4(1), 3-28

References

Hanushek, E. A., Peterson, P. E., Talpey, L. M., & Woessmann, L. (2019). The achievement gap fails to close: Half century of testing shows persistent divide between haves and have-nots. Education Next, 19(3), 8-18. https://www.educationnext.org/achievement-gap-fails-close-half-century-testing-shows-persistent-divide/

Hara, S. R., & Burke, D. J. (1998). Parent involvement: The key to improved student achievement. School Community Journal, 8(2), 9-19.

Harrison, A., Smith, H., Hulse, T., & Ottmar, E. (2020) Spacing Out! Manipulating Spatial Features in Mathematical Expressions Affects Performance. Journal of Numerical Cognition, 6(2), pp. 186-203. https://doi.org/10.5964/jnc.v6i2.243

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. https://doi.org/10.1007/s40593-014-0024-x

Hoadley, C. (2018). A short history of the learning sciences. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), International handbook of the learning sciences (pp. 11-23). Routledge. https://www.academia.edu/36699899/A_short_history_of_the_Learning_Sciences

Hollands, F., & Bakir, I. (2015). Efficiency of automated detectors of learner engagement and affect compared with traditional observation methods. New York, NY: Center for Benefit-Cost Studies of Education, Teachers College, Columbia University

Holstein, K., & Doroudi, S. (2019). Fairness and equity in learning analytics systems (FairLAK). In Companion Proceedings of the Ninth International Learning Analytics & Knowledge Conference (LAK '19). Association for Computing Machinery. http://kenholstein.com/FairLAK_proceedings.pdf

Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V. (2018). The classroom as a dashboard: Co-designing wearable cognitive augmentation for K-12 teachers. In Proceedings of the Eighth International Conference on Learning Analytics and Knowledge (LAK '18) (pp. 79-88). Association for Computing Machinery. https://doi.org/10.1145/3170358.3170377

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), Proceedings of the 20th International Conference on Artificial Intelligence in Education (pp. 157-171). Springer. https://doi.org/10.1007/978-3-030-23204-7_14

Howell, J. A., Roberts, L. D., & Mancini, V. O. (2018). Learning analytics messages: Impact of grade, sender, comparative information and message style on student affect and academic resilience. Computers in Human Behavior, 89, 8-15. https://doi.org/10.1016/j.chb.2018.07.021

References

Huang, Y., Aleven, V., McLaughlin, E., & Koedinger, K. (2020). A general multi-method approach to design-loop adaptivity in intelligent tutoring systems. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), Proceedings of the 21st international Conference on Artificial Intelligence in Education (pp. 124-129). Springer. https://doi.org/10.1007/978-3-030-52240-7_23

Ingebrand, S. W., & Connor, C. M. (2016). Assessment-to-instruction (A2i): An online platform for supporting individualized early literacy instruction. In S. A. Crossley & D. S. McNamara (Eds.), Adaptive educational technologies for literacy instruction (pp. 33-48). Routledge.

Jaggars, S. S., Motz, B. A., Rivera, M. D., Heckler, A., Quick, J.D., Hance, E. A., & Karwischa, C. (2021). The Digital Divide Among College Students: Lessons Learned From the COVID-19 Emergency Transition. Midwestern Higher Education Compact.

Jiang, W., Pardos, Z. A., & Wei, Q. (2019). Goal-based course recommendation. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge (pp. 36-45).

Joksimović, S., Dowell, N., Skrypnyk, O., Kovanović, V., Gašević, D., Dawson, S., & Graesser, A. C. (2015). How do you connect? Analysis of Social Capital Accumulation in connectivist MOOCs. In Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (pp. 64-68)

Kantar, R., McNulty, K., Snow, E. L., Wainess, R., Doshi, S. D., Walker, D. B., & Emery, M. A. (2018). The rise of immersive cognitive assessments: Towards simulation-based assessment for evaluating applicants. In C. P. Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. du Boulay (Eds.), Proceedings of the 19th international conference on Artificial Intelligence in Education (pp. 438-441). Springer. https://doi.org/10.1007/978-3-319-93846-2_82

Karumbaiah, S., Lizarralde, R., Allessio, D., Woolf, B., Arroyo, I., & Wixon, N. (2017). Addressing student behavior and affect with empathy and growth mindset. In X. Hu, T. Barnes, A. Hershkovitz, & L. Paquette (Eds.), Proceedings of the 10th international conference on Educational Data Mining (pp. 96-103). International Educational Data Mining Society. http://educationaldatamining.org/EDM2017/proc_files/papers/paper_79.pdf

Karumbaiah, S., Ocumpaugh, J., & Baker, R. S. (2019). The influence of school demographics on the relationship between students' help-seeking behavior and performance and motivational measures. In C. F. Lynch, A. Merceron, M. Desmarais, & R. Nkambou (Eds.), Proceedings of the 12th international conference on Educational Data Mining (pp. 99-108). International Educational Data Mining Society. https://files.eric.ed.gov/fulltext/ED599236.pdf

Khachatryan, G. A. (2020). Instruction modeling: Developing and implementing blended learning programs. Oxford University Press.

References

Khachatryan, G. A., Romashov, A. V., Khachatryan, A. R., Gaudino, S. J., Khachatryan, J. M., Guarian, K. R., & Yufa, N. V. (2014). Reasoning Mind Genie 2: An intelligent tutoring system as a vehicle for international transfer of instructional methods in mathematics. International Journal of Artificial Intelligence in Education, 24(3), 333-382.

Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016). How deep is knowledge tracing?. In T. Barnes, M. Chi, & M. Feng (Eds.), Proceedings of the Ninth International Conference on Educational Data Mining (pp. 94-101). International Educational Data Mining Society.

Khan, I., & Pardo, A. (2016). Data2U: Scalable real time student feedback in active learning environments. In LAK '16: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (pp. 249-253). Association for Computing Machinery. https://doi.org/10.1145/2883851.2883911

Kim, N. J., Belland, B. R., & Walker, A. E. (2018). Effectiveness of computer-based scaffolding in the context of problem-based learning for STEM education: Bayesian meta-analysis. Educational Psychology Review, 30(2), 397-429. https://doi.org/10.1007/s10648-017-9419-1

Kim, Y. J., Almond, R. G., & Shute, V. J. (2016). Applying evidence-centered design for the development of game-based assessments in physics playground. International Journal of Testing, 16(2), 142-163. https://doi.org/10.1080/15305058.2015.1108322

Kizilcec, R. F., Saltarelli, A. J., Reich, J., & Cohen, G. L. (2017). Closing global achievement gaps in MOOCs. Science, 355(6322), 251-252. https://doi.org/10.1126/science.aag2063

Klinkenberg, S., Straatemeier, M. & van der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. Computers & Education, 57(2), 1813-1824. https://doi.org/10.1016/j.compedu.2011.02.003

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. Educational Psychology Review, 19(3), 239-264.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. Science, 342(6161), 935-937. https://doi.org/10.1126/science.1238056

Koedinger, K. R., & Corbett, A. (2006). Cognitive tutors: Technology bringing learning sciences to the classroom. In K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 61-77). Cambridge University Press. https://wiki.rice.edu/confluence/download/attachments/2765648/KoedingerCorbett05.pdf

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. Cognitive Science, 36(5), 757-798. https://doi.org/10.1111/j.1551-6709.2012.01245.x

References

Koedinger, K. R., McLaughlin, E. A., Jia, J. Z., & Bier, N. L. (2016). Is the doer effect a causal relationship? How can we tell and why it's important. In Proceedings of the sixth international conference on Learning Analytics & Knowledge (LAK '16) (pp. 388-397). Association for Computing Machinery. https://doi.org/10.1145/2883851.2883957

Kreijns, K. (2004). Sociable CSCL environments. Social affordances, sociability, and social presence. Unpublished Masters Thesis, Open University of the Netherlands.

Kroeze, K. A., van den Berg, S. M., Lazonder, A. W., Veldkamp, B. P., & de Jong, T. (2019). Automated feedback can improve hypothesis quality. Frontiers in Education, 3 (116), 1-14. https://doi.org/10.3389/feduc.2018.00116

Krumm, A. E., Beattie, R., Takahashi, S., D'Angelo, C., Feng, M., & Cheng, B. (2016). Practical measurement and productive persistence: Strategies for using digital learning system data to drive improvement. Journal of Learning Analytics, 3(2), 116-138.

Kulik, C. L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. Review of Educational Research, 60(2), 265-299. https://doi.org/10.3102/0034654306000226

Lahti, H., Seitamaa-Hakkarainen, P., & Hakkarainen, K. (2004). Collaboration patterns in computer supported collaborative designing. Design Studies, 25(4), 351-371. https://doi.org/10.1016/j.destud.2003.12.001

Laru, J., & Järvelä, S. (2008). Social patterns in mobile technology mediated collaboration among members of the professional distance education community. Educational Media International, 45(1), 17-32. https://doi.org/10.1080/09523980701847131

Li, H., Gobert, J., Dickler, R., & Moussavi, R. (2018). The impact of multiple real-time scaffolding experiences on science inquiry practices. In R. Nkambou, R. Azevedo, & J. Vassileva (Eds.), 14th International Conference on Intelligent Tutoring Systems Proceedings (pp. 99-109). Springer. https://doi.org/10.1007/978-3-319-91464-0_10

Liang, C., Ye, J., Wu, Z., Pursel, B., & Giles, C. L. (2017). Recovering concept prerequisite relations from university course dependencies. In Proceedings of the Seventh Symposium on Educational Advances in Artificial Intelligence (pp. 4786-4791). Association for the Advancement of Artificial Intelligence. https://clgiles.ist.psu.edu/pubs/AAAI2017-course.pdf

Lipka, J., Hogan, M. P., Webster, J. P., Yanez, E., Adams, B., Clark, S., & Lacy, D. (2005). Math in a cultural context: Two case studies of a successful culturally based math project. Anthropology & Education Quarterly, 36(4), 367-385

References

Liu, R., Koedinger, K., Stamper, J., & Pavlik, P. (2017). Sharing and reusing data and analytic methods with LearnSphere. In T. Barnes, M. Chi, & M. Feng (Eds.), Proceedings of the Ninth International Conference on Educational Data Mining (pp. 328-331). International Educational Data Mining Society. http://dev.stamper.org/publications/LAK19_Koedinger_Stamper_et_al.pdf

Lomas, J. D., Forlizzi, J., Poonwala, N., Patel, N., Shodhan, S., Patel, K., Koedinger, K., & Brunskill, E. (2016). Interface design optimization as a multi-armed bandit problem. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16) (pp. 4142-4153). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858425

Lomas, J. D., Patel, K., Forlizzi, J. L., & Koedinger, K. R. (2013). Optimizing challenge in an educational game using large-scale design experiments. In R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, & G. Olson (Eds.), Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13) (pp. 89-98). Association for Computing Machinery. https://doi.org/10.1145/2470654.2470668

Lovett, M. C. (1998). Cognitive task analysis in service of intelligent tutoring system design: A case study in statistics. In B. P. Goettl, H. M. Halff, C. L. Redfield, & V. J. Shute (Eds.), 4th International Conference on Intelligent Tutoring Systems Proceedings (pp. 234-243). Springer. https://doi.org/10.1007/3-540-68716-5_29

Makhlouf, J., & Mine, T. (2020). Analysis of click-stream data to predict STEM careers from student usage of an intelligent tutoring system. Journal of Educational Data Mining, 12(2), 1-18. https://doi.org/10.5281/zenodo.4008050

Martin, J., & VanLehn, K. (1995). Student assessment using Bayesian nets. International Journal of Human-Computer Studies, 42(6), 575-591. https://doi.org/10.1006/ijhc.1995.1025

Martinez-Maldonado, R., Dimitriadis, Y., Martinez-Monés, A., Kay, J., & Yacef, K. (2013). Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop. International Journal of Computer-Supported Collaborative Learning, 8(4), 455-485. https://doi.org/10.1007/s11412-013-9184-1

Martinez-Maldonado, R., Kay, J., Yacef, K., & Schwendimann, B. (2012). An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment. In International Conference on Intelligent Tutoring Systems Proceedings (pp. 482-492).

Mayer, S. E., Kalil, A., Oreopoulos, P., & Gallegos, S. (2015). Using behavioral insights to increase parental engagement: The parents and children together (PACT) intervention (No. w21602). National Bureau of Economic Research.

References

McBroom, J., Yacef, K., Koprinska, I., & Curran, J. R. (2018). A data-driven method for helping teachers improve feedback in computer programming automated tutors. In C. P. Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. du Boulay (Eds.), Proceedings of the 19th International Conference on Artificial Intelligence in Education (pp. 324-337). Springer. https://doi.org/10.1007/978-3-319-93843-1_24

McLaren, B. M., Lim, S. J., & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In Proceedings of the 30th Annual Conference of the Cognitive Science Society (pp. 2176-2181)

Miller, W. L., Baker, R. S., Labrum, M. J., Petsche, K., Liu, Y. H., & Wagner, A. (2015) Automated detection of proactive remediation by teachers in reasoning mind classrooms. In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK '15), (pp. 290-294). Association for Computing Machinery. https://doi.org/10.1145/2723576.2723607

Milliron, M. D., Malcolm, L., & Kil, D. (2014). Insight and action analytics: Three case studies to consider. Research & Practice in Assessment, 9, 70-89. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.941.5324&rep=rep1&type=pdf

Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. Journal of Educational Data Mining, 4(1), 11-48. https://doi.org/10.5281/zenodo.3554641

Mitrovic, A. (2010). Modeling domains and students with constraint-based modeling. In R. Nkambou, R. Mizoguchi, & J. Bourdeau (Eds.), Advances in Intelligent Tutoring Systems (pp. 63-80). Springer. http://gdac.uqam.ca/inf7470/LivreVersionFinal/03080063.pdf

Motz, B. A., Carvalho, P. F., de Leeuw, J. R., & Goldstone, R. L. (2018). Embedding experiments: Staking causal inference in authentic educational contexts. Journal of Learning Analytics, 5(2), 47-59

Mozer, M. C., & Lindsey, R. V. (2016). Predicting and improving memory retention: Psychological theory matters in the big data era. In M. N. Jones (Ed.), Frontiers of Cognitive Psychology. Big Data in Cognitive Science (pp. 34-64). Routledge/Taylor & Francis Group. https://www.cs.colorado.edu/~mozer/Research/Selected%20Publications/reprints/MozerLindsey2017.pdf

Noel, R., Riquelme, F., Mac Lean, R., Merino, E., Cechinel, C., Barcelos, T. S., Villarroel, R., & Munoz, R. (2018). Exploring collaborative writing of user stories with multimodal learning analytics: A case study on a software engineering course. IEEE Access, 6, 67783-67798. https://doi.org/10.1109/ACCESS.2018.2876801

Nouri, J., Åkerfeldt, A., Fors, U., & Stockholm, S. S. (2017). Assessing Collaborative Problem Solving Skills in Technology-Enhanced Learning Environments-The PISA Framework and Modes of Communication. International Journal of Emerging Technologies in Learning, 12(4)

References

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for educational data mining models: A case study in affect detection. British Journal of Educational Technology, 45(3), 487-501. https://doi.org/10.1111/bjet.12156

Ogan, A., Walker, E., Baker, R., Rodrigo, M. M. T., Soriano, J. C., & Castro, M. J. (2015). Towards understanding how to assess help-seeking behavior across cultures. International Journal of Artificial Intelligence in Education, 25(2), 229-248. https://doi.org/10.1007/s40593-014-0034-8

Ostrow, K. S., & Heffernan, N. T. (2016). Studying learning at scale with the ASSISTments TestBed. In Proceedings of the Third ACM Conference on Learning @ Scale (L@S '16) (pp. 333-334). Association for Computing Machinery. https://doi.org/10.1145/2876034.2893404

Pan, L., Li, C., Li, J., & Tang, J. (2017). Prerequisite relation learning for concepts in MOOCs. In R. Barzilay & M. Y. Kan (Eds.), Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (pp. 1447-1456). Association for Computational Linguistics. https://doi.org/10.18653/v1/P17-1133

Paquette, L., Ocumpaugh, J., Li, Z., Andres, A., & Baker, R. (2020). Who's learning? Using demographics in EDM research. Journal of Educational Data Mining, 12(3), 1-30. https://doi.org/10.5281/zenodo.4143612

Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. British Journal of Educational Technology, 50(1), 128-138. https://doi.org/10.1111/bjet.12592

Park, K., So, H. J., & Cha, H. (2019). Digital equity and accessible MOOCs: Accessibility evaluations of mobile MOOCs for learners with visual impairments. Australasian Journal of Educational Technology, 35(6), 48-63. https://doi.org/10.14742/ajet.5521

Pathak, P. A. (2017). What really matters in designing school choice mechanisms. Advances in Economics and Econometrics, 1, 176-214.

Patikorn, T., & Heffernan, N. T. (2020). Effectiveness of Crowd-Sourcing On-Demand Assistance from Teachers in Online Learning Platforms. In Proceedings of the Seventh ACM Conference on Learning@ Scale (pp. 115-124).

Pavlik, P. I., Cen, H., & Koedinger, K. R. (2009). Performance factors analysis – A new alternative to knowledge tracing. In V. Dimitrova & R. Mizoguchi (Eds.), Proceedings of the 14th International Conference on Artificial Intelligence in Education (pp. 531-538). Association for Computing Machinery. https://doi.org/10.3233/978-1-60750-028-5-531

Pinkard, N. (2001). Rappin'Reader and Say Say Oh Playmate: Using children's childhood songs as literacy scaffolds in computer-based learning environments. Journal of Educational Computing Research, 25(1), 17-34.

References

Rauf, D. S. (2020, October 2). The new, tough expectations education companies face on race and diversity. EdWeek Market Brief. https://marketbrief.edweek.org/market-trends/new-tough-expectations-education-companies-face-race-diversity/

Rawson, K. A., Dunlosky, J., & Sciartelli, S. M. (2013). The power of successive relearning: Improving performance on course exams and long-term retention. Educational Psychology Review, 25 (4), 523-548. https://doi.org/10.1007/s10648-013-9240-4

Razzaq, R., Ostrow, K. S., & Heffernan, N. T. (2020). Effect of immediate feedback on math achievement at the high school level. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), Proceedings of the 21st international conference on Artificial Intelligence in Education (pp. 263-267). Springer. https://doi.org/10.1007/978-3-030-52240-7_48

Ritter, S., Yudelson, M., Fancsali, S. E., & Berman, S. R. (2016). How mastery learning works at scale. In Proceedings of the Third (2016) ACM Conference on Learning @ Scale (L@S '16) (pp. 71-79). Association for Computing Machinery. https://doi.org/10.1145/2876034.2876039

Rodrigo, M.M.T., Baker, R.S.J.d., McLaren, B., Jayme, A., Dy, T. (2012) Development of a Workbench to Address the Educational Data Mining Bottleneck. Proceedings of the 5th International Conference on Educational Data Mining, 152-155.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. Learning and Instruction, 21 (2), 267-280. https://doi.org/10.1016/j.learninstruc.2010.07.004

Roll, I., Briseno, A., Yee, N., & Welsh, A. (2014). Not a magic bullet: the effect of scaffolding on knowledge and attitudes in online simulations. Boulder, CO: International Society of the Learning Sciences

Roscoe, R. D., Snow, E. L., Allen, L. K., & McNamara, D. S. (2015). Automated detection of essay revising patterns: Application for intelligent feedback in a writing tutor. Technology, Instruction, Cognition, and Learning, 10(1), 59-79.

Roscoe, R. D., Snow, E. L., & McNamara, D. S. (2013). Feedback and revising in an intelligent tutoring system for writing strategies. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), Proceedings of the 16th International Conference on Artificial Intelligence in Education (pp. 259-268). Springer. https://doi.org/10.1007/978-3-642-39112-5_27

Rosé, C. P., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. International Journal of Computer-Supported Collaborative Learning, 3(3), 237-271. https://doi.org/10.1007/s11412-007-9034-0

References

Rowe, E., Asbell-Clarke, J., Baker, R. S., Eagle, M., Hicks, A.G., Barnes, T. M., Brown, R. A., & Edwards, T. (2017) Assessing implicit science learning in digital games. Computers in Human Behavior, 76, 617-630. https://doi.org/10.1016/j.chb.2017.03.043

Rowe, E., Asbell-Clarke, J., Bardar, E., Almeda, M.V., Baker, R.S., Scruggs, R., Gasca, S. (2019) Advancing Research in Game-Based Learning Assessment: Tools and Methods for Measuring Implicit Learning. In E. Kennedy, Y. Qian (Eds.) Advancing Educational Research With Emerging Technology, pp. 99-123.

Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. Journal of Educational and Behavioral Statistics, 29(1), 103-116.

Rupp, A. A., Nugent, R., & Nelson, B. (2012). Evidence-centered design for diagnostic assessment within digital learning environments: Integrating modern psychometrics and educational data mining. Journal of Educational Data Mining, 4(1), 1-10. https://doi.org/10.5281/zenodo.3554639

Savi, A., Williams, J. J., Maris, G. K., & van der Maas, H. L. (2017). The role of A/B tests in the study of large-scale online learning. Retrieved March 10, 2021 from https://osf.io/83jsg/

San Pedro, M. O. Z., Baker, R. S. J. d., Bowers, A. J., Heffernan, N. T. (2013) Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), Proceedings of the 6th International Conference on Educational Data Mining (pp. 177-184). International Educational Data Mining Society. https://www.upenn.edu/learninganalytics/ryanbaker/EDM2013_SBBH.pdf

San Pedro, M. O., Baker, R. S., Heffernan, N. T., & Ocumpaugh, J. L. (2015). Exploring college major choice and middle school student behavior, affect and learning: What happens to students who game the system?. In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK '15) (pp. 36-40). Association for Computing Machinery. https://doi.org/10.1145/2723576.2723610

San Pedro, M. O. Z., Liu, R., & McKinniss, T. L. (2019). Developing Game-Based Models of Cooperation, Persistence and Problem Solving from Collaborative Gameplay. In International Conference on Artificial Intelligence in Education (pp. 247-251).

Sao Pedro, M. A., Gobert, J. D., & Betts, C. G. (2014). Towards scalable assessment of performance-based skills: Generalizing a detector of systematic science inquiry to a simulation with a complex structure. In International Conference on Intelligent Tutoring Systems Proceedings (pp. 591-600).

Sass, T. R., Zimmer, R. W., Gill, B. P., & Booker, T. K. (2016). Charter high schools' effects on long-term attainment and earnings. Journal of Policy Analysis and Management, 35(3), 683-706.

Schofield, J. W. (1995). Computers and classroom culture. Cambridge University Press.

References

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Schwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. Journal of Research in Science Teaching, 46 (6), 632-654. https://doi.org/10.1002/tea.20311

Settles, B., & Meeder, B. (2016). A trainable spaced repetition model for language learning. In K. Erk, & N. A. Smith (Eds.), Proceedings of the 54th annual meeting of the Association for Computational Linguistics (pp. 1848-1858). Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-1

Shute, V. J., & Torres, R. J. (2012). Where streams converge: Using evidence-centered design to assess Quest to Learn. In M. Mayrath, J. Clarke-Midura, & D. H. Robinson (Eds.), Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research, (pp. 91-124). Information Age Publishing. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.466.3018&rep=rep1&type=pdf

Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. Computers & Education, 80, 58-67. https://doi.org/10.1016/j.compedu.2014.08.013

Singh, R. P. (2018). Learning analytics: Potential, protection, and privacy in the educational system. In M. K. Singh, Z. Zerihun, & N. Singh (Eds.), Impact of learning analytics on curriculum design and student performance (pp. 1-18). IGI Global. https://doi.org/10.4018/978-1-5225-5369-4.ch001

Slater, S., & Baker, R. S. (2018) Degree of error in Bayesian knowledge tracing estimates from differences in sample sizes. Behaviormetrika, 45 (2), 475-493. https://doi.org/10.1007/s41237-018-0072-x

Sparks, J. R., & Deane, P. (2015). Cognitively based assessment of research and inquiry skills: Defining a key practice in the English language arts. ETS Research Report Series, 2015(2), 1-55

Stamper, J., & Pardos, Z. A. (2016). The 2010 KDD Cup Competition Dataset: Engaging the machine learning community in predictive learning analytics. Journal of Learning Analytics, 3(2), 312-316

Stigler, J. W., & Hiebert, J. (1999). The teaching gap: Best ideas from the world's teachers for improving education in the classroom. Simon and Schuster.
T
aghipour, K., & Ng, H. T. (2016). A neural approach to automated essay scoring. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (pp. 1882-1891).

Tipton, E. (2014) How generalizable is your experiment? Comparing a sample and population through a generalizability index. Journal of Educational and Behavioral Statistics, 39 (6), 478-501. https://doi.org/10.3102/1076998614558486

References

Tuhiwai Smith, L. (2012). Decolonizing methodologies: Research and indigenous peoples (2nd ed.). Zed Books.

Uncapher, M. R., & Cheng, B. H. (2019). Science of learning and learning engineering: Advancing the relationship between learning sciences and teaching practice. In R. Feldman (Ed.), Learning science:

Theory, research, and practice, (pp. 157182). McGraw Hill.

VanLehn, K. (1996). Conceptual and meta learning during coached problem solving. In International Conference on Intelligent Tutoring Systems Proceedings (pp. 29-47). Springer, Berlin, Heidelberg.

Ventura, M., Chang, M., Foltz, P., Mukhi, N., Yarbro, J., Salverda, A. P., Behrens, J., Ahn, J. Ma, T., Dhamecha, T. I., Marvaniya, S., Watson, P., D'helon, C., Tejwani, R., & Afzal, S. (2018). Preliminary evaluations of a dialogue-based digital tutor. In C. P. Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. du Boulay (Eds.), Proceedings of the 19th International Conference on Artificial Intelligence in Education (pp. 480-483). Springer. https://doi.org/10.1007/978-3-319-93846-2_90

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. Computers & Education, 46 (1), 71-95. https://doi.org/10.1016/j.compedu.2005.04.003

Whitehill, J., Williams, J., Lopez, G., Coleman, C., & Reich, J. (2015). Beyond prediction: First steps toward automatic intervention in MOOC student stopout. In O. C. Santos, J. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, M. C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), Proceedings of the 8th International Conference on Educational Data Mining (pp. 171-178). International Educational Data Mining Society. https://doi.org/10.2139/ssrn.2611750

Wilson, K., & Nichols, Z. (2015). The Knewton platform. A general-purpose adaptive learning infrastructure. Knewton. New York, NY, USA. https://www.profijt.nu/wp-content/uploads/2015/09/20150902-White-paper-The-Knewton-Platform.pdf

Wolfe, C. R., Widmer, C. L., Reyna, V. F., Hu, X., Cedillos, E. M., Fisher, C. R., Brust-Renck, P. G., Williams, T. C., Vannucchi, I. D., & Weil, A. M. (2013). The development and analysis of tutorial dialogues in AutoTutor Lite. Behavior Research Methods, 45(3), 623-636. https://doi.org/10.3758/s13428-013-0352-z

Wolfman-Arent, A. (2020, April 1). Coronavirus shutdown reveals inequity of student internet access across Pa. Philadelphia, PA: WHYY. https://whyy.org/articles/coronavirus-shutdown-reveals-inequity-of-student-internet-access-across-pa/

Xhakaj, F., Aleven, V., & McLaren, B. M. (2017). Effects of a teacher dashboard for an intelligent tutoring system on teacher knowledge, lesson planning, lessons and student learning. In European Conference on Technology Enhanced Learning (pp. 315-329).

References

Xia, M., Asano, Y., Williams, J. J., Qu, H., & Ma, X. (2020). Using information visualization to promote students' reflection on "gaming the system" in online learning. In Proceedings of the Seventh ACM Conference on Learning @ Scale (L@S '20) (pp. 37-49). Association for Computing Machinery. https://doi.org/10.1145/3386527.3405924

Yeung, C. K., & Yeung, D. Y. (2018). Addressing two problems in deep knowledge tracing via prediction-consistent regularization. In Proceedings of the Fifth Annual ACM Conference on Learning @ Scale (L@S '18) (pp. 1-10). Association for Computing Machinery. https://doi.org/10.1145/3231644.3231647

Zeidenberg, M. (2015). Valuable learning or "spinning their wheels"? Understanding excess credits earned by community college associate degree completers. Community College Review, 43(2), 123-141. https://doi.org/10.1177/0091552115571595

Zhang, J., Shi, X., King, I., & Yeung, D. Y. (2017). Dynamic key-value memory networks for knowledge tracing. In Proceedings of the 26th International Conference Companion on World Wide Web (pp. 765-774). International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3038912.3052580

Zhou, G., Yang, X., Azizsoltani, H., Barnes, T., & Chi, M. (2020). Improving student-system interaction through data-driven explanations of hierarchical reinforcement learning induced pedagogical policies. In Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '20) (pp. 284-292). Association for Computing Machinery. https://doi.org/10.1145/3340631.3394848

Zhu, M., Liu, O. L., & Lee, H. S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. Computers & Education, 143, 103668. https://doi.org/10.1016/j.compedu.2019.103668

Zou, X., Ma, W., Ma, Z., & Baker, R. S. (2019). Towards helping teachers select optimal content for students. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), Proceedings of the 20th International Conference on Artificial Intelligence in Education, (pp. 413-417). Springer. https://doi.org/10.1007/978-3-030-23207-8_76

Appendix: Participants

Each of these individuals either participated in a virtual session, sent recommendations or comments via email, or spoke directly with the report author. Participants do not necessarily agree with all aspects of this report.

Elizabeth Albro
Pathikrit Banerjee
Michael Binger
Gautam Biswas
Anthony Botelho
Christopher Brooks
Emma Brunskill
Paulo Carvalho
Catherine Cavanaugh
Samuel Crane
Scott Crossley
Kristen DiCerbo
Phillip Grimaldi
Sunil Gunderia
Neil Heffernan
Andrew Jones
Rene Kizilcec
Janet Kolodner
Diane Litman
Bruce McLaren
Benjamin Motz
Phil Poekert
Steve Ritter
Erica Snow
Jim Stigler
Anne Trumbore
Melina Uncapher
John Whitmer
1 participant who requested anonymity