Learning Engineering: A View on Where the Field is at, Where it's Going,

and the Research Needed

Ryan S. Baker[1], Ulrich Boser[2] and Erica L. Snow[3]

[1] Graduate School of Education, University of Pennsylvania

[2] The Learning Agency

[3] Independent Consultant

Author Note

Correspondence concerning this article should be addressed to Ryan S. Baker (ryanshaunbaker@gmail.com), Graduate School of Education, University of Pennsylvania, Philadelphia, PA 19104 and Ulrich Boser (ulrichboser@gmail.com), The Learning Agency, Washington DC 20001, United States.

Abstract

Although the idea of learning engineering dates back to the 1960s, there has been an explosion of interest in the area in the last decade. This interest has been driven by an expansion in the computational methods available both for scaled data analysis and for much faster experimentation and iteration on student learning experiences. This article describes the findings of a virtual convening brought together to discuss the potential of learning engineering and the key opportunities available for learning engineering over the next decades. We focus the many possibilities into ten key opportunities for the field, which in turn group into three broad areas of opportunity. We discuss the state of the current art in these ten opportunities and key points of leverage. In these cases, a relatively modest shift in the field's priorities and work may have an outsized impact.

Keywords: Learning Engineering, Learning, Mastery, Feedback

Introduction

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. Learning engineering applies a rigorous combination of theory, data, and analysis to develop and improve educational systems and methodologies to produce enduring and high-quality learning. In terms of academic disciplines, learning engineering brings together a combination of computer science, data science, cognitive psychology, behavioral science, education, and instructional design. The emerging learning engineering community meets in an increasing number of places, including annual conferences on Adaptive Instructional Systems, the users meetings of the Generalized Intelligent Framework for Tutoring, the Learner Data Institute, the Advanced Distributed Learning community, IEEE ICICLE (standards meetings and one past conference), as well as at more established conferences such as

Artificial Intelligence and Education, Learning Analytics and Knowledge, Educational Data Mining, ACM

Learning @ Scale, Learning Sciences, and the International Conference on Computers in Education.

Though the idea of learning engineering was first introduced by Herbert Simon in the 1960s

(Simon, 1967), the uptake of learning engineering by the broader field of education has been slow among

both researchers and educational technology developers. However, today the use of learning engineering

has increased considerably. This trend has been supported by better tools for authoring and

experimenting with virtual and blended learning environments, a leap forward in the instrumentation of

these learning technologies, the advent and low cost of cloud computing, and advances in both data

science and educational research and design methods. As will be discussed below, there are now many

examples of how a data-intensive learning engineering approach has the potential to benefit learners.

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 article, 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 the

uptake of learning engineering and how these challenges can be addressed, with an eye towards the

research still needed for learning engineering to reach its full potential. To summarize, this article

attempts to identify key areas within the science and engineering of learning that could lead to major

improvements in educational outcomes. We also discuss which enabling advances could be particularly

important to speeding progress in educational research and development.

The structure of the remainder of the article is as follows: We first discuss the convening that

formed the source of our recommendations. Second, we give an executive summary of the convening's

recommendations along with a table summarizing the recommendations further. We then discuss

definitions of learning engineering and some key early successes. The bulk of the paper follows – a detailed

exploration of opportunities within each of the ten categories of recommendation. The paper concludes with a discussion of challenges that cut across the recommendations.

Source of Recommendations

This article's recommendations came about through a community convening where around 30 stakeholders in the field were invited to come together to discuss the potential of the emerging field of learning engineering and key challenges and opportunities for the field, the 2020 Asynchronous Virtual Convening on Learning Engineering. The convening was intended to build on past convenings bringing together researchers and practitioners to discuss more specific aspects of learning engineering, such as the annual convenings organized by the Generalized Intelligent Framework for Tutoring (GIFT) community which resulted in several books on key considerations in learning engineering (available at www.gifttutoring.org).

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 a range of individuals at times of their convenience. It also allowed for multiple rounds of soliciting different opinions, which allowed us to realize in many cases that a key perspective was missing and then solicit it. In this convening, a set of questions was posed to a group of 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 convening or offered written comments. The organizers followed up with clarifying questions on some of the most intriguing ideas, and put together this report summarizing the findings. This report attempts to represent the perspective of many

stakeholders while bringing the ideas into a single voice and set of coherent recommendations. The participants represented a range of different backgrounds and perspectives coming from the places where learning engineering is conducted, studied, and supported: academia (education, computer science, information science, linguistics, and other fields), large and established educational technology companies and publishers, non-profit R&D centers, start-up companies, government, and philanthropy. Participants were recruited on the basis of leadership and expertise in learning engineering or closely related fields. Multiple rounds of recruitment occurred – in several cases, participants in the convening noted that specific expertise was missing and recommended another participant for inclusion. A full list of participants is given in Table A1 (in Appendix A).

Though a range of stakeholders were involved in the convening, teachers, school leaders, parents, and other members of broader society were not represented in the convening (except to the degree that many of the participants were themselves parents and some were former teachers). This article represents the perspectives of 30 stakeholders working in learning engineering and related areas or working with learning engineers in various fashions. Although we attempted to include a cross-section of perspectives, both disciplinary and organizational, there are inherently limitations within an intensive but small-scale effort of this nature. The choice was made to focus on stakeholders who are already familiar with learning engineering in some fashion and therefore are more familiar with its potentials and opportunities. This choice almost certainly omitted key ideas from teachers and school leaders who will be impacted by learning engineering projects. Making fuller efforts to include these voices (e.g. Holstein, McLaren, & Aleven, 2019) will expand the potential of the field to solve the right problems. In addition, due to the substantial differences in educational systems and needs for learning engineering around the world, the decision was made early in the process to focus on the educational system in the United States. As such, most (though not all) of the stakeholders were based in the United States, a limitation to the scope of our findings.

High-Level Findings Summary

The 2020 Asynchronous Virtual Convening on Learning Engineering identified ten key areas of opportunity for learning engineering, which in turn were divided into three broad areas: Better Learning Engineering, Supporting Human Processes, and Better Learning Technologies. Each of these opportunities represented areas where progress had already occurred and successful examples existed. The goal and challenge are to take this work forward and increase its scale of application. The recommendations around this are summarized in Table 1, and will be discussed in detail through the remainder of the paper.

Better Learning Engineering involves improving the process of learning engineering itself to support faster progress across the field. Within the area of Better Learning Engineering, four types of opportunity were identified. First: Enhance R&D Infrastructure in Widely Deployed Platforms, extending automated experimentation and educational data mining to a wider number of learning platforms. Second: Build Components to Create Next-Generation Learning Technologies Faster. Developing a new learning platform with advanced adaptivity currently takes years of effort, and there is substantial duplication of effort across learning platforms. These challenges could be addressed 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.

Third: Use Learning Engineering to Support Diversity and Enhance Equity, moving from research findings on learning technologies and machine-learned models obtained on convenience samples to equitable approaches demonstrated to generalize to broader and more diverse groups of learners. The importance of learning engineering repairing rather than sustaining or amplifying inequities in our society cannot be overstated. Fourth: Bring Learning Engineering to Domain-Based Education Research, enabling

this large community to work with higher-quality rapid research tools and more efficiently share their results.

Supporting Human Processes consists of better engineering the processes that rely upon human judgment, decision making, and activity in education. Within the area of Supporting Human Processes, three types of opportunity were also identified. First: Enhance Human-Computer Systems, 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. Second: Better Engineer Learning System Implementation in Schools. Many learning systems and curricula work well with motivated teachers, supportive administrations, and ongoing support from developers but fail when extended to a broader range of classrooms. Learning engineering can play a role in determining which practices around the use of learning technology are both effective and scalable and using learning analytics to design automated scaffolds for effective practices. Third: Improve Recommendation, Assignment, and Advising Systems, using data science to develop systems that proactively analyze student trajectories and make recommendations that increase the likelihood of successful graduation and life outcomes.

Finally, opportunities in Better Learning Technologies involve improving student-facing learning technologies to improve the effectiveness and overall quality of learning experiences. Within the area of Better Learning Technologies, three types of opportunity were identified. First: Optimize for Robust Learning and Long-Term Achievement, switching from rapid innovation cycles that focus on short-term learning gains to longitudinal work that verifies that designs and algorithms benefit students over a longer period of time and prepare them to learn in new situations. Second: Support Learning 21st-Century Skills and Collaboration, using learning analytics to develop reliable and valid measures of complex skills such as communication, critical thinking, and collaboration and produce learning experiences that support their development. Finally: Improve Support for Student Engagement, leveraging the advances in automatically

detecting student affect and behavioral disengagement and embedding them in engagement/affective interventions that are both effective and acceptable to teachers, school leaders, parents, and students.

Table 1.

Recommendations Summary

Top Ten Opportunities	Examples
BETTER LEARNING ENGINEERING	
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
2. 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
3. Learning Engineering to Support Diversity and Enhance 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
4. 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
SUPPORT HUMAN PROCESSES	
5. 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
6. 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
7. 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
BETTER LEARNING TECHNOLOGIES	
8. 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
9. 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
10. 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

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 (Simon, 1967). There exist several different definitions and perspectives on what learning engineering is or should be, and which disciplinary roots are most essential to this emerging interdisciplinary field. Nonetheless, there is general agreement that learning engineering involves 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.

Learning engineering is about design (Means, 2018). It leverages the learning sciences and cognitive science (Rosé et al., 2019), but also computer science and the rapid developments in the data available and the learning analytics methods available to analyze it (Shimmei & Matsuda, 2020). Both the *science* and the *engineering* of learning have often been treated as subsumed within the broader umbrella of the learning sciences (Nathan & Alibali, 2010). The emergence of learning engineering as an area in its own right creates an opportunity to distinguish learning engineering from learning science. Although learning scientists have long conducted their research through designing artifacts and learning experiences (Cobb et al., 2003), the foremost goal of learning science is the investigation of basic research questions and the development of theory. By contrast, the foremost goal of learning engineering is the enhancement of learning outcomes and related outcomes at scale. If we think of learning engineering as applying the ideas of engineering to learning and ultimately developing a practice of engineering learning, then learning engineering becomes about both effectiveness and efficiency (Dede, Richards, & Saxberg, 2018).

In these terms, our field's efforts to transform learning into an engineering discipline are in their beginnings. Indeed, we are only recently starting to see the emergence of organizations dedicated to applying the rigor of engineering practice to learning (discussed above). Ultimately, the metaphor of

engineering suggests a possible future for education that is reliable, predictable, scalable, and based on known and standardized toolkits and specifications. We remain far from this future, and considerable development and research are needed to reach this point. Developing the tools needed to truly make learning an engineering discipline is a goal that underpins several of the recommendations within this manuscript.

Learning Engineering: Past Successes

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, has already led to 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). Work using KLI has extended well beyond the original team that developed KLI (e.g. Bergamin & Hirt, 2018; Borracci et al., 2020).

Learning engineering and its antecedent work have 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 in systems like SQL-Tutor 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. This approach then bridges naturally into giving students feedback on the errors that they make during problem solving (Mitrovic & Ohlsson, 2016). 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, used in Reasoning Mind, is the practice of engineering automated replications of the instructional practices of expert teachers (Khachatryan, 2020).

Example: 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 (Hattie, 2012) but is often prohibitively time-consuming for teachers. By using learning engineering to iterative improve automating feedback, it is possible to scalably give students timely feedback that is adaptive to their student's current state of knowledge in a domain and topic (van der Kleij et al., 2017).

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). Dashboards used by students 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. The best approach to feedback and 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 effective 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 learning by providing personalized prompts and activity recommendations (Grawemeyer et al., 2017).

Example: Educational Data Mining and Learning Analytics: Better measurement of learning as it occurs

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, with emerging interest by both data scientists (Fischer et al., 2020) and psychometricians (Bergner & Von Davier, 2019)

Simply developing the ability to measure learning as it changed (i.e. Pelanek, 2017) 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; Henderson et al., 2020) 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., 2018; 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 (Moore et al., 2020) or by systematically searching for less-effective learning content (Baker, Gowda, & Salamin, 2018; Peddycord-Liu et al., 2018). This work has also led to the development of interventions and modifications to pedagogy that improve outcomes. While almost all of these interventions are relatively small in scope, 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 in the Learnta system has shown how providing recommendations to instructors can help them select material for their students to work on (Zou et al., 2019). The same type of approaches can show what does not work, also key to iteration. Using UpGrade from Carnegie Learning, the Playpower Labs team found that adding a specific set of "gamification" features to an intelligent tutor actually reduced learner engagement by 15 percent in one context (Lomas et al., 2020).

Top High-Leverage Opportunities

In the following sections, we discuss the findings of the 2020 Asynchronous Virtual Convening on Learning Engineering in terms of ten high-level recommendations as to where the high-leverage opportunities are for learning engineering (see table 1 for a summary). As discussed above, these recommendations are grouped into three broad categories: Enhancing Learning Engineering, Supporting Human Processes, and Developing Better Learning Technologies.

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

One major step towards increasing the positive impact of learning engineering is to improve the infrastructure of the field, the supports available for conducting learning engineering. In this sense, we consider the field's R&D infrastructure to be all of the enabling tools and technologies that make it easier and faster to conduct high-quality R&D. This includes software used to build, study, and iterate learning systems. It also includes software and processes enabling automated experimentation, software that can be quickly embedded in new learning systems to enhance their functionality, and tools and processes for data security, privacy, and compliance. It furthermore includes shared data sets to bring in a wider community of researchers who can help iterate design.

If high-quality learning engineering research can be conducted faster and with less effort, then the benefits of learning engineering can become available faster and more widely. There has already been considerable investment in research and development infrastructure for learning engineering, and the impacts are seen in the rapid deployment of studies through the ASSISTments platform (Ostrow & Heffernan, 2016) compared to the previous more intensive LearnLab model (Koedinger, Corbett, & Perfetti, 2012). The LearnLab model was in turn already 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 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.

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. Currently, even when researchers share their code, it is often difficult to get it to run correctly (Boettiger, 2015). Both LearnSphere (Liu et al., 2017) and the MORF platform (Gardner et al., 2018) have attempted to create ecosystems that function in this fashion. Although LearnSphere is widely used in general, neither platform's ecosystem has achieved widespread use. This may be due both to constraints on what kinds of research is possible in these platforms, as well as 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 increase understanding of 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 (as in Tuovinen & Sweller, 1999)? When is a worked example better than a hint (see McLaren et al., 2014, for instance)? Do students from different cultural backgrounds respond differently and benefit differently from the same seemingly cognitive learning support (as in 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 seem ready 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 still 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.

Beyond this, 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 Savi 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 important to learning engineering. The same experimental infrastructure that supports scientific discovery may also hold benefits for refining and progressively improving a learning system. When doing so, attention must be paid to making sure that a comprehensive set of measures (engagement and longer-term outcomes) are applied when evaluating the results of A/B tests to understand why one condition works better than another.

This infrastructure may help the field 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 have an additive effect or are in fact counterproductive when combined (Koedinger, Booth, & Klahr, 2013; Cromley et al., 2020). Answering this type of question needs considerable data. 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 reinforcement learning algorithms and can be plugged into a variety of learning platforms will speed the process of improving learning platforms. Some variants on reinforcement learning can also help avoid "local maxima," where iteratively making small improvements leads to an improved system, but a different part of the solution space would have been better still.

This sort of experimentation should not be limited to learning scientists and engineers. Teachers can improve their teaching practice by conducting more systematic investigations of their own teaching, a practice seen for decades in Japan (Stigler & Hiebert, 1999). Partnering teachers with learning engineers in these endeavors will benefit both groups.

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 (see Lomas et al., 2013, for instance).

However, this type of infrastructure presents challenges. The first challenge is 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. Broadening the field in this fashion will accelerate the rate of discoveries beneficial to learning engineering.

When it comes to funding, there should also be greater support for private companies opening up their data. A number of organizations currently 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 industry 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 rather than adopting "lowest common denominator" data schemas that work in all platforms but

discard the most useful data. For example, it may be beneficial to identify which variables have been useful in past research efforts and begin to capture them more systematically.

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 make it so that most learning scientists and learning engineers can only ask certain questions, then largely those are the questions that will be asked.

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 (Bettini & Riboni, 2015), 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. For example, efforts to ensure equity in learning systems (recommendation 3) will become infeasible if demographic data is discarded in the name of privacy.

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. 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; replicating this work across platforms therefore has considerable possible benefits. Compliance issues become even more challenging when dealing with cross-border research, where different countries often have different expectations and rules for human subjects protections and privacy. 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).

2. 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 (see, for example, 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. General-purpose architectures such as the GIFT architecture (Sottilare et al., 2017) offer another reasonable approach to sharing components, but require developers to adopt these architectures wholesale. 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, no education-specific algorithm (not even the widely-used Bayesian Knowledge Tracing; Corbett & Anderson, 1995) 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.

Some already widely-used algorithms, like Bayesian Knowledge Tracing (BKT; Corbett & Anderson, 1995) and ELO (Klinkenberg et al., 2011), can quickly and relatively easily be built into robust, full-featured, and implementation-quality toolkits. Beyond that, there are more recently developed algorithms that offer benefits such as the ability to fit complex relationships between items (Zhang et al., 2017), the ability to represent partial overlap between skills (Pavlik et al., 2020), and consideration of memory and forgetting (Mozer & Lindsey, 2016; Settles & Meeder, 2016; Pavlik et al., 2020). 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, 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. There is a lack of production-grade toolkits for this type of modeling that can be quickly leveraged by practitioners.

One additional area where reusable software components would speed progress is in natural language processing. 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 detecting the depth and quality of student participation within on-line discussion forums (Farrow, Moore, & Gasevic, 2020), to the creation of automated conversational agents in systems such as Jill Watson (Ventura et al., 2018), to adapting content for differing student reading levels, as is seen in the Newsela platform (Rand, 2020). 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).

Three broad areas of natural language processing software components could be particularly valuable to learning engineering. First, development to make it easier to integrate existing algorithms

smoothly and seamlessly into educational technologies via APIs, including algorithms for measuring readability (Crossley et al., 2017b), text cohesion and sophistication (Crossley, Kyle, & McNamara, 2016), and algorithms for sentiment analysis (Crossley, Kyle, & McNamara, 2017a). 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, work to develop linguistic analysis resources such as keyword 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.

3. Better Learning Engineering: Learning Engineering to Support Diversity and Enhance Equity

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). The importance of learning engineering repairing rather than sustaining or amplifying inequities in our society cannot be overstated. The need to support diversity and enhance equity interacts in key ways with all of the other recommendations in this report, and learning engineering will not reach its full potential if it does not play a key role in addressing this vital challenge for society.

One approach to making instruction more equitable 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 principle producing a more equitable learning experience through technology-enhanced innovation (Aguilar, 2018).

However, it is not a given that individualized learning will steer instruction and assessment towards equity. 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 (often 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. See recommendation 1 for a discussion of privacy and security issues surrounding the collection and use of this type of data.

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). Similarly, 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). Developing designs that function well where they are applied is much easier when developers conduct participatory and inclusive design, involving members of the communities impacted (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). MOOC outcomes are also generally 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.

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

Designing a high-quality learning system or curricula for a domain depends on having a high-quality model of the content and learning needs in that domain. This need is magnified for adaptive learning systems, which often need to make real-time decisions based on such a model. Ideally, such a model represents the content of a domain, the structure of how that content interrelates (which skills are prerequisite to other skills) and the concepts with which a disproportionate number of students struggle (including misconceptions and preconceptions).

However, the current state of domain modeling in learning engineering is highly uneven between domains, particularly in terms of modeling the higher-level relationships between content (such as prerequisite graphs). There has been considerable work on modeling mathematical domains, with some areas of mathematics having full prerequisite 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 specific concepts (e.g. Bailey & Heritage, 2014; Van Rijn et al., 2014) and there are Bayesian Network models that compose meta-skills but do not model skill prerequisites formally (Shute & Rahimi, 2021). Work in biology has also represented how different concepts interrelate but in a way that supports fine-grained question generation rather than in a coarser fashion that supports curricular sequencing (Olney et al., 2010, 2012). 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 domains such as history or social studies, where it may not be feasible to identify clear prerequisite relationships.

However, recently approaches have been proposed that may be able to capture prerequisites for knowledge graphs more generally, from data sources such as student interaction data, the textual content of learning materials, and even university course catalogs (Liang et al., 2017; Pan et al., 2017; Chen et al., 2018). Providing funding for these approaches has the potential to speed the development and deployment of knowledge graphs to a broader range of domains and grain-sizes, particularly if developers keep humans in the loop to identify spurious correlations. It may also be possible to distill draft versions

of knowledge graphs from curricular documents developed by educational agencies, schools, and standards organizations (e.g. Pruitt, 2014; Sahami & Roach, 2014).

One way to speed efforts in this area would be to create a network of R&D teams 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 (Lomas et al., 2016). The PhET platform in particular has attracted strong attention from domain-based educational researchers for studying learning in their domains (e.g. Correia et al., 2019; Perkins, 2010; Wieman et al., 2010; Yuliati et al., 2018). By bringing domain-based educational researchers together in platforms like this one, we can crowd-source the development of prerequisite graphs and other domain structure models that can then be used at scale.

5. Supporting 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. In best cases, it is possible 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). Examples of this include 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 (Feng & Heffernan, 2006).

Perhaps the greatest immediate opportunity for enhancing human-computer systems in education 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 are used by instructors in a considerable variety of ways (Bull, 2020, p.440), from identifying students in need of immediate learning support (e.g. Miller et al., 2015), to selecting and resdesigning content (Lazarinis & Retalis, 2007), to identifying common mistakes across students (Yacef, 2005), to supporting practices around classroom orchestration (Tissenbaum & Slotta, 2019). Dashboards support teachers in providing students with a broader range of types of feedback and support (Knoop-van Campen, Wise, & Molenaar, 2021). 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 for teachers do not provide 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, & Gobert, 2020). Here, the challenge and opportunity 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. Instead, they want real-time recommendations about when to provide additional support to students and what kind of

support to provide (Martinez-Maldonado et al., 2014; 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. Going further, a key topic 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 and depends on student meta-cognition, and is therefore used unevenly (Fancsali et al., 2018). Current integration of human tutoring also requires tutors to get up to speed quickly based only on information directly provided

by the student. Eventually, through learning engineering, the field 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.

6. Supporting 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 (Wilson et al., 2018). 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, ALEKS, Algebra Nation, and Dreambox made mastery learning a key part of their approach and scaled to hundreds of thousands of learners. 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.

But what are the right practices to help teachers develop? 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). There has been increasing effort to learning from and scale the strategies that exemplary teachers use (Perkins et al., 2012). 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. By also integrating data on teachers' professional development experiences, learning engineers can study which professional development experiences lead to changes in practice and better outcomes for learners.

Hence, creating data systems that connect data on teachers' professional development, their classroom practices, and students' learning experiences will act as a key enabling factor for research and development to improve implementation. Once this infrastructure is in place, support for work to develop a taxonomy of teacher practices in classrooms using learning technology will facilitate the study of 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. Next-generation approaches will leverage far more information than previous very aggregate value-add approaches (compare to Konstantopoulos, 2014) 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.

7. Supporting Human Processes: 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). These approaches 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. These

recommendations may be able to 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); 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 optimally deliver recommendations of this nature (e.g. Castleman & Meyer, 2020). But there are only a few examples of integrating all three of these components together (e.g. Jiang, Pardos, & Wei, 2019).

The key challenge is to take models developed with the purpose of prediction and re-purpose them for a different use, recommendation. Next, 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 and 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 challenge to achieving this goal is the current low degree of explanation provided along with recommendations. Many 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 improved through increased support for research and development on re-purposing prediction models in this space for use in recommendation. There are

two key steps to this. First, research and development 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 (Kay, Zapata-Rivera, & Conati, 2020) -- will require projects that bring together experts in machine learning, human-computer interaction, and education.

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 goal can be achieved 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 (e.g., 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, choices that are more realistic and more likely to lead to personal and career success.

8. 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 the 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, and 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.

However, 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. It becomes even more difficult 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. 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 (Schmidt & Bjork, 1992; Rohrer, 2015; Braithwaite & Goldstone, 2015; Rawson et al., 2013), this limitation in current practice carries risks of worsening outcomes for students rather than improving them. We need to better understand the longer-term impacts of learning engineering decisions. 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 (Rohrer, 2015), and the lack of interleaving of related skills (Braithwaite & Goldstone, 2015). Many learning systems still use massed practice of skills taught in a block, particularly in mathematics, although an increasing number of learning systems are using learning engineering and data to optimize spaced practice (Butler et al., 2014; Settles & Meeder, 2016). However, in doing so, it is important to take into account specific cases where initial massed practice is beneficial, such as the learning of characteristic perceptual features (Carvalho & Goldstone, 2017).

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.

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.

9. 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., 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.

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 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 (Järvelä et al., 2019; 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 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, research and development could further examine the viability of wearable devices that measure social interactions (Montanari et al., 2017). 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; Snow et al., 2015); 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 the field can successfully 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 areas of learning engineering 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. 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 accelerates progress. Attempts to bring together large numbers of researchers and developers 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. Instead, it may be worthwhile to establish 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.

10. 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, for immediate learning (Craig et al., 2004; Cocea et al., 2009), course performance (Cheng et al., 2011; Fincham et al., 2019), and 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 (see reviews in Calvo & D'Mello, 2010; Baker & Rossi, 2013; Baker & Ocumpaugh, 2014; Kovanovic et al., 2016). However, though the technology exists to measure engagement and affect, the technology is not yet in place to reliably use these measurements to improve student experiences. 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 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. Several self-report instruments exist for capturing affect and engagement in real-time or near real-time (Sabourin, Mott, & Lester, 2011; Smith et al., 2018; Hutt et al., 2019). Developing a software plug-in that implements a standard self-report instrument for engagement and affect that has been validated across learner populations will increase the feasibility of collecting large-scale remote data. This self-report data can then be used to develop detectors that recognize student engagement and affect in real-time 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, labeled by culturally appropriate trained coders (Okur et al., 2018).

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 empathic and conversational agents (D'Mello et al., 2009; Robison, McQuiggan, & Lester, 2009; Karumbaiah et al., 2017), to visualizations of student engagement (Arroyo et al., 2007; Xia et al., 2020), to automated feedback on the quality of discussion forum posts (Butcher et al., 2020), to messages provided between learning activities (DeFalco et al., 2018) or to students who have ceased participation in a learning activity (Whitehill et al., 2015).

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 have been successful at reducing disengagement or improving affect are not acceptable to students or teachers, reducing their potential to scale. Greater support for work to understand these individuals' 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.

Conclusion

This article summarizes the findings of the 2020 Asynchronous Virtual Convening on Learning Engineering. In this article, we have outlined ten key areas of opportunity for research and development in learning engineering, and within those areas propose 33 high-potential directions for work, summarized in Table 1. There are overlaps between many of these potential directions. 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 the other areas of opportunity in this article. Encouraging R&D teams to consider multiple of these opportunities in a single project will help to expand coverage and bring the field forward. As this article 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.

Despite considerable thought around how to scale learning technologies (Clarke & Dede, 2009), scaling up remains a challenge for many key innovations that have been developed or enhanced through learning engineering. 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.

Dissemination of ideas remains as much a challenge as scaling up specific learning systems. Many of the findings of learning engineering could apply in new learning platforms and in non-technological learning situations, but remain 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, sharing ideas, and 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 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 the research and development directions within this document, we also recommend continued efforts to enhance connections between research and practice.

Overall, the process of bringing together the 2020 Asynchronous Virtual Convening on Learning Engineering and compiling this report has indicated that learning engineering has considerable promise for enhancing learning experiences, enriching learning, and supporting better long-term achievements by

learners. Strides have already been made; we are at the beginning of the field's journey towards transforming education.

References

Adair, A., Dickler, R., & Gobert, J. (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., 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

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.

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

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

Bettini, C., & Riboni, D. (2015). Privacy protection in pervasive systems: State of the art and technical challenges. Pervasive and Mobile Computing, 17, 159-174.

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

Boettiger, C. (2015). An introduction to docker for reproducible 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

Braithwaite, D. W., & Goldstone, R. L. (2015). Effects of variation and prior knowledge on abstract concept learning. Cognition and Instruction, 33(3), 226-256.

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.

Bull, S. (2020). There are open learner models about!. IEEE Transactions on Learning Technologies, 13 (2), 425-448

Butcher, T., Read, M. F., Jensen, A. E., Morel, G. M., Nagurney, A., & Smith, P. A. (2020). Using an AI-Supported Online Discussion Forum to Deepen Learning. In Handbook of research on online discussion-based teaching methods (pp. 380-408). IGI Global

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

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1699.

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

Cheng, C. K., Paré, D. E., Collimore, L. M., & Joordens, S. (2011). Assessing the effectiveness of a voluntary online discussion forum on improving students' course performance. Computers & Education, 56 (1), 253-261

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

Clarke, J., & Dede, C. (2009). Design for scalability: A case study of the River City curriculum. Journal of Science Education and Technology, 18(4), 353-365.

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational researcher*, *32*(1), 9-13.

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

Correia, A. P., Koehler, N., Thompson, A., & Phye, G. (2019). The application of PhET simulation to teach gas behavior on the submicroscopic level: Secondary school students' perceptions. Research in Science & Technological Education, 37(2), 193-217.

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

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. Behavior research methods, 48(4), 1227-1237.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2017a). Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. Behavior research methods, 49(3), 803-821

Crossley, S. A., Skalicky, S., Dascalu, M., McNamara, D. S., & Kyle, K. (2017b). Predicting text comprehension, processing, and familiarity in adult readers: New approaches to readability formulas. Discourse Processes, 54(5-6), 340-359.

Cromley, J. G., Perez, T., Kaplan, A., Dai, T., Mara, K., & Balsai, M. J. (2020). Combined Cognitive–Motivational Modules Delivered Via an LMS Increase Undergraduate Biology Grades.Technology, Mind, and Behavior, 1 (2).

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

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

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

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). 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.

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

Farrow, E., Moore, J., & Gašević, D. (2020). Dialogue attributes that inform depth and quality of participation in course discussion forums. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (pp. 129-134)

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

Fincham, E., Whitelock-Wainwright, A., Kovanović, V., Joksimović, S., van Staalduinen, J. P., & Gašević, D. (2019). Counting clicks is not enough: Validating a theorized model of engagement in learning analytics. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge (pp. 501-510).

Fischer, C., Pardos, Z.A., Baker, R.S., Williams, J.J., Smyth, P., Yu, R., Slater, S., Baker, R., Warschauer, M. (2020) Mining Big Data in Education: Affordances and Challenges. Review of Research in Education, 44 (1), 130-160.

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

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

Hattie, J. (2012) Visible Learning for Teachers: Maximizing Impact on Learning. Milton Park, UK: Routledge.

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

Henderson, N., Kumaran, V., Min, W., Mott, B., Wu, Z., Boulden, D., ... & Lester, J. (2020). Enhancing Student Competency Models for Game-Based Learning with a Hybrid Stealth Assessment Framework. Proceedings of the International Conference on Educational Data Mining.

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

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

Hutt, S., Grafsgaard, J. F., & D'Mello, S. K. (2019). Time to scale: Generalizable affect detection for tens of thousands of students across an entire school year. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-14)

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.

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2019). What multimodal data can tell us about the students' regulation of their learning process. Learning and Instruction, 4.

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

Kay, J., Zapata-Rivera, D., & Conati, C. (2020). The GIFT of scrutable learner models: why and how. *Design Recommendations for Intelligent Tutoring Systems*, *8*, 25-40.

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

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.

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

Knoop-van Campen, C. A., Wise, A., & Molenaar, I. (2021). The equalizing effect of teacher dashboards on feedback in K-12 classrooms. Interactive Learning Environments, 1-17.

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. 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

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

Konstantopoulos, S. (2014). Teacher effects, value-added models, and accountability. Teachers College Record, 116(1).

Kovanovic, V. Gasevic, D., Dawson, S., Joksimovic, S., Baker, R.S., Hatala, M. (2016) Does Time-on-task Estimation Matter? Implications on Validity of Learning Analytics Findings. Journal of Learning Analytics, 2 (3), 81-110.

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

Lazarinis, F., & Retalis, S. (2007). Analyze me: open learner model in an adaptive web testing system. International Journal of Artificial Intelligence in Education, 17(3), 255-271.

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

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

Lomas, D., Raval, D., Ritter, S., van Dijk, V., Dumoulin, D., Geraedts, P., ... & Patel, N. (2020) Should We Add a Progress Meter? How A/B Testing Can Support Rapid Cycles of Data-Informed Design. Proceedings of the ACM Learning @ Scale Workshop on Educational A/B Testing.

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

Martinez-Maldonado, R., Clayphan, A., Yacef, K., Kay, J. (2014) MTFeedback: Providing Notifications to Enhance Teacher Awareness of Small Group Work in the Classroom. *IEEE Transactions on Learning Technologies, 8,* 187-200.

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.

McLaren, B. M., van Gog, T., Ganoe, C., Yaron, D., & Karabinos, M. (2014). Exploring the assistance dilemma: Comparing instructional support in examples and problems. Proceedings of the International conference on Intelligent Tutoring Systems (pp. 354-361).

Means, B. (2018). Tinkering toward a learning utopia. In C. Dede, J. Richards, B. Saxberg (Eds.), *Learning Engineering for Online Education: Theoretical Contexts and Design-Based Examples.* Ch. 3.

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., & Ohlsson, S. (2016). Implementing CBM: SQL-Tutor after fifteen years. International Journal of Artificial Intelligence in Education, 26(1), 150-159.

Montanari, A., Nawaz, S., Mascolo, C., & Sailer, K. (2017, March). A study of bluetooth low energy performance for human proximity detection in the workplace. Proceedings of the 2017 IEEE International Conference on Pervasive Computing and Communications (PerCom) (pp. 90-99). IEEE.

Moore, S., Stamper, J., Bier, N., & Blink, M. J. (2020). A Human-Centered Approach to Data Driven Iterative Course Improvement. Proceedings of the International Conference on Remote Engineering and Virtual Instrumentation (pp. 742-761).

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

Nathan, M. J., & Wagner Alibali, M. (2010). Learning sciences. Wiley Interdisciplinary Reviews: Cognitive Science, 1(3), 329-345.

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)

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

Okur, E., Aslan, S., Alyuz, N., Esme, A.A., Baker, R.S. (2018) Role of Socio-Cultural Differences in Labeling Students' Affective States. Proceedings of the 19th International Conference on Artificial Intelligence in Education, 367-380.

Olney, A. M. (2010). Extraction of concept maps from textbooks for domain modeling. In *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 390-392). Springer, Berlin, Heidelberg.

Olney, A. M., Graesser, A. C., & Person, N. K. (2012). Question generation from concept maps. *Dialogue & Discourse*, *3*(2), 75-99.

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

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., Eglington, L. G., & Harrell-Williams, L. M. (2020). Generalized Knowledge Tracing: A constrained framework for learner modeling. arXiv e-prints, arXiv-2005.

Peddycord-Liu, Z., Harred, R., Karamarkovich, S., Barnes, T., Lynch, C., & Rutherford, T. (2018). Learning Curve Analysis in a Large-Scale, Drill-and-Practice Serious Math Game: Where Is Learning Support Needed?. Proceedings of the International Conference on Artificial Intelligence in Education (pp. 436-449).

Pelánek, R. (2017). Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. User Modeling and User-Adapted Interaction, 27(3), 313-350.

Perkins, K., Moore, E., Podolefsky, N., Lancaster, K., & Denison, C. (2012). Towards research-based strategies for using PhET simulations in middle school physical science classes. American Institute of Physics Conference Proceedings (Vol. 1413, No. 1, pp. 295-298).

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.

Pruitt, S. L. (2014). The next generation science standards: The features and challenges. Journal of Science Teacher Education, 25(2), 145-156.

Rand, C. (2020) The effect of altered Lexile levels of the same text on reading comprehension. Paper presented at the 70th Annual Conference of the Literacy Research Association.

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

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

Robison, J., McQuiggan, S., & Lester, J. (2009). Evaluating the consequences of affective feedback in intelligent tutoring systems. Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (pp. 1-6).

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.

Rohrer, D. (2015). Student instruction should be distributed over long time periods. Educational Psychology Review, 27(4), 635-643.

Roll, I., Butler, D., Yee, N., Welsh, A., Perez, S., Briseno, A., ... & Bonn, D. (2018). Understanding the impact of guiding inquiry: The relationship between directive support, student attributes, and transfer of knowledge, attitudes, and behaviours in inquiry learning. Instructional Science, 46(1), 77-104.

Rosé, C. P., McLaughlin, E. A., Liu, R., & Koedinger, K. R. (2019). Explanatory learner models: Why machine learning (alone) is not the answer. *British Journal of Educational Technology*, *50*(6), 2943-2958

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

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.

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

Sabourin, J., Mott, B., & Lester, J. C. (2011). Modeling learner affect with theoretically grounded dynamic Bayesian networks. Proceedings of the International Conference on Affective Computing and Intelligent Interactio*n* (pp. 286-295).

Sahami, M., & Roach, S. (2014). Computer science curricula 2013 released. Communications of the ACM, 57(6), 5-5.

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.

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.

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

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological science*, *3*(4), 207-218.

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

Shimmei, M., & Matsuda, N. (2020). Learning a Policy Primes Quality Control: Towards Evidence-Based Automation of Learning Engineering. *Proceedings of the International Conference on Educational Data Mining.*

Shute, V. J., & Rahimi, S. (2021). Stealth assessment of creativity in a physics video game. Computers in Human Behavior, 116, 106647.

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

Simon, H.A. (1967) The job of a college president. *Educational Record, 48,* 68-78.

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

Smith, G. K., Mills, C., Paxton, A., & Christoff, K. (2018). Mind-wandering rates fluctuate across the day: evidence from an experience-sampling study. Cognitive research: principles and implications, 3(1), 1-20.

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). An updated concept for a Generalized Intelligent Framework for Tutoring (GIFT). GIFTtutoring. org, 1-19.

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

Snow, E. L., Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2015). Does agency matter?: Exploring the impact of controlled behaviors within a game-based environment. *Computers & Education*, *82*, 378-392.

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.

Taghipour, 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

Tissenbaum, M., & Slotta, J. (2019). Supporting classroom orchestration with real-time feedback: A role for teacher dashboards and real-time agents. *International Journal of Computer-Supported Collaborative Learning*, *14*(3), 325-351.

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

Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. Journal of Educational Psychology, 91(2), 334-341. doi:10.1037/0022-0663.91.2.334

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.

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

Van Rijn, P. W., Graf, E. A., & Deane, P. (2014). Empirical recovery of argumentation learning progressions in scenario-based assessments of English language arts. Psicología Educativa, 20(2), 109-115

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

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

Wieman, C. E., Adams, W. K., Loeblein, P., & Perkins, K. K. (2010). Teaching physics using PhET simulations. The Physics Teacher, 48(4), 225-227.

Wilson, C. D., Reichsman, F., Mutch-Jones, K., Gardner, A., Marchi, L., Kowalski, S., Lord, T., & Dorsey, C. (2018). Teacher implementation and the impact of game-based science curriculum materials. Journal of Science Education and Technology, 27(4), 285-305

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).

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

Yacef, K. (2005). The Logic-ITA in the classroom: a medium scale experiment. International Journal of Artificial Intelligence in Education, 15(1), 41-62.

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

Yuliati, L., Riantoni, C., & Mufti, N. (2018). Problem Solving Skills on Direct Current Electricity through Inquiry-Based Learning with PhET Simulations. International Journal of Instruction, 11(4), 123-138.

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 A: Participants

Each of these participants took part in the 2020 Asynchronous Virtual Convening process in some fashion. Participants do not necessarily agree with all aspects of this report.

Table A1.

Table of Participants

Elizabeth Albro
Pathrikrit 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
Anonymous Participant