The Current Trade-off Between Privacy and Equity in Educational Technology

Ryan S. Baker

March 8, 2022

Abstract/Synopsis

Current educational technologies can be very effective at improving learning outcomes and student engagement. These technologies increasingly rely upon artificial intelligence to determine what a learner knows, whether they are engaged, and what to do in response to what the system discovers about the student. Making these technologies effective for all learners depends on taking into account the risk of algorithmic bias, where algorithms work less well for learners in some demographic groups. However, specific privacy protection measures currently being adopted can hinder attempts to identify and address algorithmic bias. This chapter will discuss ways that new data analysis and data protection technologies may be able to improve the trade-off between the important values of privacy and equity, offering recommendations for state educational agencies and school districts.

Keywords: Educational technology, privacy, equity, algorithmic bias

Today, the best educational technologies can be very effective at engaging learners and increasing learning outcomes. Educational games such as Zoombinis (Asbell-Clarke et al. 2020), Physics Playground (Shute et al. 2013), and Dragonbox (Siew et al. 2016) can make difficult concepts in computational thinking, physics, and mathematics accessible to students. Intelligent tutoring systems like Mathia (Pane et al. 2014) and ALEKS (Huang et al. 2016) help students learn mathematical reasoning skills, targeting students' time to material they have not yet learned and providing scaffolding for learning. Research on computer supported collaborative learning has supported the enhancement of discussion forums where students learn together (Jeong et al. 2019).

However, there is growing awareness that educational technologies may not work equally effectively for all learners. Repeatedly, technologies and pedagogies that work effectively in some settings fail to function effectively when taken to new settings (Orr 2015). Differences in implementation and usage between schools clearly play a large role in this phenomenon (Tipton & Olson 2018). However, there is also evidence that even the findings these technologies are developed based on may not generalize across all groups of students (Karumbaiah et al, forthcoming).

Within the broader span of educational technologies, there is increasing use of artificial intelligence in education. Artificially intelligent educational technologies can be highly effective at promoting learning (see review in VanLehn 2011), but the algorithms that these systems depend upon can also have algorithmic biases (Baker & Hawn, forthcoming), despite the best intentions of their developers (Holstein et al. 2019). Algorithmic bias refers to the situation when an algorithm performs worse for members of some sub-sets of the population than others, meaning that the predictions and recommendations that follow from the algorithm may be harmful to some even if they are positive, on average. Biases have been documented for many educational technologies (Baker & Hawn, forthcoming).

In order to reduce algorithmic bias, it is necessary to have data on student demographic membership and to retain data for long enough for student outcomes to be clear. However, increasing concerns about student privacy have led to an increasing push for educational technology vendors to not collect identifying data or demographic data, to discard this data after a year, or not to share this data.

Unfortunately, as this chapter will argue, the push to delete or never collect identifying and demographic data is likely to make it much more difficult to investigate inequities and to improve effectiveness for the learners being negatively impacted.

In this chapter, I will discuss each of these trends, considering the value of both protecting privacy and reducing algorithmic bias, and how specific privacy protection measures can hinder attempts to identify and address algorithmic bias. I will then discuss ways that new data analysis and data protection technologies may be able to improve the trade-off between these important values. Finally, the chapter offers recommendations for state educational agencies and school districts.

The Promise of Artificially Intelligent Educational Technology

The first two decades of the 21st century have seen the demonstration, validation, and scaling of several artificially intelligent educational technologies. A relatively large share of the expanded utilization of educational technology during the pandemic consisted of videoconferencing to support continued human teaching, and learning management systems and/or courseware where students watched videos and completed or submitted assignments

(Francom et al. 2021). Nonetheless, the user base of artificially intelligent educational technologies expanded as well.

Some of the key early visions for how artificial intelligence could be used in education suggested that learning systems could leverage the same strategies used by expert tutors in one-on-one in-person tutoring, coining the term *intelligent tutoring systems* for this type of technology (Merrill, Reiser, Ranney, & Trafton 1992; Lepper et al. 1993; McArthur, Stasz, & Zmuidzinas 1990). This led to visions of learning software as perceptive as teachers (Self 1999) and expert tutors (Shute & Psotka 1994). Popular science fiction books such as *Diamond Age* (Stephenson 2003) envisioned one-one real-time learning experiences that would guide children through both their academic and life skills development, from early childhood into adulthood.

Although contemporary learning technologies are not yet as sophisticated as some of these researchers initially hoped (Baker 2016), nonetheless these systems have been an overall substantial success. Contemporary scaled systems tend to each be successful at capturing a single (not always the same) dimension of the sophisticated systems envisioned decades ago (Baker 2016).

The first intelligent tutoring system to clearly document efficacy in real-world classrooms and scale was the Cognitive Tutor (now called Mathia). Developed over the span of more than a decade by cognitive psychologists at Carnegie Mellon University, Mathia now covers the majority of middle school and high school mathematics. Mathia has been the subject of many experimental and quasi-experimental studies, from its earliest usage in Pittsburgh (Koedinger et al. 1997) to nationwide RCTs (Pane et al. 2010, 2014). Recent studies have suggested that some Mathia curricula are highly effective (Pane et al. 2014), results have been mixed for other Mathia curricula (Pane et al. 2010), implementation fidelity and teaching practices make a major difference for its effectiveness (Sales et al. 2016; Sales & Pane 2020), and teachers tend to achieve better results once they are experienced at using it in their classrooms (Pane et al. 2014). When used by experienced teachers in the fashion it was designed for, Mathia leads to deeper conceptual understanding of mathematics and better problem-solving skill, as well as better performance on standardized examinations. Mathia's success is dependent in part on its ability to recognize student knowledge in real-time, information it uses to decide when to advance the student within its curriculum (Corbett 2001).

Several other intelligent tutoring systems and related systems have also scaled and provided evidence for efficacy. ALEKS uses automated algorithms to determine which topics a student is prepared to work on (Cosyn et al. 2021), and has been shown to reduce equity gaps (Huang et al. 2016) and to improve overall outcomes in some studies (Mojarad et al., forthcoming) but also to be very dependent on proper implementation to work well (Phillips et al. 2020). The intelligent simulation Inq-ITS can automatically detect multiple aspects of student science inquiry skill in line with the Next Generation Science Standards and uses this information to scaffold the development of those skills, leading to knowledge that generalizes across contexts (Li et al. 2018). SQL-Tutor has been used by over a hundred thousand students to learn database programming worldwide, using a complex model of domain knowledge to identify student errors (Mitrovic & Ohlsson 2016). A range of other examples of the effectiveness of intelligent tutoring systems can now be found in the literature, with several recent meta-analyses (VanLehn 2011; Steenbergen-Hu & Cooper 2013; Ma et al. 2014; Kulik & Fletcher 2016; Xu et al. 2019).

The artificially intelligent learning systems that are widely used in K-12 classrooms, both in the United States and internationally, typically rely upon AI technology and algorithms from the 1990s or on more recent updates to the same paradigms of adaptivity made possible by those algorithms. The last two decades have seen an explosion of advancement in what can be detected about students in real time, including their disengaged behaviors (Baker & Rossi 2013), their emotion (D'Mello & Graesser 2010), and their self-regulated learning skills (Aleven et al. 2006; Maldonado-Mahauad et al. 2018). These algorithms capture aspects of behavior and student state that can be easy for humans to recognize but hard to explain, leading to the increased use of machine learning algorithms which learn to replicate human judgments. Several experimental lessons and systems have been developed which leverage these algorithms to adapt in real-time to increase educational effectiveness. For example, affectively aware systems have been adapted to help frustrated students improve their self-efficacy (DeFalco et al. 2018), and systems respond to student disengagement by visualizing a student's recent behavior for them (Arroyo et al. 2007; Xia et al. 2020).

Another recent trend is towards systems that use *reinforcement learning* algorithms (reinforcement for the algorithm, not for the child) to learn how to teach more effectively. These systems experiment with their own decisions, figuring out which content and supports to offer to a specific student in a specific situation (Singla et al. 2021). Although K-12 and undergraduate use of these systems is primarily in the early experimental stages (Ausin et al. 2020; MacLellan & Gupta 2021), Amazon now uses algorithms of this nature to increase the time-efficiency of their online training materials for employees (Bassen et al. 2020). In general, adaptive learning technologies make it possible for systems to differentiate their learning support for different students, and reinforcement learning makes it faster for systems to figure out how to do so.

Over the next decades, it seems reasonable to expect increasing usage of more sophisticated automated artificial intelligence in K-12 education in the U.S. and worldwide.

The Risk of Algorithmic Bias

As the use of machine learning and reinforcement learning for education increases, some of the risks now known to exist for these types of algorithms emerges in education as well. Perhaps foremost among these risks is the risk of *algorithmic bias,* where a computer algorithm either replicates a bias found in human behavior and decision-making, or through poor design actually learns to demonstrate new biases on its own. Algorithmic bias has been demonstrated in a range of contexts, from criminal justice (Angwin et al. 2016), to medicine (O'Reilly-Shah et al. 2020), to computer vision (Klare et al. 2012), to hiring (Garcia 2016).

Algorithmic biases have also been documented in the educational use of algorithms. Algorithms have been documented to perform worse for specific demographic groups for a range of applications – predicting student dropout (Anderson et al. 2019), detecting emotion (Ocumpaugh et al. 2014), automated essay scoring (Bridgeman et al. 2012), and many others. Baker and Hawn (forthcoming) review the published literature on algorithmic bias in education, and find that biases manifest not only for variables known to be involved in algorithmic bias in general (race, ethnicity, and gender) but for a range of other variables (rural learners, native language, parental educational background, international students, military-connected students).

Baker and Hawn's review investigates which groups are impacted by algorithmic bias in education. Kizilcec and Lee (forthcoming) demonstrate that much of this bias occurs due to flaws in the algorithms used. However, Karumbaiah (2021) argues that "upstream" sources of bias – stemming from data collection, study design, and the choice of theory – are an equally important source of bias. She notes that educational technologies are often designed using findings derived using data from well-represented populations, and that these findings may not apply to historically marginalized populations (Karumbaiah et al., forthcoming). As another example, upstream bias also emerges when AI algorithms are trained on small or unrepresentative samples, so the resulting predictions that look good on average are not actually as good for specific sub-sections of the population.

However, the findings reviewed by Baker and Hawn (forthcoming) and Karumbaiah (2021) can only be obtained in a context where researchers have access to student demographic variables. Despite the increasing magnitude of research on algorithmic bias in education, most research on educational algorithms does not study the impacts on different groups. In fact, as a recent review by Paquette and colleagues (2020) indicates, most research on educational algorithms does not even mention the demographic attributes of the populations contributing data, much less study whether there is differential effectiveness or impact.

It has been known for decades that our existing educational system produces poorer outcomes for members of historically marginalized communities (see review in Gordon 2007). If the next generation of educational technologies are based on algorithms that are less effective for members of these communities, we will end up perpetuating or even magnifying these inequities. Similarly, if the next generation of educational technologies are based on research findings that only apply to non-marginalized populations, we will perpetuate or even magnify the inequities that exist. If we are to use advanced artificial intelligence technologies to benefit learners, we need to develop these technologies so that they work for all learners and particularly historically underserved learners. We can only do so by developing our models on diverse and representative populations of learners -- and we can only be certain we have avoided algorithmic bias if we look for it. Doing so requires collecting large data sets that include demographic data and other measures of individual characteristics that serve as proxies for the full suite of preferences and differences between learners.

However, another ongoing trend puts this possibility at risk. This trend, entered into with generally very positive motivations and goals, nonetheless creates risks that the field and community of developers will develop a generation of educational technologies that are less effective for historically underrepresented learners. This trend is the push towards prioritizing privacy.

The Push Towards Prioritizing Privacy

Recent years have seen considerable concern about student privacy, both in academia (Slade & Prinsloo 2013; Lynch 2016; Klose et al. 2020) and in public advocacy (NASSP, n.d.; Student Privacy Compass 2021). Both researchers and public advocates have expressed concern about the increasing amount of educational data becoming available at scale, and its uses for both commercial and scientific purposes. Klose et al. (2020) have noted that educational databases, when hacked, have been used for purposes such as identity theft. There has also been concern that educational data could be used to advertise services to students (Lynch 2016; Golightly 2020). Many have also speculated that long-term use of educational data could lead to decisions being made about individuals years or even decades after a poor decision is made (Zeide 2017).

Some have called for deidentification as a way to address these concerns (Ho 2017), and key data sharing initiatives such as the Pittsburgh Science of Learning Center DataShop (Koedinger et al. 2008) and the edX research data eXchange (edX 2021) remove all student identifiers prior to data sharing. However, other researchers have noted that re-identification remains possible for data sets even after student identifiers have been removed, noting a case where a class could be identified based on when they took a field trip (Yacobsen et al. 2021).

This has led some researchers to argue for the use of privacy-protection approaches where student demographic identity variables are removed from data as well (Klose et al. 2020), at minimum in cases where a student is not representative of the class or overall data set they are part of (Bayardo & Agrawal 2005).

Others have sought to put stringent limitations on the sharing of deidentified data – for example, the Student Data Privacy Consortium's template agreement (2021) requires that each time a deidentified data set is shared, it must be approved by each local education agency. While this agreement protects the school district's rights and oversight, the agreement also makes it very onerous to share large multi-school district datasets with a broad span of researchers. Going even further, some have argued for the deletion of all educational data for any purpose after the school semester has ended, and the agreement not to use data for any secondary analysis purpose (EdWeek 2017; see discussion in Laird, Quay-de la Vallee, & Mahesh 2019).

As privacy initiatives go forward, and school districts increasingly mandate that vendors do not collect or use demographic data or put into place agreements that make data sharing or usage onerous, there are likely reductions of risk to student privacy. This reduction of risk does not come for free, however; it is accompanied by an inability to test for and reduce algorithmic bias. In other words, the question is not whether AI will be biased at first – as the process of fine tuning algorithms inevitably involves predictions that deviate from reality in ways that are correlated with individual characteristics – but whether we will have the systems in place to collect the information and the incentives to ensure that the information is used to reduce these biases and make AI algorithms better.

Requiring a vendor to delete student data, or making it difficult to share data, makes it harder to use data for inappropriate purposes. It also makes it much more difficult to hold that vendor accountable for algorithmic biases in their product. Many educational technology vendors have demonstrated responsibility in identifying algorithmic bias in their products, publishing evidence about these limitations and flaws and working with the scientific community to correct them (Bridgeman et al. 2013; Loukina & Buzick 2017; Christie et al. 2019; Baker, Berning, & Gowda 2020). Poorly-designed data privacy agreements make it much harder for external researchers to participate in this process of critique, increasing the risk that algorithmic bias will reduce the effectiveness of educational technologies for historically underrepresented populations. The same is likely for requirements that prevent the collection and use of demographic and identity data – perhaps even more so, since rapid research processes could still find and correct algorithmic bias in data required to be deleted, but if data is never collected, it cannot be used.

The arguments around data privacy are often couched in the language of non-maleficence, the goal of doing no harm. Violations of student privacy create clear opportunities for harm, and this author has no intention of minimizing their importance. However, another non-maleficence goal should also be considered – the goal of preventing algorithmic bias. Educational technology should be effective for all students, and in particular for students from historically marginalized populations. The noble goal of avoiding harm through privacy violations must be balanced with the goal of a fair and equitable educational system.

Alternative Ways to Protect Privacy While Improving Algorithmic Effectiveness

The current situation appears to leave us with a difficult choice: should we take action to protect student privacy, or should we move instead in the direction of using data to identify and fix algorithmic biases? But perhaps the question should not be which of these non-maleficent values we prefer, but instead whether there is a way to have both. Can we protect student privacy while also collecting and using the data that we need to avoid algorithmic biases?

Fortunately, it appears that we can. There are now approaches that enable us to collect, retain, and use the data we need to address algorithmic biases – while reducing (though not entirely eliminating) risks to privacy. Essentially, these approaches come down to carefully managing data rather than eliminating it (or the feasibility of using it) entirely.

School districts today retain a great deal of data on learners. This data is necessarily personally identifying – schools need to be able to track their students' grades, disciplinary incidents, and standardized examination scores. This data contains information that could be embarrassing or problematic if it were released – often much more so than the data held by educational technology vendors. School district data security is often imperfect – hacks, malware, and ransomware has plagued school districts in recent years (Lopez 2021), much as it has impacted broader society. And yet, we neither call for school districts to discard all their data on an annual basis, nor prevent school districts from collecting demographic information, nor try to make it very difficult for school districts to let external researchers use that data for appropriate research uses. The risk of security breaches in educational vendors is real. At the same time, it is not immediately obvious that a company – which can be put out of business by lawsuits – has less of an interest in data security than a school district does.

The key is to develop secure ways to make data available for research, without redacting that data to the point where it is no longer useful to test for algorithmic bias. There are legitimate concerns around data sharing, but the goal ultimately is the *use* of data, not sharing it. An example of this distinction can be seen in our data infrastructure (named MORF) at the University of Pennsylvania. We currently store our data from Massive Online Open Courses in a repository that enables research on complete, unredacted data while making it highly difficult (hopefully infeasible) to extract that data (Gardner et al. 2018). Our open-source infrastructure allows researchers to submit data analysis programs to the infrastructure; a wide range of data analysis programs (across programming languages and data analysis tools) can be used, but only a restricted set of output functions can be used. New output functions cannot be added to the system without careful hand review by the technical team. This functionality allows researchers to run a range of analyses on student data without exposing that student data directly to researchers, reducing the risk of privacy breaches.

Our prototype has been used to attempt to replicate a number of past findings and analyses of algorithms (Andres et al. 2017; Gardner et al. 2019). It is imperfect in terms of usability and has thus far been difficult for researchers to learn to use outside of our team and a small number of pilot universities. However, it establishes a paradigm where vendors and school districts could make their data available for analysis – analysis on algorithmic bias, using a full range of demographic and identity variables – with substantially reduced risk of data disclosure and privacy violation.

Our infrastructure may not be the eventual infrastructure adopted by this sector (first research projects seldom are), but this overall paradigm – analysis on secured servers, with no direct access to personal data – has the potential to find a better compromise between privacy and reducing algorithmic bias than solutions that emphasize one goal over the other.

Recommendations for State Educational Agencies and School Districts

The need to find a trade-off between the non-maleficent goals of protecting privacy and avoiding algorithmic bias leads this author to offer a few recommendations to state educational agencies and school districts:

1. **Provide demographic data to vendors for the purpose of checking for algorithmic bias.** If a technology is leading to worse outcomes for some demographic groups or making worse predictions for those demographic groups, districts adopting that technology and community members need to know that (and we need to fix the problem!). Of course, this data should be transmitted and stored in a secure fashion, but it is needed if we are to verify whether algorithmic bias is occurring.
2. **Incentivize vendors to conduct algorithmic bias audits, or conduct them directly.** Where the expertise is available, school districts and state agencies can conduct audits for algorithmic bias. However, this may not be feasible in many cases: not every school district will have the capacity to conduct an algorithmic bias audit. As the developer of the algorithm, it is generally easier for the vendor to do an algorithmic bias audit, if they have the necessary demographic data. There are also economies of scale in doing so. External researchers may also be able to conduct these audits, a practice often seen for evaluations of efficacy. Best practices for the audit should be used, of course, with measures and groups agreed to in advance, and evidence on how the analysis was conducted being open for inspection.
3. **Rather than asking vendors to delete data, ask them to secure it.** The practice of requiring vendors to delete data at the end of every school year reduces the risk of privacy leaks, but also makes it impossible to check for longer-term equity issues. Many algorithms' accuracy and impacts can only be determined using longitudinal data (i.e. the accuracy of a predictive analytics model), so data should be kept until it is clear it will no longer be needed for this purpose.

4. **Encourage vendors to adopt data infrastructures that enable privacy-protecting analyses.** Accountability can be increased by allowing school district/state researchers and other external researchers to inspect and/or conduct algorithmic bias analyses. Doing so with current data infrastructures is often difficult to do securely, and the transfer of data creates privacy risks. Encouraging vendors to develop or adopt data infrastructure that enables analysis while securing data (like the aforementioned repository at the University of Pennsylvania) will enable a better trade-off between privacy and reducing algorithmic bias than is currently possible today.

Conclusions

In this chapter, I have discussed the trade-off between two concerns, both coming from a value of non-maleficence: protecting privacy and preventing algorithmic bias. I have briefly reviewed the push towards protecting privacy in education, and some of the unintended consequences that current steps to protect privacy may be causing. I have discussed the emerging evidence that algorithmic bias is a major challenge to the effectiveness of new AI technologies for education. These technologies are in many cases beneficial overall, but may be less effective for specific groups of students, due to algorithmic bias. Current steps to protect privacy may make it much more difficult to prevent algorithmic bias.

This chapter details why specific privacy protection measures may make it harder to identify or address algorithmic bias. It discusses the possibility that new approaches to data storage and use may help to create a better balance between protecting privacy and preventing algorithmic bias. These new data storage and analysis technologies are still in the early stages of research and development, but given the fast developments ongoing in data science, they may be available to vendors, school districts, and state education agencies fairly soon. In tandem, conducting research on which variables should be collected will help us to comprehensively prevent algorithmic bias. There remains a need for further study of which groups are impacted by algorithmic bias in which situations (Baker & Hawn, forthcoming). Simply making it possible to use the variables that are already being collected would make a positive difference; a move towards collecting a broader research-based range of variables will make it possible to more comprehensively prevent algorithmic bias.

Overall, the right approach to balancing privacy and equity remains open to question, but both goals must be attended to. Otherwise, the promise of artificial intelligence for education may instead result in educational technologies that amplify the inequities that already exist in our educational system.

References

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

Anderson, H., Boodhwani, A., & Baker, R. S. (2019). Assessing the Fairness of Graduation Predictions. Proceedings of the 12th International Conference on Educational Data Mining, 488–491.

Andres, J.M.L., Baker, R.S., Siemens, G., Gasevic, D., Spann, C.A. (2017) Replicating 21 Findings on Student Success in Online Learning. *Technology, Instruction, Cognition, and Learning, 10* (4), 313-333.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine Bias: there's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminalsentencing%0A

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., ... & Woolf, B. P. (2007, June). Repairing disengagement with non-invasive interventions. In *AIED* (Vol. 2007, pp. 195-202).

Asbell-Clarke, J., Rowe, E., Almeda, V., Edwards, T., Bardar, E., Gasca, S., ... & Scruggs, R. (2021). The development of students' computational thinking practices in elementary-and middle-school classes using the learning game, Zoombinis. *Computers in Human Behavior*, *115*, 106587.

Ausin, M. S., Maniktala, M., Barnes, T., & Chi, M. (2020, July). Exploring the impact of simple explanations and agency on batch deep reinforcement learning induced pedagogical policies. In *International Conference on Artificial Intelligence in Education* (pp. 472-485). Springer, Cham.

Baker, R.S. (2016) Stupid Tutoring Systems, Intelligent Humans. *International Journal of Artificial Intelligence and Education*, 26 (2), 600-614.

Baker, R. S., Berning, A., & Gowda, S. M. (2020). Differentiating Military-Connected and NonMilitary-Connected Students: Predictors of Graduation and SAT Score. EdArXiv. https://doi.org/10.35542/osf.io/cetxj

Baker, R.S., Hawn, M.A. (in press) Algorithmic Bias in Education. To appear in *International Journal of Artificial Intelligence and Education*.

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*. U.S. Army Research Lab, Orlando, FL, pp. 155-166, 2013

Bassen, J., Balaji, B., Schaarschmidt, M., Thille, C., Painter, J., Zimmaro, D., ... & Mitchell, J. C. (2020, April). Reinforcement learning for the adaptive scheduling of educational activities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Bayardo, R. J., & Agrawal, R. (2005). Data privacy through optimal k-anonymization. In *21st International conference on data engineering (ICDE'05)* (pp. 217-228).

Bridgeman, B., Trapani, C., & Attali, Y. (2012). Comparison of Human and Machine Scoring of Essays: Differences by Gender, Ethnicity, and Country. Applied Measurement in Education, 25(1), 27–40.

Christie, S. T., Jarratt, D. C., Olson, L. A., & Taijala, T. T. (2019). Machine-Learned School Dropout Early Warning at Scale. *Proceedings of the International Conference on Educational Data Mining*.

Corbett, A. (2001). Cognitive computer tutors: Solving the two-sigma problem. In *International Conference on User Modeling* (pp. 137-147). Springer, Berlin, Heidelberg.

Cosyn, E., Uzun, H., Doble, C., & Matayoshi, J. (2021). A practical perspective on knowledge space theory: ALEKS and its data. *Journal of Mathematical Psychology*, *101*, 102512.

D'mello, S. K., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, *20*(2), 147-187.

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.

EdWeek (2017) Maryland Dad Wants June 30 to be "National Student Data Deletion Day". Retrieved Dec 20, 2021 from https://www.edweek.org/technology/20aryland-dad-wants-june-30-to-be-national-student-data-deletion-day/2017/06

edX (2021) Using the Research Data Exchange Data Package. Retrieved Dec. 20, 2021 from https://edx.readthedocs.io/projects/devdata/en/latest/rdx/index.html

Francom, G. M., Lee, S. J., & Pinkney, H. (2021). Technologies, challenges and needs of K-12 teachers in the transition to distance learning during the COVID-19 pandemic. *TechTrends*, *65*(4), 589-601.

Garcia, M. (2016). Racist in the Machine: The Disturbing Implications of Algorithmic Bias. World Policy Journal, 33(4), 111–117.

Gardner, J., Andres-Bray, M., Brooks, C., Baker, R. (2018) MORF: A Framework for Predictive Modeling and Replication at Scale With Privacy-Restricted MOOC Data. *Proceedings of the 3rd Workshop on Open Science in Big Data.*

Gardner, J., Yang, Y., Baker, R., Brooks, C. (2019) Modeling and Experimental Design for MOOC Dropout Prediction: A Replication Perspective. *Proceedings of the 12th International Conference on Educational Data Mining*, 49-58.

Golightly, D. (2020) Google, New Mexico AG Spar Over Chromebook Student Data Collection. https://www.androidheadlines.com/2020/02/google-new-mexico-attorney-general-lawsuit-student-data-collection-chromebook.html

Gordon, E. W. (2007). *Narrowing the Achievement Gap: Strategies for Educating Latino, Black, and Asian Students*. Springer Science & Business Media.

Ho, A. (2017). Advancing educational research and student privacy in the “big data” era. In *Workshop on Big Data in Education: Balancing the Benefits of Educational Research and Student Privacy* (pp. 1-18). National Academy of Education Washington, DC

Holstein, K., Wortman Vaughan, J., Daumé, H., Dudik, M., & Wallach, H. (2019). Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–16.

Huang, X., Craig, S. D., Xie, J., Graesser, A., & Hu, X. (2016). Intelligent tutoring systems work as a math gap reducer in 6th grade after-school program. *Learning and Individual Differences*, *47*, 258-265.

Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of computer-supported collaborative learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational research review*, *28*, 100284.

Karumbaiah, S. (2021) The Upstream Sources of Bias: Investigating Theory, Design, and Methods Shaping Adaptive Learning Systems. Unpublished Doctoral Dissertation, University of Pennsylvania.

Karumbaiah, S., Ocumpaugh, J., Baker, R.S. (in press) Context Matters: Differing Implications of Motivation and Help-Seeking in Educational Technology. To appear in *International Journal of Artificial Intelligence and Education.*

Kizilcec, R. F., & Lee, H. (in press). Algorithmic Fairness in Education. Algorithmic Fairness in Education. In W. Holmes & K. Porayska-Pomsta (Eds.), Ethics in Artificial Intelligence in Education. Abingdon-on-Thames, UK: Taylor & Francis.

Klare, B. F., Burge, M. J., Klontz, J. C., Bruegge, R. W. V., & Jain, A. K. (2012). Face Recognition Performance: Role of Demographic Information. IEEE Transactions on Information Forensics and Security, 7(6), 1789–1801.

Klose, M., Desai, V., Song, Y., & Gehringer, E. (2020). EDM and Privacy: Ethics and Legalities of Data Collection, Usage, and Storage. Proceedings of the International Conference on Educational Data Mining.

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

Koedinger, K., Cunningham, K., Skogsholm, A., & Leber, B. (2008). An open repository and analysis tools for fine-grained, longitudinal learner data. *Proceedings of the International Conference on Educational Data Mining.*

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of educational research*, *86*(1), 42-78.

Laird, E., Quay-de la Vallee, H., Mahesh, A. (2019) *Balancing the Scale of Student Data Deletion and Retention in Education*. Washington, DC: Center for Democracy and Technology Retrieved Dec. 20, 2021 from https://cdt.org/wp-content/uploads/2019/03/Student-Privacy-Deletion-Report.pdf

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

Li, H., Gobert, J., Dickler, R., & Moussavi, R. (2018). The impact of multiple real-time scaffolding experiences on science inquiry practices. In International Conference on Intelligent Tutoring Systems (pp. 99-109). Springer.

Lopez, Gil Abraham, "INVESTIGATING THE RANSOMWARE INFECTION RATE OF K12 SCHOOL DISTRICTS DURING THE COVID PANDEMIC" (2021). *Electronic Theses, Projects, and Dissertations*. 1317. https://scholarworks.lib.csusb.edu/etd/1317

Loukina, A., & Buzick, H. (2017). Use of Automated Scoring in Spoken Language Assessments for Test Takers With Speech Impairments. *ETS Research Report Series*, 2017(1), 1–10.

Lynch, C. F. (2017). Who prophets from big data in education? New insights and new challenges. *Theory and Research in Education*, *15*(3), 249-271.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of educational psychology*, *106*(4), 901.

MacLellan, C. J., & Gupta, A. (2021). Learning Expert Models for Educationally Relevant Tasks Using Reinforcement Learning. *International Educational Data Mining Society*.

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*, *80*, 179-196.

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

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

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

Mojarad, S., Baker, R.S., Essa, A., Stalzer, S. (in press) Replicating Studying Adaptive Learning Efficacy using Propensity Score Matching and Inverse Probability of Treatment Weighting. To appear in *Journal of Interactive Learning Research*.

National Association of Secondary School Principals (NASSP) (n.d.) Student Data Privacy. Retrieved Dec. 20, 2021 from https://www.nassp.org/top-issues-in-education/position-statements/student-data-privacy/

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.

O'Reilly-Shah, V. N., Gentry, K. R., Walters, A. M., Zivot, J., Anderson, C. T., & Tighe, P. J. (2020). Bias and ethical considerations in machine learning and the automation of perioperative risk assessment. British Journal of Anaesthesia, 125(6), 843–846.

Orr, L. L. (2015). 2014 Rossi award lecture: Beyond internal validity. *Evaluation review*, *39*(2), 167-178.

Pane, J. F., McCaffrey, D. F., Slaughter, M. E., Steele, J. L., & Ikemoto, G. S. (2010). An experiment to evaluate the efficacy of cognitive tutor geometry. *Journal of Research on Educational Effectiveness*, *3*(3), 254-281.

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

Phillips, A., Pane, J. F., Reumann-Moore, R., & Shenbanjo, O. (2020). Implementing an adaptive intelligent tutoring system as an instructional supplement. *Educational Technology Research and Development*, *68*(3), 1409-1437.

Sales, A. C., Wilks, A., & Pane, J. F. (2016). Student Usage Predicts Treatment Effect Heterogeneity in the Cognitive Tutor Algebra I Program. *Proceedings of the International Conference on Educational Data Mining*.

Sales, A. C., & Pane, J. F. (2020). The Effect of Teachers Reassigning Students to New Cognitive Tutor Sections. *Proceedings of the International Conference on Educational Data Mining*.

Self, J. (1999). The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. *International journal of artificial intelligence in education*, *10*(3-4), 350-364.

Siew, N. M., Geofrey, J., & Lee, B. N. (2016). Students' algebraic thinking and attitudes towards algebra: the effects of game-based learning using Dragonbox 12+ App. *The Research Journal of Mathematics and Technology*, *5*(1), 66-79.

Singla, A., Rafferty, A. N., Radanovic, G., & Heffernan, N. T. (2021). Reinforcement Learning for Education: Opportunities and Challenges. *Proceedings of the International Conference on Educational Data Mining*.

Shute, V. J., & Psotka, J. (1994). *Intelligent Tutoring Systems: Past, Present, and Future*. Technical Report, Armstrong Lab, Brooks Air Force Base. US Air Force.

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in newton's playground. *The Journal of Educational Research*, *106*(6), 423-430.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510-1529.

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of educational psychology*, *105*(4), 970.

Stephenson, N. (2003). *The Diamond Age: Or, a Young Lady's Illustrated Primer*. Spectra.

Student Data Privacy Consortium (2021) *Standard Student Data Privacy Agreement (NDPA Standard Version 1.0) Version 1r7*. Retrieved Dec 21, 2021 from https://cdn.ymaws.com/www.a4l.org/resource/resmgr/files/sdpc-publicdocs/final_sdpc_ndpa_v1-7.pdf

Student Privacy Compass (2021) Student Privacy Primer. Accessed Dec. 20, 2021 from https://studentprivacycompass.org/resource/student-privacy-primer/

Tipton, E., & Olsen, R. B. (2018). A review of statistical methods for generalizing from evaluations of educational interventions. *Educational Researcher*, *47*(8), 516-524.

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

Xia, M., Asano, Y., Williams, J. J., Qu, H., & Ma, X. (2020, August). Using Information Visualization to Promote Students' Reflection on" Gaming the System" in Online Learning. In *Proceedings of the Seventh ACM Conference on Learning@ Scale* (pp. 37-49).

Xu, Z., Wijekumar, K., Ramirez, G., Hu, X., & Irey, R. (2019). The effectiveness of intelligent tutoring systems on K-12 students' reading comprehension: A meta-analysis. *British Journal of Educational Technology*, *50*(6), 3119-3137.

Yacobson, E., Fuhrman, O., Hershkovitz, S., & Alexandron, G. (2021). De-identification is Insufficient to Protect Student Privacy, or–What Can a Field Trip Reveal?. *Journal of Learning Analytics*, *8*(2), 83-92.

Zeide, E. (2017). The structural consequences of big data-driven education. *Big Data*, *5*(2), 164-172.