Fairness and Equity in Learning Analytics Systems (FairLAK)

Kenneth Holstein\(^1\) and Shayan Doroudi\(^2\)

\(^1\)Human Computer Interaction Institute, \(^2\)Computer Science Department
Carnegie Mellon University
{kjholste, shayand}@cs.cmu.edu

**ABSTRACT:** The potential for data-driven algorithmic systems to amplify existing social inequities, or create new ones, is receiving increasing popular and academic attention. A surge of recent work, across multiple researcher and practitioner communities, has focused on the development of design strategies and algorithmic methods to monitor and mitigate bias in such systems. Yet relatively little of this work has addressed the unique challenges raised in the design, development, and real-world deployment of learning analytics systems. This interactive workshop aims to provide a venue for researchers and practitioners to share work-in-progress related to fairness and equity in the design of learning analytics and to develop new research and design collaborations around these topics. The workshop will begin with a brief overview of research in fair AI and machine learning, followed by presentations of accepted and invited contributions. In addition, a key outcome of the workshop will be a research agenda for the LAK community, around fairness and equity. Workshop participants will collaboratively construct this agenda through a sequence of small- and whole-group design activities. At the end of the workshop, participating researchers and practitioners will then explore opportunities for collaboration around specific research and design thrusts within this agenda.

**Keywords:** fairness; equity; algorithmic bias; real-world impact; critical perspectives; human factors; design; ethics; AI; machine learning; cross-disciplinarity

1 BACKGROUND

Data-driven algorithmic systems increasingly influence every facet of our lives, including the quality of healthcare we receive, who receives a job or a loan, whose livelihoods are automated away, who is released from jail, and who is subjected to increased policing (e.g., Barocas & Selbst, 2016; Veale, Van Kleek, & Binns, 2018). In recent years, the potential of such systems to amplify existing social inequities, or even to create new ones, has received a surge of popular and academic attention. It is now commonplace to see popular press articles about algorithmic bias in high-stakes applications such as loan granting, hiring, recidivism prediction, and predictive policing (e.g., Giang, 2018; Lohr, 2018). Interdisciplinary research communities have emerged with a focus on understanding and mitigating such risks – most notably the Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML)\(^1\) and the nascent FAT* community\(^2\).

\(^1\) https://www.fatml.org/

\(^2\) https://fatconference.org/
Despite this widespread attention to fairness and bias in data-driven algorithmic systems, communities such as FAT/ML and FAT* have thus far tended to focus heavily on a relatively small set of high-stakes application domains such as the examples mentioned above (Green & Hu, 2018; Holstein et al., 2018; Veale et al., 2018). In particular, relatively little work has focused on educational contexts, where increasing use of learning analytics and AI raises unique challenges not commonly faced in other domains (Ocumpaugh, Baker, Gowda, Hansen & Reich, 2015; Heffernan, & Heffernan, 2014; Ito, 2017). For example, while most existing fairness auditing and “de-biasing” methods require access to sensitive demographic information (e.g., age, race, gender) at an individual-level, such information is often unavailable to learning analytics practitioners in practice (Holstein et al., 2018; Kilbertus et al., 2018). In addition, it can sometimes be challenging to define what “equitable” outcomes might look like (Hansen & Reich, 2015; Ito, 2017), in contexts where a learning analytics system results in disparate outcomes across student subpopulations (e.g., students coming in with lower or higher prior knowledge).

The Learning Analytics and Knowledge (LAK) community has long been interested in the ethical dimensions of data-driven educational systems (e.g., Draschler et al., 2015; Sclater & Bailey, 2015; Tsai & Gasevic, 2017). However, the focus has often been on institutional and policy level considerations, including concerns around data ownership and privacy. As multidisciplinary conversations around algorithmic fairness and bias proceed at a rapid pace, it is critical that they are not proceeding without us. It is crucial not only that the learning analytics community is aware of advances in understanding and mitigating undesirable algorithmic bias, but also that our community is actively contributing to these conversations. In addition to advancing the field of learning analytics, such direct engagement may help push the broader literature on algorithmic fairness forward, by presenting domain-specific nuances that need to be addressed or by challenging some of the literature’s core assumptions from an educational perspective. This workshop is particularly well suited for this year’s LAK conference, given the theme of promoting inclusion and success.

2 WORKSHOP OBJECTIVES AND INTENDED OUTCOMES

The primary goals of this workshop are as follows:

A. Cross-disciplinary ‘translation’: Introduce LAK researchers and practitioners to the state-of-the-art in fairness and bias in data-driven algorithmic systems.

B. A venue to share relevant research and practice: Provide a venue for researchers and practitioners to share in-progress research/design work or on-the-ground experiences related to algorithmic fairness and bias in learning analytics systems.

C. Visioning / Developing a research agenda: Collaboratively develop a research agenda for more equitable learning analytics, based on the open problems and directions identified by workshop participants.

D. Researcher and practitioner ‘matchmaking’: Helping participants identify opportunities for fruitful researcher-researcher and/or researcher-practitioner collaborations.

We will disseminate the shared research agenda developed at the workshop, along with other workshop outcomes, via a Twitter hashtag (#FairLAK). In addition, outcomes will be disseminated via
one or more blog posts (which will also be shared over social media, such as Twitter) and through a potential joint paper with workshop participants for LAK 2020 or the Journal of Learning Analytics.

3 WORKSHOP ORGANIZATION

Type of event: Half-day workshop

Type of participation: Participation for the first FairLAK workshop will be ‘mixed’: both participants with a paper submission (following an open call) and other interested members of the LAK community will be welcome to attend.

Schedule:

A. Introductions and background (~30 minutes): Workshop organizers will present high-level workshop objectives. Participants will briefly introduce themselves and share their personal objectives for the workshop. Then the organizers will provide a rapid overview of existing work on fairness in data-driven algorithmic systems. Researchers and practitioners will learn about existing, state-of-the-art methods (from FAT/ML and related literatures in machine learning, statistics, and human-computer interaction) to audit real-world learning analytics systems for potentially harmful biases, and strategies/methods to mitigate such biases.

B. Presentations of accepted and invited contributions (~80 minutes): Three accepted presentations and three invited presentations (8 minutes each, with 5 minutes for questions and discussion)

C. Collaborative group work (60 minutes):

C.1 Small-group discussions: Problem-finding (20 minutes): Participants will identify pressing open issues around fairness and equity in learning analytics systems, collecting issues on sticky notes in small-group discussions

C.2 Whole-group discussion: Sharing open problems and envisioning possible solutions (20 minutes): Groups will share the issues they have identified, synthesizing issues through affinity diagramming

C.3 Small-group discussions: Turning ‘possible solutions’ into research agendas for the LAK community (20 minutes): Groups will gather around particular areas of the growing affinity diagram (dynamically and self-selected, based on areas of interest), to discuss specific issues that interest them in greater detail – this time generating ideas for possible solutions and/or research projects

D. Synthesis, speed dating, and next steps (40 minutes):

D.1 Whole-group discussion: Developing a shared research agenda for fair learning analytics (20 minutes): Based on the activities above, the organizers will help groups
synthesize their ideas into a shared research agenda (i.e., a call to action for the LAK community, consisting of several concrete research and design directions)

**D.2 Speed Dating and Closing Notes** (20 minutes): Researchers and practitioners will circulate throughout the room, engaging in brief conversations with others to begin exploring concrete opportunities for collaboration

**REFERENCES**


Ethics in Praxis: Socio-Technical Integration Research in Learning Analytics

Kyle M. L. Jones
Indiana University-Indianapolis (IUPUI)
kmlj@iupui.edu

Chase McCoy
Indiana University-Bloomington
chamccoy@indiana.edu

ABSTRACT: Learning analytics create beneficial opportunities to reimagine educational institutions, pedagogy, and learning experiences. However, it is unclear whether or not the processes that will create these benefits take into consideration ethical issues, such as fairness and privacy. For learning analytics to be legitimate, there is a need to make sure that data practices and technological designs limit downstream harms as much as possible. To do so, some have argued that transparency, auditing, and participatory design can achieve this goal; we argue that there is an opportunity to address ethical concerns upstream using a method from science and technology studies: Socio-technical integration research.

Keywords: Ethics, fairness, interventions, socio-technical integration research, social science methods

1 INTRODUCTION

Universities are enmeshed in ubiquitous information technologies that serve their educational aims and the administration of highly bureaucratic institutions. These assemblages of databases, applications, systems, sensors, and other technical artifacts have led to an undeniable increase in data quantity and with it the potential to transform data into actionable insights using machine learning, descriptive statistics, and predictive models. Often under the umbrella term of learning analytics, researchers, practitioners, and administrators are working to explore how methods derived from data science can potentially impact, if not transform, higher education. While learning analytics present a significant opportunity to examine pedagogy and learning outcomes, in addition to institutional structure and management practices, the positive benefits learning analytics may reap come with significant ethical questions. Research has emerged to address these questions, but the scholarly field and practitioner discipline need methods to identify and resolve these problems in the day-to-day work of learning analytics—not just at a theoretical level.

Our approach in this paper is to present a method from science and technology studies (STS)—socio-technical integration research, or STIR—and its potential to identify and influence ethics in praxis. What we mean by “ethics in praxis” is not just the instantiation of a moral choice in everyday learning analytics work. Instead, our focus is on how a STIR participant (a learning analytics
practitioner) defines, justifies, and acts on a moral perspective, embedding that perspective in and using that perspective to guide the work the practitioner does as a means to ends aligned with learning analytics. In brief, a STIR study situates social scientists alongside learning analytics practitioners to engage the latter in questions about their ethics in praxis and the social consequences of their work.

We begin this paper by arguing that the ethical questions surrounding learning analytics have been focused primarily on privacy concerns, and rightfully so. However, there are other problems worth investigation, including how practitioners make choices that protect students from harmful consequences and treat them fairly. Instead of focusing on effects downstream of learning analytics, we contend that addressing ethics in praxis upstream could be useful; to do just that, we outline the STIR method. Potential applications of STIR studies with learning analytics practitioners follow. Finally, we summarize our STIR study of an institutional researcher and conclude the paper with recommendations for the learning analytics community.

2 TOWARDS FAIRNESS IN LEARNING ANALYTICS

2.1 More Than Privacy

Information ethics scholars and learning analytics researchers have taken up some ethical concerns as they relate to informational privacy, but less so questions of fairness. Naturally, the creation of new information flows—many of which contain granular, identifiable data about student life—in support of learning analytics have raised student privacy concerns, and this area of the literature has demonstrated conceptual and theoretical rigor in ways that are having notable impacts on, inter alia, how institutions grapple with privacy problems in their policies (see Pardo & Siemens, 2014; Rubel & Jones, 2016). Other researchers are examining socio-technical solutions to scaffold important informed consent strategies in technological designs (see Prinsloo & Slade, 2015). However, the privacy literature has, with notable exceptions, only touched on issues of fairness (see Prinsloo & Slade, 2016; West, Huijser, & Heath, 2016). One way to understand fairness issues is to consider who benefits from learning analytics and whether or not the distribution is just.

2.2 Fairness and Just Distributions of Benefits

The capture and analysis of data representing students' social, intellectual, and physical behaviors that drive learning analytics lead us to ask two important questions related to fairness. First, what benefits accrue, for whom are they distributed, and is the distribution justifiable? We can easily imagine situations where data derived from student life is used to support administrative aims (e.g., efficiency, effectiveness, political gains)—but not positive learning outcomes and experiences for those whose lives are made transparent for data analysis purposes. Our second question homes in on processes informing learning analytics. Regardless of the actual benefits and how they are distributed, will the processes by which learning analytics insights are created directly benefit students and protect them from harm? If learning analytics are not beneficent and attuned to particular harmful consequences, then they cannot be considered fair practices. The first question attends to distributive justice, while this second question raises concerns about procedural justice and ethics in praxis—our focus for this paper.
2.3. Examining Downstream Effects

One way to shore up the legitimacy of learning analytics is to bring to light data practices and their results (e.g., interventions, predictions) to determine whether or not such things are justifiable. In information ethics and critical data studies, research efforts have focused on improving transparency around black-boxed data artifacts (see Citron & Pasquale, 2014). The general argument for doing so is a Brandeisian one: Transparency will hold those who create, distribute, and implement data artifacts more accountable; consequently, accountability will resolve discriminatory and/or deceptive practices and increase fairness (Ananny & Crawford, 2016). One weakness of this approach is that it tends to place its attention on downstream practices and artifacts that are already established and mature. As a result, technological recommendations and policy suggestions attempt to slow down and reverse that which has technological momentum. While we support this type of research and these ongoing initiatives, we also believe directing research on learning analytics upstream could lead to fairer, ethically sensitive technological designs and practices.

2.4. Addressing Ethics in Praxis Upstream

Important efforts are being made by researchers and designers to design learning analytics systems and data artifacts with particular users in mind, and other work—such as that which takes a participatory/co-design strategy—develops learning analytics hand-in-hand with actual users in the design stage. For instance, Zhu, Yu, Halfaker, and Terveen (2018, p. 2) suggest a novel “value-sensitive algorithm design” process, which “engages relevant stakeholders in the early stages of algorithm creation and incorporates stakeholders’ tacit values, knowledge, and insights into the abstract and analytical process of creating an algorithm.” These efforts are crucial for identifying and resolving ethical problems upstream before they are baked into learning analytics technologies. However, successful these approaches may become, they cannot fully account for socially situated practice.

Socio-technical user studies have time and again demonstrated that tool use is dependent on the social context in which the user is situated (see Oudshoorn & Pinch, 2003). And while participatory/co-design/value-sensitive design strategies of learning analytics can account for one aspect of upstream ethics, particular uses (or non-uses as may be the case) of these technologies depends on conditions, norms, and values that are sometimes hard to identify and account for in design. Moreover, not all practices depend on specific learning analytics technologies. In fact, it is still commonplace that data visualizations, statistical models, and other analytic practices are done using off-the-shelf applications (e.g., Tableau, SPSS, Excel). As a result, upstream interventions also need to address how practitioners interact with tools in support of learning analytics and account for the social context in which practitioners work day to day.

An approach of this sort requires researchers to get into the very spaces and places where consequential decisions are made about how to make students data and consider them as data artifacts (Jones & McCoy, 2018). More importantly, such an approach would need to go beyond descriptive studies of what is happening and move towards actively intervening in analytic work. In so doing, practitioners would be prompted by researchers to become reflexive about their practices.
and the consequences thereof to make responsive practical and ethical modulations. We argue that the socio-technical integration research (STIR) method can lead to positive upstream engagement and useful modulations.

3 THE SOCIO-TECHNICAL INTEGRATION RESEARCH METHOD

3.1 Who and What to STIR

In socio-technical integration research (STIR), social scientists embed themselves within a research context to actively engage with researchers by probing and encouraging them to reflect on the societal dimensions and implications of their practices. STIR was initially developed to provide laboratory scientists the opportunity to pair with social scientists in order to enable collaboration between them and to aid laboratory scientists in unpacking “the social and ethical dimensions of research and innovation in real time and to document and analyse [sic] the results” (Fisher, n.d., p. 76).

Social scientists STIR participants (practitioners participating in a study) to provide opportunities for them to reflect on what they are doing, why they are doing it, and how they could do things differently, with the end goal being that participants will actively modulate their behavior by considering the social aspects of their work. During their time together, the STIR researcher seeks to elicit “reflexive awareness,” or an attentiveness to “the nested processes, structures, interactions, and interdependencies, both immediate and more removed, within which they operate” (Fisher, Mahajani, & Mitcham, 2006, p. 492) for the STIR participants. Participants are encouraged to reflect upon three areas of their practice: considerations; alternatives, and outcomes.

Considerations refer to the particulars of their practice, including the goals and values of their work, as well as the social, political, and technological resources from which they draw for support. Alternatives are practices that differ from the participants’ current ones but could impact the trajectory of their work if they were to be adopted. Finally, with outcomes participants are encouraged to reflect upon the outcomes of their work and if different decisions, approach, resources, and people could influence their practice. As STIR participants reflect upon these three areas and related societal and ethical dimensions, opportunities emerge for participants to recognize their socio-ethical position, which in turn leads to "goal-directed" (p. 492) modulations that directly impact the participant’s current practices.

3.2 Modulations: De Facto, Reflexive, and Deliberate

Modulations in STIR occur in three stages: de facto, reflexive, and deliberate. De facto modulations are the implicit societal and ethical dimensions that shape research participants’ everyday work practices and exist prior to a STIR. The STIR approach assumes that participants do not actively reflect on whether these dimensions are efficacious or in alignment with their norms and values or those of the social context that guides their practices, because there is no incentive to do so. Reflexive modulations are those that arise because of heightened awareness as the participant is probed to consider the societal and ethical dimensions of their practice and the consequences. In these cases, these dimensions are made explicit by the participant, and they begin to notice how
social influences (e.g., actors, politics, values, resources, etc.) interact with their given practice. In the final stage, deliberate modulations, participants act upon their reflexive modulations to make changes to their practices. These deliberate modulations may simply influence the efficiency and effectiveness of their work; however, deeper level modulations—which is the goal of STIR-ing—lead to altered goals, objectives, and assumptions of a project due to an enhanced awareness to the societal implications of their practice.

4 STIR AND LEARNING ANALYTICS

Social scientists can use socio-technical integration research (STIR) to uncover the societal and ethical dimensions of learning analytics practitioners as they build systems, develop data-based artifacts, and deploy analytic strategies (e.g., algorithms, models); see Figure 1.

![Figure 1. A model showing how a STIR of a learning analytics practitioner could lead to de facto, reflexive, and deliberate modulations.](image)

But given that learning analytics is often embedded in and supportive of complex institutional bureaucracies (e.g., higher education), defining who is a learning analytics practitioner can be challenging. Unlike original STIR studies where it was quite obvious that the laboratory was the context and bench scientists were the participants, with learning analytics it can be difficult to make these methodological choices. Below, we make some recommendations for STIR-ing learning analytics practitioners.

Academics. In the spirit of original STIR studies, STIR-ing learning analytics could be done with research teams building learning analytics artifacts. Such individuals represent academic
departments and cross-institutional collaborations. Findings may reveal varying degrees of ethical sensitivity among different types of researchers (e.g., doctoral students, postdocs, tenured faculty).

Mathematical and Computational Scientists. Individuals responsible for programmatic and algorithmic code effectively write some of the rules of individual behaviors and determine the information they use to evaluate themselves (van Dijk & Poell, 2013). STIR-ing these practitioners could help them better understand how they embed their values and that of the institutions for whom they are designing learning analytics systems.

Interface and User Experience Technologists and Instructional Designers. Practitioners focused on human-computer interaction processes are steeped in affective and persuasive computing methods, which are often used to elicit particular user responses using design strategies and messaging campaigns (e.g., nudging). A STIR study of these practitioners could surface the ethical justifications designers make to, say, limit choice sets or educate students about predictive scores.

Educational Technologists and Instructional Designers. Technologists and designers in educational institutions are in unique positions to educate instructors on how to use learning analytics tools. STIR-ing these individuals could raise their awareness about student privacy issues, among other things.

Institutional Researchers, Registrars, and Other Information Professionals. In higher education, the deployment and successful diffusion of learning analytics tools and practices are impacted by various information professionals who access, steward, and analyze sensitive institutional information. STIR-ing these practitioners could develop interesting findings regarding their decision making around information disclosure and institutional politics.

Apropos to the last category of learning analytics practitioners above, in the next section, we will discuss a longitudinal STIR we conducted on a single institutional researcher engaged in developing learning analytics data artifacts for their institution’s administration. The study will be explicated further in a forthcoming publication, but for this workshop paper, we will briefly discuss our preliminary findings.

5 STIR-ING A LEARNING ANALYTICS PRACTITIONER

To understand how STIR can help to uncover learning analytics ethics in praxis, we conducted a STIR study of a single institutional researcher at a mid-sized public university. The participant’s responsibilities entail, among other things, conducting statistical analyses on important administrative metrics, such as retention, recruitment, and enrollment, and providing this information to their institution’s administration. The STIR focused on assessing the potential value of the approach for uncovering and better understanding this practitioner’s upstream privacy practices.

Over four months, we conducted 12 in-person and virtual interviews with the participant. We developed a STIR interview protocol based on elements in Figure 1 to guide the participant to reflect on their privacy practices and those of their staff within their office. The interviews sought to elicit
from the participant the considerations, alternatives, and outcomes of their work, and to uncover the three types of modulations and instances where their practices were modified to more explicitly consider privacy or, at the least, brought about ideas for future privacy-focused initiatives. Furthermore, during the interviews, the participant often shared data artifacts, such as an ongoing project on enrollment projections and trends, while discussing the data practices associated with their everyday work.

The participant’s de facto modulations revealed that they value privacy in their work in regards to ensuring that they and their staff follow privacy policies set by FERPA and their institution. However, the participant’s reflexive modulations uncovered that these guiding policies insufficiently address privacy issues in practice. The participant became aware that many of their office’s and institution’s actual privacy practices are not addressed in the policies, particularly, for example, in regards to how identifiable student data should be distributed throughout the institution, and who should be allowed access to sensitive student information. The participant reflected on how this policy lacuna has led to data access and distribution practices that differ between them and their colleagues throughout their institution.

The learning analytics practitioner’s reflexive modulations gave rise to deliberate modulations. Here, not only did the participant became aware of the need to have more formal institutional and departmental policies to guide privacy in praxis, but they began the process of documenting their privacy practices. By working with other institutional actors engaged in learning analytics, conversations within in the practitioner’s institution and within their office have begun around creating explicit institutional and departmental documents on with whom and how data should be shared within their institution. Furthermore, the participant stated the planned to establish opportunities, such as at an office retreat or during team meetings, for their staff to document their privacy practices.

6 CONCLUSION

In this paper, we have argued that socio-technical integration research (STIR) presents new opportunities to investigate how learning analytics practitioners define, justify, and act on their moral perspectives—their ethics in praxis. Our study suggestions and the summary of our forthcoming research provide insights into what STIR may accomplish. Yet, the learning analytics community may benefit from more structure to begin STIR studies of their own and, more importantly, adopt a reflective perspective about their ethics in praxis.

To increase the adoption of the STIR method, we see an opportunity to develop a multi-faceted research and training agenda. First, social scientists addressing ethical issues associated with learning analytics could develop a research agenda to further explore STIR and related intervention methods, as well as plan strategic STIR studies. Such an agenda could be developed at a pre-conference workshop or special research retreat, among other things. Should this agenda gain traction, learning analytics and STIR experts could develop training materials for non-STIR experts. While STIR is a rigorous method, we believe that it does not take advanced qualitative research training to learn its intricacies and apply its techniques. Non-research learning analytics practitioners could learn how to STIR and conduct STIR evaluations at their place of work.
The ethical issues associated with learning analytics are many and consequential. From privacy to discrimination, bias to fairness, and many others, these concerns deserve serious attention to ensure that learning analytics technologies are designed and deployed in ways that further the educational mission of higher education and protect its primary stakeholder group—students—and others from harm. Scholarly efforts to date have cataloged many of these issues, and in so doing they have recommended sound policy principles. While there is still more work to do on this front, it is arguably time to shift efforts to focus on how such ethical concerns materialize and are accounted for in everyday practice.

REFERENCES


Early-adopter Iteration Bias and Research-praxis Bias in the Learning Analytics Ecosystem

Michael J. Meaney  
University of Cambridge  
Arizona State University  
mjm234@cam.ac.uk

Tom Fikes, Ph.D.  
Arizona State University  
tgfikes@asu.edu

ABSTRACT: By devising a conceptual framework of the learning analytics ecosystem, we identify two types of bias that may stymie the efforts of leveraging learning analytics to produce fair and equitable virtual learning environments. First, Early-adopter Iteration Bias may lead learning analytics to derive insights about optimal course design based on preferences and behavior patterns of more prepared, lower need learners. Second, Research-praxis Bias prevents practitioners from properly utilizing insights derived from learning analytics and research.

Keywords: Educational equity, Human computer interaction, Interaction design, Design bias

1 INTRODUCTION

In the context of open-scale courses (including Massive Open Online Courses, or MOOCs), the learning analytics ecosystem has the potential to provide valuable insights into teaching and learning for virtual learning environments (VLEs) (Nguyen et al., 2017). This may be especially valuable for scaling low-barrier, individualized learning experiences that can reach traditionally underrepresented populations or other high-need students (Aguilar, 2018). The broader educational systems in which learning analytics are embedded, however, give rise to multiple sources of bias that may stymie the efforts to develop these courses into fair and equitable VLEs. First, an Early-adopter Iteration Bias may unintentionally lead to design recommendations that serve already well-educated and well-represented learners (Meaney and Fikes, 2018). Because analytics and research inform practice, if the data are not adequately disaggregated, and heterogenous effects considered, conclusions will be biased toward the majority and drive the innovation and optimization of the courses to further favor these students, potentially disadvantaging underrepresented learners, or other high-need students. Second, Research-praxis Bias, whereby the producers of VLEs do not properly benefit from learning analytics and research insights into VLEs, might further prevent VLEs from meeting the needs of underrepresented or other high-needs learners (Meaney, 2018). A depiction of the learning analytics ecosystem that highlights these sources of design bias is illustrated in Figure 1.
Figure 1: A model of the learning analytics ecosystem illustrating two sources of bias. The universe of students who could benefit from VLEs contains a high proportion of less prepared, higher needs students. *Early-adopter Iteration Bias* describes the situation in which courses designed for traditional higher-education students lead students from more prepared, lower need backgrounds to disproportionately enter VLEs and then succeed at higher rates. The data corpus produced by VLEs reflects the population of more prepared, lower need learners; and learning analytics and research conducted on this corpus produces results biased toward the majority. *Research-praxis Bias* describes the situation in which producers of VLEs receive insights from learning analytics and the research community that is driven by the more prepared, lower need majority, leading to innovation and optimization of VLE design that is even further away from the needs of less prepared, higher needs students. This is further complicated by the general disconnect between the research and practice communities.

2 EARLY-ADOPTER ITERATION BIAS

Early-adopter Iteration Bias is a conceptual model we are introducing to account for a series of processes and constraints that optimize open-scale course production for more prepared, lower need learners. The intuition is grounded in Rogers’ (2010) notion that early adopters of technology will often have population characteristics different to that technology’s later users, which may be the actual target population. Learning analytics of massive data sets have focused on behavior patterns of the average student, who are (we suggest) early-adopters who are more likely to be already well-educated (Rohs and Ganz, 2015; van de Oudeweetering and Agirdag, 2018). This leads optimization and design recommendations to be driven by insights derived from users less likely to need help. If future open-scale course iterations continue to be optimized based on present usage patterns of early-adopters, and if these usage patterns continue to reflect the needs and behaviors of more prepared, lower need learners, this could further exacerbate educational inequity by disadvantaging less prepared, higher need learners. Early-adopter Iteration Bias is illustrated in Figure 2.
Figure 2: The diffusion of innovations is a concept developed by Rogers (2010). The theory suggests that innovations diffuse across society along different segments of the population, sequentially: innovators, early adopters, early majority, late majority, and laggards. Rogers notes that early adopters of new technologies will more likely be well-educated and wealthier. These users have access to more and better information, coupled with a higher tolerance of risk for new products. Early adopters are also likely to have disposable income and are a more attractive target market toward which to design new products. Innovations are iterated and optimized based on data available from early adopters.

Given the disproportionate rate of already well-educated learners using open scale courses and other low barrier VLEs, it is possible that Early-adopter Iteration Bias has already entered the learning analytics ecosystem. We created a graphic highlighting the educational attainment of users studied in eight learning analytics papers over the past few years. Nearly 80% of users already held a college degree, as illustrated in Figure 3 (data cited from: Robinson et al., 2015; Dillahunt et al., 2015; Christensen et al., 2013; van de Oudeweetering and Agirdag, 2018; Ho et al., 2015; Wang et al., 2018).

Arizona State University’s Global Freshman Academy (ASU GFA) stands out for attracting a higher proportion of less prepared, higher need students. These courses offer university credit eligibility and earned-admission to ASU Online, and are intentionally designed to attract non-traditional learners without a post-secondary degree. Even still, more than half of learners in this VLE are more prepared, lower need learners.

Scaling low-barrier, individualized learning experiences that can reach traditionally underrepresented populations or other high-need students requires not only new marketing strategies, but also VLE content and pedagogy to suit the needs of these students. Analyzing data and deriving insights biased toward the majority of existing users may actually undermine this aim.
3 RESEARCH-PRAXIS BIAS

Research-praxis Bias compounds the potential problems from Early-adopter Iteration Bias, in two separate but interrelated ways. The first source of Research-praxis Bias is straightforward: practitioners who utilize the research and insights of the learning analytics community potentially embed into the design of the courses recommendations and conclusions derived from skewed data privileging behavior patterns of more prepared, lower need students.

The second source of Research-praxis Bias is more nuanced and complex. It was unveiled through recent qualitative work, a pilot study interviewing practitioners producing VLEs at ASU (Meaney, 2018), and builds off of the noted “chasm” between research and practice in the development of VLEs (Price et al., 2016; Bakharia et al., 2016). The qualitative study notes three important insights that may be worth further consideration and investigation and that, indeed, contribute to a Research-praxis Bias in the learning analytics ecosystem that may hinder aims of inclusion and equity.

First, it was discovered that relatively little is known about the production process of VLEs. Little research has been conducted to examine how the particular mindsets and processes of practitioners producing VLEs may impact design and thus, student outcomes.

Attempting to partly rectify this gap by interviewing producers of VLEs, a second, somewhat simple, but noteworthy insight was made: that it is important to not treat the practitioner community as a homogenous block. There are professors who create content; there are learning designers mediating the construction of the VLEs; and there are program managers charged with recruiting students and making the program sustainable, amongst others. These different subgroups of practitioners bring significantly different work and educational backgrounds, differing definitions of the ideal end user,
and different pedagogic paradigms to their design and production processes. These differences contribute to visions and goals for the product that are not always in alignment.

Some practitioners, for example, might take student self-regulation as a pre-requisite for successful completion of courses in a VLE; this can yield design choices less concerned with trying to equip students with study habits and time management strategies. There is some evidence that highly self-regulated users are more likely to complete courses, more likely to be older, and more likely to have a graduate degree (Kizilcec et al., 2017). These design orientations play significant roles in the production processes of these courses, and will have impact on the subsequent outcomes for heterogenous populations of learners. The learning analytics and research communities should take such differences and the resulting design dynamics into account when analyzing whether VLE designs promote educational equity.

Third, there is a noticeable variance among the practitioners’ access and utilization of theory and academic research as a guide to their work. Some practitioners have a background in critical theory and disability studies, along with other theories from their post-graduate studies, and bring these to bear as theoretical lenses to their work. Others rely on a more quantitative, behaviorist view. The academic literature and discourse about open scale courses and MOOCs is often not visible or accessible to these practitioners. This represents a challenge and opportunity for the learning analytics research community to consider how to better disseminate their findings to a practitioner audience.

Determining how to better disseminate research insights in a constructive and actionable way to the practitioner community would be a worthy goal for learning analytics research community moving forward. Additionally, it seems that the learning analytics research community might consider some of the perspectives of practitioners themselves and create a more reciprocal work arrangement. The critical theory and disabilities studies referenced by practitioners might help guide learning analytics researchers to more thoughtfully sub-group and disaggregate data in order to pay more attention to groups who might be marginalized. This approach might help ensure that specific learning needs of certain populations of users are not obscured by the generalized and averaged insights produced by big data.

There are a number of vexing challenges to bridging the divide between research and practice (Prieto et al., 2018) that are beyond the scope of this paper. We do note, however, that the divide cuts both ways: the learning analytics and research community has much to offer the practitioner community in terms of specific insights and observations regarding student behavior derived from data, and the practitioner community has much to offer in terms of knowledge of learning theory, technology development, and differentiated teaching strategies for sub-groups of learners, among other insights. These insights should influence and build off of each other, hopefully resulting in a more informed, deliberate, careful, and, ultimately, more fair and equitable, construction of courses for learners.
4 CONCLUSION

MOOCs and other open-scale VLEs were intended to broaden access to high quality post-secondary education (Agarwal, 2013). Research has shown that, instead, most users are from more prepared, lower need backgrounds (Rohs and Ganz, 2015; van de Oudeweetering and Agirdag, 2018).

Diagnosing the sources of this dissonance is of paramount importance, especially as MOOCs and open scale courses approach an inflection point. Some members of the learning analytics and research community observe an imminent shift in strategy, in part resultant from the failure to make MOOCs and open scale courses more fair and equitable. A recent article in *Science* summarized the past few years of research on these VLEs, noting that the courses “disproportionately drew their learners from affluent countries and neighborhoods, and markers of socioeconomic status were correlated with greater persistence and certification,” (Reich and Ruipérez-Valiente, 2019). The researchers assert that universities may be doubling-down on this model: after hoping to reorient higher education toward providing access to a broadly defined conception of traditionally underserved learners, “we see the field instead coalescing around a different, much older business model: helping universities outsource their online master’s degrees for professionals,” (Reich and Ruipérez-Valiente, 2019).

The research and practitioner communities may have inadvertently played a role in accelerating this shift. Our learning analytics ecosystem model hypothesizes that, despite the good intentions and noble efforts of researchers and practitioners, certain biases have unintentionally made the challenge of serving less prepared, higher needs learners more difficult. Early-adopter Iteration Bias may skew learning analytics and research toward recommendations that optimize course design for more prepared, lower need learners. Research-praxis Bias prevents the broader VLE producing community from fully utilizing the insights derived from learning analytics and research properly. We should note that, while closing the chasm between research and practice could greatly improve the design of MOOCs and open scale courses, this in itself would be insufficient; the insights derived from learning analytics and research may already be skewed as a result of Early-adopter Iteration Bias. Seeking to resolve these challenges requires a simultaneous approach.

We invite members of the learning analytics, research, and practitioner communities to reflect on this learning analytics ecosystem model and its implications with us, in the hope that doing so might help identify strategies to rectify these biases and the fairness and equity problems they may be exacerbating.

REFERENCES


STRATEGIC OMISSION AND RISK AVERSION: A BIAS-RELIABILITY TRADEOFF

David Lang
Stanford University
david.nathan.lang@stanford.edu

ABSTRACT: Whether high-stakes exams such as the SAT or College Board AP exams should penalize incorrect answers is a controversial question. In this paper, we document that penalty functions can have differential effects depending on a student’s risk tolerance. Moreover, literature shows that risk aversion tends to vary along other areas of concern such as race, gender, nationality, and socioeconomic status. In this article, we simulate Item Response Theory (IRT) data with and without a wrong answer penalty. In the presence of mild risk aversion, we find that students omit 12% more items than risk neutral individuals with identical ability. This translates into a nearly 2% difference in sum scores between the risk neutral and risk averse groups. We also find that penalty functions result in noisier estimates of student ability. These findings suggest that random guessing penalties should not be used in most circumstances, particularly for learning platforms.

Keywords: learning analytics, item response theory, risk aversion, differential item function, differential test function, simulation

1 MOTIVATION

In the past decade there have been notable shifts in the decision to penalize wrong answers in high-stakes testing. In 2010, the College Board removed its wrong answer penalty for the AP exams. The SAT has also removed this penalty from its exams in recent years.

In this paper, we explore whether learning platforms should follow suit. Many platforms implicitly or explicitly penalize guessing through either gamification mechanisms such as point systems or through hint generation. These designs often are associated with increased user engagement or performance but they may have downstream impacts on certain types of users (O’Rourke, Haimovitz, & Ballweber, 2014). Simulation may help us understand how these design features influence student behavior.

2 LITERATURE REVIEW

While much of the literature surrounding high-stakes testing has focused on bias in terms of gender and race/ethnicity, relatively little focus has been put forth into the effects of how random guessing penalties may mediate this bias. Past work points out that most exams with a penalty function are still designed so that a person who tries to maximize their average score will be indifferent to always guessing (Budescu & Bar-Hillel, 1993). Moreover, they point out that this penalty function introduces systematic biases for students. If students have a different objective (e.g. get a passing grade or get the top grade in the class), then these incentives may not hold. Other work found that there were substantial differences by gender in willingness to guess in the face of a penalty function (Baldiga, 2013). To date, there has been even less focus on how risk aversion affects the psychometric properties of these assessments.
2.1 Risk Aversion

There are three broad classifications of risk tolerance: risk-aversion, risk-prefering, and risk-neutrality. To understand these distinctions, consider a coin-flip bet where a person wins a dollar if the coin lands heads and loses a dollar if the coin lands tails. A risk averse person will never take a bet with an average payoff of zero. A risk-prefering person will always take this bet. The risk neutral person will be indifferent between taking this bet and not taking this bet.

In this paper, we model risk aversion using an exponential utility function:

\[ U(\text{points}, \text{risk}_{tolerance}) = \frac{1 - e^{-\text{points} \cdot \text{risk}_{tolerance}}}{\text{risk}_{tolerance}} \]

The components of the function are points (the number of points awarded or lost) and risk tolerance. Positive risk-tolerance parameters correspond to risk-aversion. Negative risk-tolerance parameters correspond to a risk-prefering behavior. In a testing framework, if the utility of attempting a question is positive, the examinee will attempt it. Otherwise, the examinee will omit it. This function exhibits several useful properties. First, it exhibits a constant coefficient of relative risk aversion. In decision analysis literature, this property is also known as the ‘delta property’ (Kirkwood, 1997). This property assures that an individual will have the same preferences regardless of their current wealth endowment. In a testing framework, this means that an individual’s decision to omit a particular item will not depend on one’s current score. This assumption is fairly reasonable for small scale decisions, such as one question on a forty-question exam. Additional benefits of this assumption are that it eliminates concerns with respect to item ordering effects interacting with risk aversion, and unlike other potential utility functions, this function can be transformed into a risk-averse/risk-prefering function simply by assigning a positive/negative risk tolerance value.

In terms of understanding what risk aversion looks like in the real world, most estimates suggest that individuals have positive risk tolerance and that a risk tolerance parameter of one is not unreasonable (Gandelman & Hernández-Murillo, 2014). Figure 1 shows that point estimates of risk aversion in the United States is around 1.5. The most extreme countries are the Netherland with a risk tolerance of less than a quarter and Taiwan with a risk tolerance of nearly 2.5.

3 MODEL

To assess the question of omission on exams, we simulate a forty-question exam. The exam data is modeled as Rasch data such that each individual’s true ability estimate is known to us. The probability that a student will answer an item correct can be expressed by the following formula where \( \theta_i \) corresponds to the ability of student \( i \) and \( b_j \) corresponds to the difficulty of item \( j \):

\[ \frac{1}{1 + e^{(\theta_i - b_j)}} \]
We further assume that students are aware of their ability and item difficulty but are uncertain whether or not they get the specific item correct. We also assume that they are aware of a one-quarter point penalty if they answer a question incorrectly. In this case, the students will respond to an item only if the expression below holds:

\[ P_t(\text{Correct} | \theta_i, b_j) \cdot U(1, \text{risk}_{\text{tolerance}}) + (1 - P_t(\text{Correct} | \theta_i, b_j)) \cdot U\left(-\frac{1}{4}, \text{risk}_{\text{tolerance}}\right) \geq 0 \]

We then re-estimate a person’s ability based on their responses under three separate scenarios: (1) no penalty, (2) risk-neutrality, (3) risk-aversion with a risk tolerance of 1. We then repeatedly estimate the difference between these three groups and our true ability measures to assess whether or not this biases estimates of test performances. The underlying data generation process assumes both ability and item difficulty follow the standard normal distribution.

Figure 2 illustrates the utility of responding to a question in which the student is aware of the probability they will get the question right. The horizontal line at zero identifies the locations at which students of varying risk tolerances will be indifferent to answering the question and omitting their response. Points above the zero line correspond to attempting the item. Points below the line correspond to omitting the item. The dashed-line corresponds to a risk neutral student. For risk-preferring students, students with a risk preference of three will “guess” if their probability of getting the question right is at least 3%. The most risk averse student would not respond unless they had at least a 55% chance of getting the question correct.
4 SIMULATIONS

A hundred bootstrapped simulations were run to better estimate the effects of strategic omission. Repeated simulations yields the omission rates plots below. On average, a risk-neutral simulation yields an omissions rate of 18%. In the risk-averse case, this omission rate jumps up to approximately 30%. Sum scores change relatively little with only a two percentage point difference in exam performance (Figure 3).

Figure 2 Indifference Probabilities and Utility

Figure 3 Bootstrapped Estimates of Sum Scores
4.1 Ability Measurement Error

By introducing a penalty, it introduces a large region where low ability individuals will not attempt certain items. This makes distinguishing between low ability people and very low ability people extremely difficult. From a maximum likelihood estimation perspective, this means that for each item there is a portion of the information curve where the estimate is completely flat. An illustration of that fact can be seen in Figure 4.

![Figure 4 Probability of Answering an Item Correctly as a Function of Ability and Risk Aversion](image)

We also recover individual ability estimates using a Rasch model and maximum likelihood. Estimates of these data yield unbiased estimates of an individual’s ability (See Figure 3). The mean absolute deviations of theta increases as the penalty function is introduced and as the risk-aversion increases. As such, the amount of error in ability measurements is nearly twice as large for a risk-averse population than if there were no penalties enacted on the same population of students (See Figure 5).
4.2 Reliability

So the fundamental question is why are these penalty functions used if it increases non-response rates and seems to introduce these potential claims of bias. One possible explanation is that improves measures of reliability. We compute the reliability of the generated exams using Cronbach’s alpha (Cronbach, 1951). The boxplots below show that reliability increases if students are given an incentive to omit incorrect answers. This effect still holds even if one assumes heterogeneity of risk tolerance amongst users (See Figure 3). In effect, what happens is that users who have relatively low likelihood of getting an item correct through random guessing gets their answer compressed to zero in response to a penalty. This omission, in turn, increases the reliability of an exam.
5 DISCUSSION

From a reliability perspective, penalizing exams has some benefits. Introducing penalties tends to increase the reliability of the exams. This increase in reliability comes at the cost of certain measures becoming noisy. Further, if there’s heterogeneity of risk aversion, it’s possible that the rank ordering of students could jump noticeably when an exam switches from a penalty function to an exam without a penalty function. Strategic omission makes generating distinctions between the bottom-half of the distribution very difficult. To the extent that an exam is concerned with generating a precise estimate of ability, utilizing a penalty function is ill-advised.

The only cases where a guessing penalty could make sense are when risk tolerance is a parameter that is also being trained. For instance this type of penalty function could be useful when training actuaries, financial investors, or stockbrokers. The rationale for this is that their score would be both a composition of their true ability and their risk tolerance.

5.1 Implications for Learning Analytics and Platform Design

This works suggests that penalties should not be used for assessment purposes. If individuals are penalized for wrong answers, then risk-averse users will strategically omit more responses than risk-tolerant users. In turn, this means that learning platforms would direct risk-averse users into more remedial content than similar ability students who are risk-neutral. To the extent that these populations are underserved groups (females, underrepresented minorities, and low socioeconomic status), embedding penalties for random guessing could deter these groups from interacting with the platform and replicate existing inequalities. Further, our simulations suggest that guessing penalties may make it more difficult for learning platforms to distinguish between users in the lower end of the ability distribution. These are often the groups that are of focal interest to learning analytics researchers and policy makers.

Many learning platforms reward users with points or badges for engaging with the platform and penalize users for using built-in hint generation features. Removing penalties from these contexts seem like a natural decision. Generally, these penalties should be removed when items are being used as part of a formative assessment.

If random guessing penalties are to be used in a summative assessment, there are approaches that mitigate the performance bias between risk-averse and risk neutral users. One of the design choices is to allow students to respond to multiple items before submitting a response for grading. This will allow rational agents to hedge their responses and makes risk-averse users more likely to respond so long as their knowledge is truly better than random guessing.

6 ACKNOWLEDGEMENTS

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant (#R305B140009). We also acknowledge thoughtful reviews from the conference organizers.
REFERENCES


