



# Disparities in Students' Propensity to Consent to Learning Analytics

Warren Li<sup>1</sup> · Kaiwen Sun<sup>1</sup> · Florian Schaub<sup>1</sup> · Christopher Brooks<sup>1</sup>

Accepted: 23 April 2021 / Published online: 3 June 2021  
© The Author(s) 2021

## Abstract

Use of university students' educational data for learning analytics has spurred a debate about whether and how to provide students with agency regarding data collection and use. A concern is that students opting out of learning analytics may skew predictive models, in particular if certain student populations disproportionately opt out and biases are unintentionally introduced into predictive models. We investigated university students' propensity to consent to learning analytics through an email prompt, and collected respondents' perceived benefits and privacy concerns regarding learning analytics in a subsequent online survey. In particular, we studied whether and why students' consent propensity differs among student subpopulations by sending our email prompt to a sample of 4,000 students at our institution stratified by ethnicity and gender. 272 students interacted with the email, of which 119 also completed the survey. We identified that institutional trust, concerns with the amount of data collection versus perceived benefits, and comfort with instructors' data use for learning engagement were key determinants in students' decision to participate in learning analytics. We find that students identifying ethnically as Black were significantly less likely to respond and self-reported lower levels of institutional trust. Female students reported concerns with data collection but were also more comfortable with use of their data by instructors for learning engagement purposes. Students' comments corroborate these findings and suggest that agency alone is insufficient; institutional leaders and instructors also play a large role in alleviating the issue of bias.

**Keywords** Privacy · Consent · Learning analytics · Educational equity · Institutional trust

---

This article belongs to the Topical Collection: *The FATE of AIED*  
Guest Editors: Kaška Porayska-Pomsta, Beverly Woolf, Wayne Holmes and Ken Holstein

✉ Warren Li  
liwarren@umich.edu

Extended author information available on the last page of the article.

## Introduction

In recent years, data collection in educational settings has been increasing. While this is partially attributed to institutional audit culture and its practices of benchmarking and formalizing accountability (Shore and Wright 2003; Shore 2008), the rise of technology use in classrooms (Tondeur et al. 2017; Long et al. 2017) coupled with advances in artificial intelligence and machine learning algorithms, which heavily rely on large quantities of data, has accelerated this trend. One purpose for the collection and use of this data is to create predictive models of learners, the targets of which range from academic performance to affect and engagement in class (Gardner and Brooks 2018). An application of these models include early warning systems (Macfadyen and Dawson 2010), which are used to alert advisors, instructors, administrators, or students themselves if a student appears to be struggling so that they can be supported before they fall significantly behind (Alhadad et al. 2015).

However, these systems often rely upon the collection of sensitive data such as demographics, grades, and interaction traces with online content (Pardo and Siemens 2014) that students are uncomfortable sharing for learning analytics (Ifenthaler and Schumacher 2016) depending on the stakeholder involved. For instance, third parties, such as Learning Management System (LMS) vendors, have also turned to developing early warning systems and products that rely on educational data, even though such data sharing arrangements may be unclear to students (Polonetsky and Jerome 2014). The manner by which data collection is conducted thereby creates a tension between institutional goals of using predictive models to support students' educational progress and retention, instructor goals of course-specific performance monitoring, and upholding commitments to learners' consent, agency and privacy (Pardo and Siemens 2014; Prinsloo and Slade 2014b).

There have been numerous calls to provide students with more agency regarding how data is used in learning analytics (Pardo and Siemens 2014; Drachsler and Greller 2016). Yet, students' privacy concerns may deter them from consenting to the use of the data in learning analytics. Moreover, biases have been shown to exist in predictive models, partly due to non-representative samples acquired during data collection (Ocuppaugh et al. 2014). As the availability of data is restricted, machine-learned models may have a reduction in accuracy which can lead to less effective interventions for some (or all) students (Li et al. 2019). This is particularly concerning since demographic gaps already exist in educational achievement (Bainbridge and Lasley 2002), and is especially true for underrepresented minorities (Bensimon 2005), those with a lower socioeconomic status (Duncan and Magnuson 2005), and between genders in certain contexts such as STEM programs (Matz et al. 2017). Not only are there outcome discrepancies, but it has also been shown that different demographics-based communities have different expectations of privacy and concerns when it comes to how their data is to be used (Cho et al. 2009). If students in the minority groups or with a particular background are more reluctant to share data, their data will be absent, which may end up biasing models in ways that are not representative of all students.

In this study, we investigate students' propensity to consent to or opt out of having their data collected and used for learning analytics. We further connect consent propensity to students' demographics, personality characteristics, privacy perceptions, as well as students' perspectives and concerns regarding learning analytics in order to understand the factors motivating students' expressed consent preferences. Linking participants' responses to demographic characteristics enables us to analyze the differences between student subpopulations and how those might translate into differential consent rates. The research questions we address are as follows:

- [RQ1:] What are students' perspectives on their educational data being used in learning analytics?
- [RQ2:] What are the population and participation characteristics of students who indicate a preference to allow or disallow their educational data to be used for learning analytics?

In "**Methods**", we describe our study to answer these questions by first ascertaining students' propensity to consent or deny use of their educational data for learning analytics with an email-based, one-question preference elicitation prompt. Respondents were subsequently invited to complete an online survey that investigated the factors behind their consent indication in order to identify key determinants. The email prompt and online survey responses were then associated with students' institutional demographic data in order to contextualize the relationship between students' demographic characteristics and their propensity to participate in learning analytics. We sent our email prompt to a sample of 4,000 students at our institution stratified by ethnicity and gender; 272 students responded to the email prompt, of whom 119 further completed the survey.

In "**Findings**", we found differences in response rate to the email prompt among genders and ethnicities. Female students were much more likely to respond than male students and, despite stratified recruitment, responses from White students were overrepresented while responses from Black students were underrepresented; there were no differences in consent behavior between genders nor ethnicities. Among respondents, we identified three important factors which play a role in students' consent expressions regarding learning analytics: student's trust in the educational institution, a student's level of concern regarding individual data collection, and a student's comfort with an instructor's use of data for improving student engagement. Certain privacy attitudes are correlated with population subgroups, most notably students' identifying as Black generally express less trust in the institution, and female students tend to have greater apprehension about personal data collection while simultaneously being comfortable with instructor use of such data to improve student engagement.

Our findings suggest that instructors may have an important role in making students feel at ease when it comes to data sharing. We discuss in "**Discussion**" how this comfort may be bolstered by being more transparent regarding who data is used and who has access to it, thereby balancing broader institutional interests of effectively educating students while maintaining individual privacy safeguards and student

agency. We also discuss limitations of the current study and routes to deepen our understanding of the rationale behind students' consent decisions.

## Background and Related Work

We discuss prior work on privacy and ethical concerns regarding learning analytics, equity and disparities in education, and sociocultural orientations in education.

### Privacy and Ethical Issues in Learning Analytics

Learning analytics relies on the collection and use of student data that may include sensitive information and confidential records, which raises privacy concerns (Drachler and Greller 2016; Ifenthaler and Schumacher 2016; Reidenberg and Schaub 2018). Meanwhile, broader changes in society emphasizing individuals' rights in data processing are reflected in new privacy regulations such as Europe's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) and California Privacy Rights Act (CPRA). In this light, discussions around the implications of collecting, using, and analyzing student data in educational contexts are becoming more critical (Prinsloo and Slade 2017; Niall 2017). Existing research has pointed out emerging privacy and ethical issues around learning analytics, including student consent and agency over student data, and their trust in learning analytics systems (Pardo and Siemens 2014; Drachler and Greller 2016; Rubel and Jones 2016).

Student consent is critical not only to demonstrate respect for them and their decisions, but also to support important values such as autonomy and freedom of choices (Sedenberg and Hoffmann 2016). Also, considering student consent is to acknowledge students' rights and voluntary collaboration to allow the collection and use of student data by learning analytics in supporting student learning (Slade and Prinsloo 2013). It is an ethical approach when institutions include codes of conduct that guide informed consent, data collection purposes, and transparency of data use to minimize potential harm and allegations of misuse (Land and Bayne 2005; Slade and Prinsloo 2013).

Prior studies have found several sociodemographic characteristics contributing to disparities among demographic groups when it comes to their consent to participation in research such as age (Jacobsen et al. 2004; Benfante et al. 1989), gender (Ramos et al. 2004; Pirzada et al. 2004), socioeconomic status (Boshuizen et al. 2006; Gordon et al. 1959), and ethnicity (Moorman et al. 1999, 2004). Li et al. (2019) found that student consent or opt-out decisions can affect the predictive power of learning analytics models for different student subpopulations. In our study, we quantified students' participation and consent rates for learning analytics by demographic groups, which is important for contextualizing the differential effects identified by Li et al. We further investigate the underlying reasons as to why students choose to consent or opt out of learning analytics, and how these factors are linked to demographic characteristics and personality traits.

Meanwhile, consent is closely related to autonomy and agency (Alexander 1996). Student agency is characterized as students being able to hold themselves accountable to make decisions in learning processes, which is critical to students' learning engagement and pursuit of learning goals (Deakin Crick and Goldspink 2014; Seifert 2004). To increase student agency and empowerment to participate in learning analytics, students should be viewed as collaborators in learning analytics rather than data producers or service receivers (Buchanan 2011; Kruse and Pongsajapan 2012), and Sun et al. (2019) found that students demand more agency in decisions regarding how data about them is used.

On the other hand, current consent practices face challenges and critiques as consent is often perceived as an operational act rather than being understood and assented with moral legitimacy (Barocas and Nissenbaum 2014). Barocas and Nissenbaum (2009) identified that in the online behavioral advertising context, consent often neither sufficiently capture users' agreement to tracking and targeting nor conveys meaningful notice that could facilitate users' choices due to the disconnection of privacy policies from different parties (e.g., data publishers, contracted third parties), the changing nature of privacy policies, and the lack of data flow transparency to users. As a result, accompanied with the asymmetrical power relationships with companies, people could feel powerless toward the inevitable privacy violations, a social phenomenon described by Draper and Turow (2019) as digital resignation. Obfuscatory consent practice may confuse people and discourage them from demanding agency (Draper and Turow 2019; Ellison and Ellison 2009).

Furthermore, students' trust in learning analytics systems plays a critical role in supporting an educational ecosystem that maximizes the experiences of different stakeholders such as learners and educators (Drachler and Greller 2016), creates reciprocal relationships between the institution and the students to encourage students to share their data for learning benefits (Slade et al. 2019), and facilitates the establishment of reliable analytics systems (Petersen 2012). Prior work has also found several factors positively influencing students' trust in learning analytics, such as protecting data to avoid unauthorized access or distribution, proper storing of historic data, data de-identification, valuing student privacy, achieving consensus on data collection purposes, and transparency of data collection (Pardo and Siemens 2014; Clarke and Nelson 2013; Drachler and Greller 2016; Slade and Prinsloo 2013; Beattie et al. 2014). Recent work shows that students inherently trust and expect their institution to properly and ethically use student data (Slade et al. 2019). Our study explores whether students' trust in the institution affect students' propensity to consent to learning analytics.

## Equity and Disparities in Education

Equity has long been a fundamental concept in education. Simon et al. (2007) describe equity in education as twofold: fairness and inclusion. Equity as fairness illustrates that an individual's socioeconomic status should not affect their chances to pursue education. Equity as inclusion acknowledges the basic need to complete compulsory education in order to acquire the skills needed in society.

Racial equality remains a controversial issue in education due to disparities in academic outcomes and limited access to opportunities and resources for students of color (Noguera 2016). Students from Hispanic/Latinx, African American, American Indian, and Pacific Islander groups are underrepresented at all levels of higher education from undergraduate majors to graduate program pursuits, particularly in STEM-related fields (Hanson 2008; Cook and Córdova 2007).

Even with successful graduation from higher education, individuals from minority groups are less likely to consider pursuing research careers (DePass and Chubin 2008).

In the late-90s, Ladson-Billings (1998) stated that “*the intersection of race and property [is] a central construct in understanding a critical race theoretical approach to education*”. While the fundamental belief in critical race theory (CRT) is to “*recognize the experiential knowledge of people of color*” (Matsuda 2018), education scholars hold the same belief and have recognized more aspects in CRT that align consistently with education equity goals (Ladson-Billings 1998; Dixson and Rousseau 2005) such as, the inherency of CRT to historical and contextual analysis, its challenges to mainstream neutrality, objectivity, color-blindness, and merit, as well as its values on the opinions of people of color (Crenshaw et al. 1995; Matsuda 2018).

Gender equity in education has also been a subject of national debate as shown when the American Association of University Women (AAUW) published *The AAUW Report: How Schools Shortchange Girls* (Bailey et al. 1992), which marks a series of efforts to support gender equity through the introduction of topics such as race and gender on campus and girls in science and technology (Corbett et al. 2008). Gender disparities have been shown to play a role in students’ academic performances, school experiences, education outcome, and barriers while achieving their educational goals (Buchmann et al. 2008; McWhirter 1997; Grossman and Grossman 1994).

As learning analytics aims to support teaching and learning for all students (Diaz and Brown 2012), discussions arise around the fair use of learning analytics (Prinsloo and Slade 2014a; Roberts et al. 2017) and predictive models (Dwork et al. 2018; Friedler et al. 2019; Liu et al. 2018; Gardner et al. 2019) due to the potential biases and lack of impartiality in such algorithms (Cofone 2018; Richardson et al. 2019); this could lead to inaccurate modeling for populations that are not well represented (Li et al. 2019; Ocumpaugh et al. 2014). When coupled with the fact that minorities are already less likely to consent in numerous contexts as we described in “*Privacy and Ethical Issues in Learning Analytics*”, it becomes crucial to understand how characteristics such as gender and race affect students’ consent propensity in the context of learning analytics to avoid developing models that inadvertently widen disparities.

## Sociocultural Orientations in Education

Learning science research has established that students’ academic performance is related to factors such as their personality traits (Zhou 2015), cultural background (Niles 1995), and competitiveness and cooperativeness (Baumann and Harvey

2018). An individual's competitiveness and cooperativeness is part of their social interdependence orientation (Johnson et al. 1998; Johnson and Norem-Hebeisen 1979), and such characteristics are associated with one's gender, cognitive and social development (e.g., perception and response in group settings), attitudes toward the educational institution and relevant people in that environment (e.g., other students and teachers), and perspective-taking ability (Madsen 1967; Johnson and Engelhard 1992). More specifically, positive social interdependence is (cooperation) established when individuals from a group share common goals and their collective actions affect the group outcomes (Johnson and Johnson 1991; Deutsch 1949). In other words, people cooperate when they realize that they would not accomplish the goal without everyone working towards it (Johnson et al. 1998; Johnson and Johnson 1991). Relatedly, people's gender (Ramos et al. 2004; Pirzada et al. 2004) and perceived contributions of their consent to research benefits (Kim et al. 2017) (as an example of perspective-taking) have been shown to affect consent. We therefore explore whether students' competitiveness/cooperativeness, as a representation of their various underlying cognitive and social developments, would be a factor influencing their willingness to consent, as well as their perspectives on data collection and use.

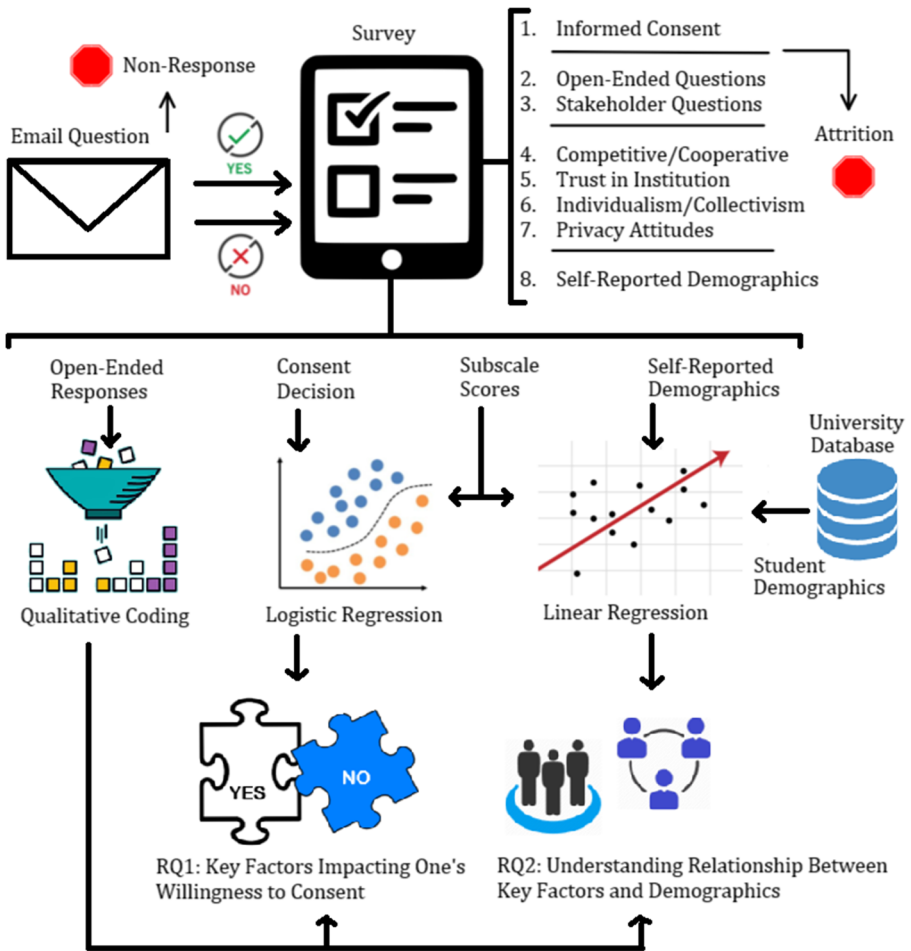
Furthermore, the culture aspect of social orientations can reflect an individual's decision-making considerations and motivation to succeed (Johnson and Engelhard 1992; Triandis 2018). Among different dimensions of culture measurements, individualism-collectivism (IND-COL) has been the most studied (Hofstede 1984; Cozma 2011). IND-COL can be a key characteristic of an individual's racial identity (Nobles 2006), influences how people prioritize personal goals versus group goals (Schwartz 1990; Yamaguchi 1994), and can be used as a framework to analyze if one feels connected to and responsible for the group they belong to (e.g., students' perception of their roles and responsibilities as students) (Taylor and Moghaddam 1994; Triandis et al. 1988). Carson (2009) also identified that collectivism is reflected in students' belief of education purpose and the way they evaluate academic success. As we discussed in "[Privacy and Ethical Issues in Learning Analytics](#)", consent is closely related to one's agency and student agency relies on students being considered collaborators in learning analytics. Thus, we investigate if there is a relationship between students' sense of responsibility to contribute their data (as a form of collectivism) and their consent practices.

## Methods

Our study investigated two primary research questions: (RQ1) What are students perspectives on their educational data being used by learning analytics systems in the form of predictive models? and (RQ2) What are the population characteristics of students who indicate they would consent or opt-out of participating in such uses? In order to investigate these questions, we distributed an email-based preference elicitation prompt to students asking them whether or not they would hypothetically agree to have their data used in learning analytics systems. Upon selecting either *yes* or *no* to indicate their consent preference, students were redirected to an online survey

that asked about the student’s rationale behind their consent indication and perspectives regarding their data being used for learning analytics in different contexts and by different stakeholders. We further elicited relevant personality characteristics and attitudes that might impact students’ propensity to consent. Responses were then linked with institutional demographic data to identify correlations with consent. This study design is summarized in Fig. 1.

The study team comes from an interdisciplinary background and has a variety of experiences with student data. Dr. Brooks, for instance, has been a part of the institutional stewardship chain for student data related to learning technologies, which is adjacent to the data we collected. In addition, Dr. Schaub has been involved in



**Fig. 1** Our study consisted of an email prompt sent to students that included links to the online survey, which consisted of multiple components. Email and survey responses were linked to institutional student records. The analysis methods are also shown with the corresponding data needed for each approach in order to address our research questions



institutional processes related to privacy and learning analytics, and the whole study team has been involved in student modeling and educational data science research at the institution including qualitative and quantitative approaches in the past. Next, we explain each part of the study design in greater detail. Our study has been approved by our Institutional Review Board.

### Measuring Privacy Perceptions, Personal Traits, and Decision to Consent

We sent a one-question email prompt to a stratified student sample to understand student's consent decision regarding data being used by learning analytics. Li et al. (2019) found that the use of either “opt-out” or “opt-in” wording leads to different response rates from participants. Thus, we prepared two variants of the email prompt, shown in Fig. 2.

Students only saw one framing and we conducted pilot testing to ensure that the wording did not lead to confusion. Once a student clicked either the yes or no link, the response was logged with an identifier to link it with their corresponding institutional demographic records. Identifiers were subsequently discarded before analysis. Regardless of response, respondents were then directed to a debrief that explained the purpose of the study, an informed consent form, and an invitation to participate in an optional online survey. Participants who completed the online survey were compensated \$5.

The email prompt allowed us to ascertain propensities for students' consent to learning analytics data use. Our online survey further explored why such decisions were made. Note that we intentionally used a broad consent message in order to study the factors and pre-conceived notions about learning analytics that influence students' consent decision. We are not advocating for this prompt as an exemplar for broadly soliciting data consent decisions on live systems.

For the survey questions (see Appendix A for the full survey instrument), we iteratively refined the wording to minimize misinterpretation, and pilot-tested the

#### Research study on your educational data and learning analytics

We are U-M researchers conducting a study to investigate student perceptions regarding the use of students' educational data in learning analytics systems.

Universities, including the University of Michigan, are using learning analytics systems that measure, collect, analyze, and report data of learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs. Learning analytics systems use learners' data such as your grades, how you use online learning tools, your course registration information, and other information about you to achieve different goals for different university stakeholders.

If you were given a choice to opt out of having your student data used in learning analytics systems, would you? (Please click to indicate your choice.)

[Yes](#), I would opt out.

[No](#), I would not opt out.

#### Research study on your educational data and learning analytics

We are U-M researchers conducting a study to investigate student perceptions regarding the use of students' educational data in learning analytics systems.

Universities, including the University of Michigan, are using learning analytics systems that measure, collect, analyze, and report data of learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs. Learning analytics systems use learners' data such as your grades, how you use online learning tools, your course registration information, and other information about you to achieve different goals for different university stakeholders.

If you were given a choice to opt into having your student data used in such learning analytics, would you? (Please click to indicate your choice.)

[Yes](#), I would opt in.

[No](#), I would not opt in.

**Fig. 2** The email for the opt out condition is on the left and the email for the opt in condition is on the right. They are identical except for the last paragraph and consent options. Each student was sent only one of these two versions

questions with a group of about 10 undergraduate and graduate students working on privacy-related and educational technology research. While the survey contains multiple scales, most questions were Likert scale items that did not require significant cognitive load to process. We also provided fair compensation based on the average completion time of 15 min, which we do not consider to be excessive, though it is presumable that some participants exited due to length. As shown in Table 1, of the 272 people who clicked one of the options in the email prompt, 150 actually consented and started the survey, of whom 116 completed the survey, i.e., the survey completion rate is 43%.

At the beginning of the survey, participants were asked in three open-response questions to describe the important factors that affected their consent decision, perceived benefits of student data being used by learning analytics systems, and concerns with such data use. Next, we asked participants to rate their level of comfort on a seven-point scale with their educational data being used in five scenarios by different stakeholders for different purposes (e.g., “help instructors gain insights about students’ engagement”).

We further assessed students’ level of *competitiveness and cooperativeness* in the educational setting using the Social Interdependence Scale (Johnson and Norem-Hebeisen 1979) as such characteristics are associated with one’s attitudes toward the educational institution, the relevant people in that environment (e.g., other students, teachers), and perspective-taking ability (Madsen 1967; Johnson and Engelhard 1992). We aimed to explore if students’ competitiveness/cooperativeness as a representation of their various underlying attitudes would be a factor influencing their willingness to consent.

Given that students’ trust in the institution has been shown to be a fundamental factor influencing students’ learning experience (Van Maele et al. 2014), we wanted to understand whether students’ *institutional trust* might impact their consent propensity. We used Ghosh et al.’s trust scale (2001) that defines trust as students’ confidence in the institution’s ability to support students achieving learning and career goals.

Students in our institution come from diverse cultural backgrounds. We hypothesize that students’ sense of responsibility to contribute their data (as a form of

**Table 1** Response rates per condition

	Opt-in Condition <i>n</i> = 2000	Opt-out Condition <i>n</i> = 2000	Overall <i>n</i> = 4000
Emails Sent			
Link Clicks	129 (6.5%)	143 (7.2%)	272 (6.8%)
Consent	101 (78.3%)	96 (67.1%)	197 (72.4%)
Deny	28 (21.7%)	47 (32.9%)	75 (27.6%)

The first two rows shows the number of emails sent out for each condition, and the number of people that clicked on a link (i.e. made an actual consent decision). The last two rows shows the number (and percentage) of respondents who consented or denied use of their data for those who responded to either the “opt-in” or “opt-out” condition via email link. “Overall” is a sum of the first two columns (both wording conditions)

collectivism) could be a potential factor affecting their consent practice. Thus, we use the *Horizontal and Vertical Individualism and Collectivism* measurement scale (Triandis and Gelfand 1998) to evaluate horizontal individualism, vertical individualism, horizontal collectivism, and vertical collectivism. Since prior work has found students having privacy concerns regarding learning analytics (Pardo and Siemens 2014; Picciano 2012), we also included the *Internet User Information Privacy Concerns Scale* (IUIPC) (Malhotra et al. 2004).

Finally, we asked demographic questions, including gender, ethnicity, first-generation college student status, and year of study in order to understand key factors in the decision to consent with regards to demographic characteristics. While we had access to institutional demographic data for participants, which we also used in our analysis, this self-reported demographic information allowed students to self-identify gender including non-binary gender options. We further asked to specify their country of origin. However, because not all respondents to the email prompt completed the survey, we used institutional records for ethnicity and gender in our statistical analysis. For year of study, we use students' self-reported class standing.

### Recruitment & Participants

As the data is collected from a single institution, we briefly describe the University of Michigan (UM) to help contextualize the work for others seeking to apply our results. Demographic statistics are obtained from the most recent figures (University of Michigan AA 2020) published by the Office of Diversity, Equity, and Inclusion (DEI). The student body is skewed towards higher socioeconomic status; the gender composition is balanced with 50.6% identifying as man, 48.3% identifying as woman, and 1.1% as transgender or gender non-conforming. UM is a large four-year, primarily residential, majority undergraduate, full-time, more selective university with lower transfer-in rates and very high research activity (Carnegie Classification IHE 2017). Out of approximately 46,000 students, the mean age of students is 22.7 with 7.9% coming from backgrounds where neither parent or guardian has attended college. 75.0% of students were born in the US and the ethnic composition is as follows: 4.3% African-American or Black, 24.2% Asian-American or Asian, 6.3% Hispanic or Latinx, 1.7% Middle Eastern or North African, 0.1% Native American or Alaskan Native, 57.9% White, 1.0% Other; 4.6% specified one or more of the previous categories.

The institution has a history of advancing DEI and its stance is that DEI is key to individual flourishing, educational excellence, and the advancement of knowledge (University of Michigan AA n.d.). In 2017, the university established the Learning Analytics Guiding Principles (University of Michigan AA 2017) that define learning analytics, and set respect, transparency, accountability, empowerment, and continuous consideration as UM's core tenants regarding research in this field. The Center for Academic Innovation (University of Michigan AA 2020) also develops projects to extend academic excellence and provide sustainable solutions to advance learning, facilitate problem solving, foster equity and inclusivity, and increase access and affordability.

Students were recruited based on specific demographic characteristics in the institutional database containing students academic records and demographic details. For each email variant (opt-in versus opt-out wording conditions) we recruited 2,000 students with each student receiving only one version of the survey (total emails sent was 4,000). Each sample of 2,000 students was selected using a disproportionate sampling method in order to ensure a balanced data set. The population was first divided into 5 strata based upon the ethnic categories listed in the institutional data (White/Caucasian, Asian, Black/African, Hispanic/Latinx, Other, which included those who indicated two or more ethnicities, Hawaiians, and Native Americans). Each stratum was also balanced with respect to gender.<sup>1</sup> This meant that each ethnicity-gender group had  $n = 552$  participants with the exception of Black/African students ( $n = 342$  for both males and females) due to scarcity.

### Quantitative Analysis: Identifying Factors in Consent Decisions

We used a logistic regression to control for the various factors outlined and to identify which considerations are most important in students' decision to consent. For the scales used, we computed a composite score based on items within each of its subscales. This procedure compressed the number of survey scale features to 24. Note that the five comfort rating questions were used as-is (a discrete value in the set 1 thru 7, inclusive). Full details of this analysis method along with models are found in the accompanying computer code in <https://osf.io/sg4rk/>.

### Ensuring Data Quality and Correctness

To minimize errors due to low-quality answers, we checked survey responses for speeding and straightlining. Manual review of particularly fast and slow responses revealed no anomalies. Therefore, we choose to keep all responses.

We identified outliers and influential points by plotting Studentized residuals and Cook's distance for each observation, using an absolute value  $>2$  and  $4/(N - k - 1)$  as thresholds respectively, where  $N$  is the total number of observations and  $k$  is the number of explanatory variables. Studentized residuals are the residuals divided by estimates of the standard deviation, while Cook's distance summarizes the effect of removing an observation on the fitted response values. This resulted in 15 flagged points. Manual inspection made it evident that 2 people had accidentally selected a different option either by accident or due to a misunderstanding of the prompt. For instance, one explicitly stated that, "I misread the choices. As it said yes I assumed it meant to opt-in, not 'yes, I would opt out.' "; such answers were corrected. The remainder of the flagged items did not reveal any other evidently concerning issues.

---

<sup>1</sup>Institutional records only provided a male/female binary indicator, but we included the option to report non-binary identities in the survey. However, as only 2 students were non-binary, it is not possible to draw conclusions based on this sample size and so the quantitative analysis relies on the institutional data set; we have included an additional discussion of these 2 cases in our qualitative findings ([Findings](#)).

Removing all of these points results in quasi-complete separation and a large shift in the coefficients. Thus, it may be the case that those who denied use of their data were considered “unusual” solely because the overwhelming majority of students consented to data use for learning analytics; 15 of only 25 respondents who did not consent are in this list. We choose to retain these points in the model as they represent important perspectives to consider.

### Model Fit and Feature Selection

We fit a logit regression model using maximum likelihood estimation. The input variables includes the 24 survey features, one for each subscale. The binary outcome variable was whether a student consented or denied the use of their data. Emphasizing the fact that we are interested in understanding specific factors, a feature selection processes was used for pruning the list of inputs into a smaller subset. This helps ensure that the significance values used to make these determinations are reliable, that confidence intervals on regression coefficients are sufficiently narrow, and violations of the linearity assumptions are addressed.

We used the variance inflation factor (VIF) as a gauge for multicollinearity and note that a number of features had a VIF above 5, indicating a problematic amount of collinearity. This is not necessarily surprising given that it is plausible to expect that some the measured concepts will be correlated to each other, particularly since we constructed composite scores based on subscales of an overarching latent trait. We alleviated this by conducting feature selection with recursive feature elimination (RFE) with 20-fold cross validation, which removes features iteratively based on feature importance, as well as backwards elimination (BE) with a threshold set at  $p < 0.05$ , which removes features in accordance with the highest  $p$ -values. We then choose the features common to both pruned models with  $p < 0.05$ .

### Quantitative Analysis: Understanding Relationships Between Key Consent Factors and Demographics

All demographic data was one-hot encoded for categorical variables (year of study, gender, and ethnicity into  $N - 1$  dichotomous variables where  $N$  is the number of categories). We choose the category with the greatest population to exclude as a reference in the linear regression model. Therefore, an input was created for each demographic listed in Table 4 with the exception of “White”, “Sophomore (2nd Year)”, “Not First Generation”, and “Male”, which was our comparison group, resulting in a total of 8 one-hot encoded columns. The input variables are these 8 columns, while the target variable is the composite score for each of the  $N_f$  key factors identified using the feature selection process described in “[Model Fit and Feature Selection](#)”. This results in a total of  $N_f$  separate ordinary least squares models, one for each target variable.

Diagnostics were conducted for each of these models. There were no indications of collinearity. Plotting the residuals against fitted values did not suggest any egregious outliers, nonlinear behavior, or major concerns regarding heteroscedasticity, which may deflate  $p$ -values due to increased variance that is unaccounted for in the model.

Thus, we are reasonably confident in our coefficients and statistical conclusions, described in “[Findings](#)”.

## Qualitative Analysis

To analyze open responses to survey questions, we engaged in successive rounds of open coding, in which one researcher went through all responses and developed an initial codebook (Saldaña 2015), followed by iterative codebook refinement by two of the authors independently coding a subset of responses and then jointly reconciling disagreement. After four iterations, high inter-rater reliability (Cohen’s  $\kappa = .77$ ) was achieved. One researcher then used the final codebook to recode all responses. The final codebook consisted of 15 themes with 29 unique codes, see [Appendix B](#).

## Findings

Beginning with the quantitative analysis, we present statistics about response rates, show the key factors from the scale items in willingness to consent and how these are correlated with demographics according to our regression models. The qualitative analysis is then presented based on students’ answer to the open-ended survey questions, laying out self-reported factors, benefits, and concerns regarding students’ views on data sharing, organized by subpopulations of students making similar statements.

## Quantitative Analysis Findings

We break down our discussion of the quantitative analysis into three parts: statistics regarding participation rates during our initial email engagement with students, results regarding our logistic regression model used to identify primary factors underlying student’s decision to share data (RQ1), and results from the linear regression models, which explain demographic correlations with each identified factor of importance (RQ2). We find that there is both a gender and ethnicity gap between groups when it comes to response rate. The key factors identified behind the decisions of students who did respond were trust in the institution, level of general concern regarding individual data collection, and comfort with instructor use of data for classroom engagement. Institutional trust was generally higher for female students and lower for students who identify as Black, while data collection concerns and comfort with instructor data use were higher for females when compared to males.

## Response to Email Prompt

[Table 1](#) describes response rates split by email wording condition. Despite the low overall response rate of 6.8% we find that, generally speaking, most people (72.4%) consent to data usage when they do respond. We do not find any effect on the participation rates (link clicks) between the opt-in and opt-out conditions. While the consent rate is somewhat lower for the opt-out condition; a two-tailed test for proportions

shows that there is no statistical difference between these conditions ( $p = 0.39$ ) when only considering those who made a selection. Therefore, for the remaining analysis, we combine the opt-in and opt-out conditions and look only at the aggregate data, given that the differences are negligible. This confirms the result in Li et al. (2019), which states that wording has no effect on participation rate, but contrasts with their findings regarding consent rate where a difference was found between conditions.

We also decompose click rates to analyze participation by subpopulation, such as ethnicity and gender, see Table 2. There is a significant difference between the number of clicks, or engagement, between male (106) and female participants (166), despite gender-based stratification in recruitment. Given those who did respond however, the consent rates do not deviate from expectation. A Chi-Squared test indicates that gender is independent of the consent rate, but this is not the case for number of clicks ( $\chi^2 = 13.24$ ,  $p = 0.0003$ , Cramer's  $V = 0.22$ ); there is a moderate association.

A similar case holds for ethnicity: the consent engagement differs quite drastically between subpopulations, especially when compared with the expected number of link clicks. While Asian and Hispanic respondents' answers align with expectation (percent deviation of -4% and -1% respectively), there is a notable overrepresentation of responses from those identifying as White (by 31%), and an underrepresentation of answers from those identifying as Black (by -40%). Once again, we find that ethnicity is not independent from the click rate ( $\chi^2 = 14.32$ ,  $p = 0.002$ , Cramer's  $V = 0.13$ ) with a medium effect size, whereas there is no such relationship with consent.

Given the aforementioned discrepancy, we ran a more particular test to see if there are true differences between the proportion of those who click on an answer within these subpopulations. Namely, we divide the sample into those who identify as Black and non-Black (case 1), as well as those who identify as White and non-White (case 2). The sample statistic in case 1 is  $-0.03$ , with a 95% confidence interval (CI) of

**Table 2** The number of link clicks and the number of people who consented to data use

	Link Clicks (Actual)	Link Clicks (Expected)	Consent (Actual)	Consent (Expected)	Emails Sent
Male	106	136	81	77	2000
Female	166	136	116	120	2000
White	98	75	71	77	1108
Asian	72	75	57	52	1104
Hispanic	74	75	52	54	1104
Black	28	47	17	20	684

The expected number of link clicks and consent rates were obtained by taking the aggregate click and consent rates, scaled by the total number of emails sent out according to subpopulation size (by gender or ethnicity). As mentioned previously in "Methods", this was balanced for gender and, to the greatest extent possible, ethnicity as well. Some adjustments were necessary due to the smaller proportion of students who identify as black

$[-0.053, -0.012]$ , corresponding to  $p = 0.002$ . For case 2, the sample statistic is 0.028, 95% CI of  $[0.011, 0.046]$ , and  $p = 0.0015$ . Therefore, Black students participate less when compared to those those who are not Black, and White students participate more when compared to non-White students.

### Identifying Primary Factors in Participation

We explore the reasons behind the differential engagement by subgroup and address RQ1 by identifying the key factors that led to students' decision to consent or deny use of their data for those who did respond. We fit a logit model where the input variables are the factors impacting students' willingness to consent with the binary outcome variable being whether or not a student consented or denied consent of the use of their data. Since the goal is to identify critical factors, our focus is not to achieve the highest predictive accuracy, and we note that this logit model is not and should not be used to generate predictions of consent without further error analyses as minority subgroup classifications may be unreliable and skewed towards the majority class distribution for imbalanced datasets. We further note that the feature selection process described below are based on coefficient significance values that are decoupled from predictions or measures of fit.

Our final model was obtained by conducting feature selection using two techniques: recursive feature elimination (RFE) with 20-fold cross validation as well as backwards elimination (BE) with a threshold set at  $p < 0.05$  as described in “[Model Fit and Feature Selection](#)”. As BE yielded a feature subset of the RFE approach, we fit our model with the set from the most stringent standards and focus our discussion on the results of BE. Specifically, BE indicates a subset of 3 factors that have an effect on the response variable and we consider the following to be impactful in students' decision to consent: one's trust in the institution, concern in the amount of personal data collected, and comfort with instructor use of data for instructional purposes. Table 3 shows summary statistics for the final model.

The odds ratios for each of these key factors may be interpreted in terms of a percent increase in the likelihood to consent to data usage given a one-point change in each particular subscale. Since all of the items in these three subscales are based on a 7-point Likert scale, a one-point increase in institutional trust means that

**Table 3** Summary of our logit model with three factors: institutional trust, the “Collection” subscale from UIIPC, and the self-developed stakeholder question regarding comfort with instructor use of data

	Coeff.	Std. error	$z$	$P >  z $	[0.025	0.975]	Odds ratio
Institutional Trust	0.843	0.328	2.571	0.0100	0.200	1.486	2.324
UIIPC Collection	-1.526	0.436	-3.495	0.0004	-2.381	-0.670	0.217
Comfort-Instructor	1.1366	0.291	3.900	0.0001	0.565	1.708	3.116

Regression coefficients, standard error, z-score, p-value, and the lower and upper bound of a 95% confidence interval are given. In addition, we compute the odds ratio by exponentiating the coefficients



students' likelihood to consent increases by 132%. A one-point increase in comfort with instructor data use leads to a 212% jump, while concern regarding data collection by one point drops the chance of students consenting by 78%.

### Demographic Correlations with Key Factors in Consent

For each of the three key factors identified—institutional trust, data collection concern, and comfort with instructor data use—we then analyze correlations with demographic characteristics. First, we provide summary demographic statistics for those who completed the survey in Table 4. Note that similar to the email engagement findings, the discrepancies between demographics is also present in the survey completion rates: students who identified as White or students who identified as Female students are overrepresented while students who identified as Black are underrepresented.

To understand the potential reasons behind these differences we address RQ2 and identify correlations between demographics and influences impacting willingness to consent by running a linear regression model per key factor identified. The summary statistics for each model, where the outcome variables are the corresponding subscale scores, are tabulated in Table 5. For institutional trust, we find that identifying as female corresponds to higher levels of trust relative to males. The opposite is true for certain ethnicities: Black students are less trusting of the institution when compared to White students. Female students also correlate with having more data collection concerns, as measured by the IUIPC collection subscale, while having greater comfort with instructor use of data for course purposes. Finally, the collection score has an inverse relationship with those identifying as Asian (that is, students who identify as Asian not be as concerned with personal data being collected), although this effect is a slightly weaker claim given the larger  $p$ -value ( $p = 0.15$ ).

**Table 4** Demographic counts for the number of students within each category for ethnicity, year of study, gender, and self-reported first-generation college student status

Demographic statistics		(Counts)	
Ethnicity		Year of study	
White	49	Freshman (1st Year)	N/A
Asian	31	Sophomore (2nd Year)	65
Hispanic	29	Junior (3rd Year)	42
Black	10	Senior (4th+ Year)	12
Gender		First-Generation College Student	
Male	44	Yes	27
Female	75	No	92

Note that Freshman are not included as the university database only takes a snapshot of records at the end of a term whereas our survey was administered at the start. The total for each category adds up to 119, which is the total number of people that completed the entirety of the survey, but does not include those who clicked on the email but chose not to proceed

**Table 5** Summary statistics for each of three linear regression models with different target variables: institutional trust, IUIPC Collection subscale, and the stakeholder question for comfort regarding instructor data use

	Coeff.	Std. Error	<i>t</i>	<i>P</i> >   <i>t</i>	[0.025	0.975]
<b>Instl. Trust</b>						
Female	0.384	0.226	1.696	0.093	−0.065	0.832
First Gen.	−0.022	0.260	−0.084	0.934	−0.538	0.494
Junior	0.067	0.230	0.293	0.770	−0.388	0.522
Senior	−0.225	0.367	−0.613	0.541	−0.952	0.502
Black	−0.672	0.405	−1.659	0.100	−1.474	0.131
Asian	−0.304	0.268	0.238	0.812	−0.479	0.610
Hispanic	0.065	0.275	0.238	0.812	−0.479	0.610
<b>IUIPC Collect.</b>						
Female	0.505	0.294	1.715	0.089	−0.078	1.088
First Gen.	0.157	0.339	0.464	0.644	−0.514	0.829
Junior	0.240	0.299	0.803	0.424	−0.352	0.832
Senior	−0.029	0.478	−0.060	0.952	−0.975	0.918
Black	−0.025	0.527	−0.048	0.962	−1.070	1.019
Asian	−0.506	0.349	−1.450	0.150	−1.197	0.186
Hispanic	0.077	0.358	0.216	0.830	−0.632	0.786
<b>Instruct. Data Use</b>						
Female	0.585	0.273	2.144	0.034	0.044	1.125
First Gen.	−0.175	0.314	−0.559	0.577	−0.797	0.447
Junior	0.190	0.277	0.686	0.494	−0.359	0.739
Senior	−0.401	0.442	−0.907	0.367	−1.278	0.476
Black	0.180	0.499	0.368	0.714	−0.788	1.147
Asian	0.202	0.323	0.624	0.534	−0.439	0.842
Hispanic	0.176	0.332	0.530	0.597	−0.481	0.833

In all three cases, the inputs are demographic categories. Note that the reference groups mentioned in “Quantitative Analysis: Understanding Relationships Between Key Consent Factors and Demographics” are excluded, as this would be a linearly dependent column. Once again, the coefficients, errors, t-statistics, p-values, and confidence interval bounds are provided. No significance levels have been explicitly chosen

## Qualitative Analysis Findings

Our analysis of students’ responses to the open-ended survey questions revealed more nuanced student perspectives on data collection and factors influencing their willingness to consent. We found that students recognized that allowing their student data to be used in learning analytics systems can beneficially contribute to improving education, supporting new research, and positively impact other students, while also expressing concerns for data privacy, data collection and ambiguity around data

**Table 6** Decision factors reported by students who consented

Factors affected students to consent	Count	Percentage
Value data contribution to improve education/learning/teaching	28	30.43%
Value data contribution for new understanding & research insights	19	20.65%
Positive impacts on other students and future generations	19	20.65%
Value data contribution to support general research	16	17.39%
Data privacy and security considerations	15	16.30%
Value data contribution to improve learning analytics systems/tools	13	14.13%
Value data contribution for generic support	13	14.13%
Neutral or no impact on student themselves	13	14.13%
Value data contribution for data completeness and accuracy	11	11.96%
No concerns	11	11.96%
Other	12	13.04%
Positive impact on the university	7	7.61%
Positive impact on other people	7	7.61%
Positive impact on student themselves	6	6.52%
Generic mentioning privacy	6	6.52%
Positive impact on faculties and staff	5	5.43%
Survey compensation	5	5.43%
Trust in the institution	3	3.26%
Trust in the researchers	2	2.17%

usage. Second, based on the patterns we have identified through statistical analysis, we report corresponding qualitative findings providing insights on students' rationales behind such patterns. Namely, students who commented on trusting the institution all consented, while those who said they distrust the institution or the researchers denied consent. Students had varying views on data collection depending on the context of use and the stakeholders involved. We further probed students' privacy perceptions and found diverse connections. Students' responses revealed that they considered instructors to be key stakeholders and users of student data.

### Reported Important Factors in Consent Decision

Among the 116 responses regarding important factors that affected students' consent decision, 92 related to positive consent decisions, and 24 denied consent. For the 92 students who consented, we identified<sup>2</sup> 19 factors that affected their decision to consent (see Table 6). 30% of students who consented valued that their data would be contributing to the improvement of education, learning, and teaching. A fifth of the students stated their support for allowing student data to be used for advancing understanding and research insights on student learning behaviors, teaching methods,

<sup>2</sup>Some of the student comments include multiple factors.

etc., and 20% expressed willingness to contribute their student data if that could help other students or future generations to learn. Some students (17.4%) pointed out the importance of supporting research, indicating research was a factor that led them to consent. Around 16% stated some levels of privacy considerations when deciding to consent, such as valuing privacy in general, assuming that student data privacy is guaranteed by default, or indicating privacy concerns while still consenting. Student data improving learning analytics systems (14.1% of the students) and believing that allowing student data to be used is a purposeful and meaningful act (14.1% of the students) were two factors valued by students. For instance, one student said that *“thinking about the greater good influenced my decision. If my student data can help improve quality of education overall, I would support its use.”* Roughly 14% had a neutral response to the use of student data, for example: *“I have nothing to lose when giving my data,”* while 12% did not identify any concerns. 12% of students believed that contributing to student data use is important to ensure data completeness and accuracy.

Of the 24 students who denied consent (see Table 7), 63% expressed privacy concerns (e.g., data breach, uncomfortable sharing student data), and 20% expressed concern over lack of transparency on how data is collected, used, and by whom. 16.7% talked about the lack of proper compensation for the use of their student data. 12.5% were concerned about potential negative impacts on them, such as *“I would be worried about my academic data being used in a way that negatively affects me.”* A few students denied consent due to their distrust in the institution or the researchers. One student noted that use of student data could harm marginalized groups, and one student noted a lack of agency and control regarding student data use.

### Perceived Benefits of Data Use in Learning Analytics

Students also identified benefits of using student data for learning analytics regardless of their consent practice (see Table 8). The top three benefits mentioned by all

**Table 7** Distribution of factors from students denied consent

Factors affected students to deny consent	Count	Percentage
Data privacy and security concerns	15	62.50%
Lack transparency of data collection, access, analysis, and purpose	5	20.83%
Lack compensation to use student data	4	16.67%
Negative impact on student themselves	3	12.50%
Misc.	3	12.50%
Distrust in the institution	2	8.33%
Distrust in the researchers	1	4.17%
Negative impact on marginalized students	1	4.17%
Data leakage concern	1	4.17%
Lack control over student data	1	4.17%

**Table 8** Distributions of benefits in using student data

Benefits of using student data for learning analytics	Count	Percentage
Value data contribution to improve education/learning/teaching	60	52.17%
Value data contribution for new understanding/research insights	44	38.26%
Benefit other students and future generations	40	34.78%
Value data contribution for data completeness and accuracy	25	21.74%
Benefit student themselves	16	13.91%
Benefit the university	10	8.70%
Value data contribution to improve learning analytics systems/tools	8	6.96%
Benefit faculties and staff	8	6.96%
Neutral or no impact on student themselves	7	6.09%
Benefit other people	6	5.22%
Value data contribution for generic support	3	2.61%
Benefit the society	2	1.74%
Misc.	2	1.74%
Value data contribution to support general research	1	0.87%
Survey compensation	1	0.87%

students were the value of contributing data to improve education (52%), supporting research (38%), and positive impact on other and future students (35%). 21.7% mentioned wanting to ensure data completeness and accurate analysis. 14% thought they might be positively affected, for instance: *“I think it’s important that my data be used to better improve and optimize our learning environments, which will benefit not only me, but the students that will be coming after me.”* Eight students (7%) supported using student data to improve learning analytics systems. Students also recognized that student data use could positively impact the university (8.7%), faculty and staff (7%), and “others” without specifying who (5.2%).

### Perceived Concerns Regarding Data Use in Learning Analytics

In terms of concerns (see Table 9), over 60% mentioned privacy and security concerns, such as *“I’m a tad concerned that my data could be leaked to the general public. I like some privacy, so I don’t really want everyone to have unfettered access to all my student data.”* 21% worried about inaccurate data interpretation. Data leakage or theft concerned 15 students (13%). 11% worried about a lack of transparency. For instance, *“I am concerned about my privacy, who has access to my student data, and the real purposes for which it is being used (i.e. more than just for optimizing learning for the future).”* While 11% wrote *“no concerns,”* another 10% pointed out negative impacts for them, such as *“I feel it would violate my privacy or that it might be used against me in some way.”* Some worried about data confidentiality (6.8%) or insufficient student agency and control (5.1%), and one student did not trust the institution to use the data responsibly.

**Table 9** Distributions of concerns in using student data

Concerns of using student data for learning analytics	Count	Percentage
Data privacy and security concerns	71	60.68%
Inaccurate interpretation/representation from the data and analytics	25	21.37%
Data breach/leakage/stolen	15	12.82%
Lack transparency of data collection, access, analysis, and purpose	13	11.11%
No concerns	13	11.11%
Negative impact on student themselves	12	10.26%
Confidentiality	8	6.84%
Misc.	8	6.84%
Lack control over student data	6	5.13%
Distrust in the institution	1	0.85%
Generic mentioning privacy (not concerns)	1	0.85%
Lack compensation to use student data	1	0.85%

### Students' Trust & Distrust in Institutions

Our quantitative analysis revealed institutional trust to be a significant predictor of students' consent. Among students who consented, four (3 females: 1 Asian, 1 White, 1 prefer not to disclose; 1 male: Black) explicitly expressed trust in the institution, mentioning its reputation, accountability, and research methods (e.g., to use data properly), respectively. Another student trusted that researchers would be "*handling my data appropriately and not abuse it.*" In contrast, two students who denied consent (1 male, 1 non-binary; both White) expressed distrust in the institution. One of them did not "*trust the university to use this data in a way that won't hurt marginalized students.*" The second did not trust that "*data wouldn't be used for commercial purposes.*" Additionally, two Black students who denied consent (1 male, 1 female) distrusted researchers because it was unclear how they might use student data, and the related privacy risks and potential harms.

Our quantitative analysis further showed that Black students generally tended to trust the institution less. Of the nine Black students who completed the online survey (6 females, 3 males), 7 consented and 2 denied (1 male, 1 female). One student who denied consent expressed distrust: "*I'm not sure what they're using the data for and I don't trust it.*" In contrast, most of the 7 consenting students focused on benefits of student data use such as "*potentially help another student in the future,*" "*be more informative and beneficial on teaching/learning methods and tools than self-report,*" and "*the university and other students can improve from analyzing my student data.*"

### Student Perspectives on Data Collection

As our quantitative analysis further showed that students' propensity to consent to learning analytics data use is negatively correlated with their concerns for data collection, we analyzed all open-ended responses that explicitly mentioned data collection.

Nine female students and one male student, from a range of ethnic backgrounds commented on data collection (comments did not differ based on gender or ethnicity). Two female students who did not consent (1 Asian, 1 White) cited discomfort with sharing personal data as the reason. One of them also expressed privacy concerns: *“I am not sure which information about me is being collected and analyzed and how will that information be applied to optimize learning...who would get access to my information and to what extent.”* The other 8 students, who all consented, expressed mixed attitudes towards data collection. The majority of them supported data collection for better understanding students’ performance, more accurate and representative results, research, and improving student learning. This suggests that students who emphasize benefits of data collection are more inclined to consent.

We further looked at students’ responses mentioning privacy to shed further light on their attitudes toward data collection. 65 students (41 consented, 24 denied) mentioned in privacy concerns in 92 individual responses. 72 of these responses expressed data collection concerns. Notably, there were no distinctive differences between answers from students who consented and those who did not. 33 of these students (20 consented: 16 females, 4 males; 13 denied: 8 females, 1 non-binary, 4 males), with diverse ethnic backgrounds (consented: 2 Black, 5 Hispanic, 6 Asian, 6 White, 1 American Indian or Alaska Native and Black); denied: 2 Black, 1 Hispanic, 3 Asian, 6 White, 1 Middle Eastern or North African), stated concerns and uncertainty about potential data misuse, and lack of transparency regarding data collection, data access, and data sharing. These students further expressed concerns about possible abuse or misuse of student data, such as by researchers, “the system”, for marketing, or to sell information. For instance, one student stated *“what would happen after my data is used for its primary purpose? Does it just sit in a database, available to anyone for other use without my knowing or consent? Does it get deleted?”*

Privacy concerns of 10 female and 5 male students (10 consented, 5 denied) focused on data security, leaks, and improper exposure. 5 students who denied consent (2 White males, 1 Hispanic/Latinx female, 1 Asian male, 1 Middle Eastern or North African female) mainly worried about student data being compromised if the system is not secure enough, others noted how general privacy and security issues factored into their decision not to consent: *“It feels like my data isn’t safe with anyone...how can I trust any group when every day there are news stories about major platforms/companies failing their users, intentionally or not?”* The 10 consenting students also expressed concerns about potential data leakage, exposure, or hacking, and how extracted data could be *“used against me”* or *“affect my future job opportunities.”* Related, 11 students, 9 male and 3 females, (9 consented, 2 denied) noted risks of being identified. The students who denied (2 White males) stated that *“the data won’t stay anonymous”* or *“more people would know my information,”* which is similar to what those who consented said. However, we observed that students who consented tended to comment with a more trusting tone, assuming that collected student data would be aggregated and anonymized. This suggests, that whether data is de-identified and handled properly affects students’ consent propensity.

We noted earlier that all students who distrusted the institution denied consent. Trust also came up in relation to privacy concerns. Six students (3 males: 2 White, 1

Black; 3 females: 2 Hispanic/Latinx, 1 White), of whom 2 consented and 4 denied, expressed their distrust or discomfort regarding data collection, use, and access. The four denying students, 3 White and 1 Black, were uncomfortable with their information being gathered and known by different parties: “*I would prefer to not have my information (i.e., classes, my learning tools) being gathered. I feel a little uncomfortable*” and “*I wouldn’t trust other people looking at my personal information.*” Thus, students’ distrust and discomfort may also explain their reluctant attitude toward data collection, which is a driving factor influencing students’ consent propensity.

### **Views on Instructor Use of Student Data**

Our quantitative analysis further revealed that students are more likely to consent to “help instructors gain insights about students’ engagement.” 9 students, of whom 7 consented (2 White female, 2 Black females, 1 Asian female, 1 Asian male, 1 male did not disclose ethnicity) and 2 denied (1 White male, 1 White female), noted benefits of student data use by instructors. Some believed instructors can use student data to improve teaching methods and optimize the learning process, others felt that it helps instructors better understand different types of students to provide more personalized support. It seems that students view instructors as key stakeholders and users of student data, and students are relatively more comfortable with such data use, which is positively related to their consent practice.

## **Discussion**

We identified three main factors that influence a student’s willingness to consent to learning analytics data use: degree of trust in the institution, concern regarding personal data collection, and comfort with instructors using data to gain insights on student engagement. We now discuss our findings’ contributions to the knowledge regarding students’ privacy perspectives and behaviors in an educational setting. First, we acknowledge limitations in extrapolating results due to our survey instrument. We then highlight how varying engagement rates may suggest that some students are not being well represented and what this implicates when building AI systems and soliciting consent decisions. Next, we explore key factors and demographic trends in our findings, paying particular attention to the importance of instructors and discrepancies in trust between subpopulations. We end by discussing how institutional contexts factor into consent.

### **Survey Instrument Limitations**

We acknowledge that an online survey is limited in trying to identify the reasons underlying consent, especially for those who are already weary of sharing data online. However, we believe this data collection is reasonably realistic (though by no means ideal) about how asking for consent may be conducted at a university: namely in an online fashion, likely via email or through a prompt for a particular learning platform. In our study, the consent message was not specific about the data collection purpose,



so as to prevent priming participants. Ideally, purposes should be clearly specified and consent should be specifically for a certain purpose.

Our study's ecological validity may be affected in that the views expressed by certain subgroups in the survey here may not apply to all students in that subgroup. A similar argument may be made for the feature selection process: there may be other considerations that students find critical to their decision, which are not reflected here. This is why we explicitly analyze relationships with ethnicity only after pruning the scope of potential consent-related factors. Yet, these findings can help narrow the scope of future work in order to further pinpoint how specific interactions, such as instructor trust (discussed in [The Role of Instructors in Key Factors for Consent](#)), relate to engagement by subgroup. We also emphasize that while the survey used to identify factors involved in students' decision to consent was one component of our study (RQ1), another key finding for RQ2 was to empirically capture a consent process that could be reasonably undertaken, as described previously, and that not responding or clicking through is an important participation characteristic that the email prompt measures.

Nonetheless, since students are hypothetically answering their consent decision on allowing learning analytics to use student data, this might not align with their actual behavior and decision in a real world data sharing context. Consequently, mentioning a specific use case (e.g., an early-warning system used to assist student learning) or incorporating deception, may influence students' consent propensity or increase the response rate due to greater perceived relevance and urgency. With that said, this must be weighed with ethical considerations and protocols, such as including a debrief immediately after and recognizing that it may have the side effect of diminishing the level of student trust in learning analytics research for some students, even if the risks are deemed to be minimal.

### **The Difficulty Posed by Varying Engagement Rates**

As shown in [“Response to Email Prompt”](#), there is a difference in response rate by subpopulation. Namely, Black students responded to the email request at a significantly smaller percentage than expected, while responses from those identifying as White were overrepresented even when we account for differences in number of emails sent to each group. The fact that underrepresentation exists supports the well-established theories described in [“Background and Related Work”](#) and is not a surprise. However, we reiterate that institutions seeking input from students regarding data use may be receiving a biased sample – not only because there are minorities in the population, but also because those who are underrepresented are even less likely to respond to a survey.

With that said, it is possible that students choose to ignore the question due to a number of factors such as the framing of the email, the perception that their decision does not matter, and for other reasons that we cannot ascertain. However, the click rate still provides key additional context, since non-participation still represents someone who the stakeholder would be making a decision for on their behalf: the data is either used, or withheld. In other words, we make no claims that intention can be derived from non-responses or the click rate, but this metric can still be linked to

demographics in order to better understand the possible range at which data may be used if we were to treat those non-responses as decisions of consent or non-consent for the purposes of training a machine-learned model per se.

Thus, to avoid potential erosion of trust and tension around data usage, it may be beneficial to further explore the addition of nudging indicators to move away from broad consent practices and lessen the gap around response rate in student elicitation surveys. For instance, a prompt might mention specific data sources and stakeholders involved. This could help students understand the implications of their decisions, but may backfire if the details are overly complex and are thereby skipped or not comprehended. We could include a short paragraph explaining that it is invaluable for underrepresented students' to provide their input to avoid biases in predictive models and to improve the educational quality for all students; social framing has been shown to impact privacy decision-making (Coventry et al. 2016). This may be presented to a random sample of students, or shown only to minority students. Other options may be varying levels of compensation based on subpopulations that are most lacking in data—paying for data that is scarce, a more costly option that brings up the broader issue of data ownership and value. It is important to note that the focus here is simply on getting people to indicate their consent preference either way, and not necessarily to nudge them to choose consent, though it is possible such changes could impact both response and consent rates (Utz et al. 2019). Consequently, depending on the intent of the actor, nudging may advance an institution's goals while shifting the responsibility of bias towards students and countering their self-interests, so it is important to design prompts carefully.

### **The Role of Instructors in Key Factors for Consent**

All three key factors we identified seem plausible and have intuitive explanations. It is probable that students who trust the institution and are more comfortable with instructors using data to improve classroom instruction would be more likely to share data. It is also understandable that concerns regarding personal data collection would decrease the likelihood of such an occurrence. However, of all the subscale measures and various stakeholders, instructor considerations are the most significant. This demonstrates that instructors may have significant influence on students' willingness to share their data, perhaps even overshadowing broader concerns about general data collection or institutional practices. It is especially plausible that institutional trust and comfort with instructor are both key consent factors that influence each other, as prior studies have shown that a sense of belonging to the university affects retention and engagement (Zepke and Leach 2010), among other factors, and that teacher-student relationships contribute to these feelings of rapport (Hagenauer and Volet 2014). It is likely that students have more opportunities to form closer relationships with instructors and interact with them on a more frequent basis, whereas institutions may be attributed to administrators and other officials whose roles and direct impact are less easily ascertained.

Researchers are often required to provide consent opportunities for research subjects, but neither instructors nor institutions have such requirements or expectations to do so. Learning analytics are increasingly embedded in the tools adopted by

institutional actors such as instructors and advisors, and privacy-related decision are made on behalf of students by institution's information officers and legal counsel. However, such norms may not reflect where institutions or society are heading. Trends in recent years such as the GDPR and CCPA/CPRA demonstrate rising interest regarding issues of privacy and consent. Questions regarding consent in the coming decade, such as the ones raised in this paper, are thus important to highlight. At UM, students regularly have to consent to data sharing when using certain external or third-party tools (e.g. Learning Tools Interoperability or LTI); it is not unfathomable that instructors may play a more direct role regarding consent practices in the near future.

Consequently, some suggestions for tangible interventions include having the instructors provide more transparency regarding student data use in the classroom. For example, by telling students what educational and demographic data is being collected, the purpose for its use, and the people who have access, it may shift students' comfort with instructors use of data, thereby changing the likelihood of student consent. Similarly, having an instructor send an email prompt asking for students to consent or deny use of their data may elicit more people agreeing to share data as opposed what may have been seen as a unilateral institutional action. Future research on the relationship between specific instructors and student trust, and whether there are correlations with various degree programs and departments across campus may yield context to better target such interventions.

### **Ethnicity and Gender Trust Gaps and The Role of Institutions**

While it is more difficult to draw concrete conclusions between demographics and key factors in willingness to consent reported in the online survey due to lower sample sizes and overall power, some of the qualitative comments provide evidence supporting our quantitative results, such as the anticorrelation between institutional trust and students identifying as Black. Yet, there remains a fair number of Black students who did indicate some level of trust in the institution according to their open-ended responses. Discrepancies between gender displayed fairly strong signals when it came to all three key factors. Specifically, we find that those who identify as female tend to trust the institution and instructors more, despite greater concerns regarding data collection as a whole. This may seem contradictory, though perhaps it is the case where those who identify as female are hesitant, but make an exception when data use is situated in an educational context due to trust in the institution and/or the instructor. Therefore, conducting a confirmatory analysis to identify deeper rationale behind these cost-benefit considerations would be beneficial.

Additional research identifying whether certain factors outweigh others may be extracted in follow-up surveys and semi-structured interviews with specific groups of learners corresponding to student characteristics identified to be correlative with key factors such as Black, female, or students at the intersection of these identities. Questions about students' personal experiences at their university, relationship with instructors and other stakeholders, as well as personal beliefs and attitudes around data collection specifically would provide insights into what agency students desire with respect to learning analytics. It may also uncover whether these decisions are

based on firmly ingrained biases or actionable concerns, and help contribute to a more realistic model of student choices and its effects on predictive modeling in learning contexts.

Lastly, we want to differentiate the restrictions between institutional ethics boards that oversee research study procedures versus institutional data consent and collection practices. In the US, institutional review boards (IRBs) approve and monitor research involving human subjects, which includes this study. Title 45 in the Code of Federal Regulations (i.e. Common Rule) contains provisions regarding how informed consent is obtained and additional protections for certain vulnerable populations. Even then, §46.104 and §46.116 lists various exceptions where consent may be waived. Regardless, such regulations do not necessarily restrict what an institution can do; institutions wishing to engage students to gain consent for learning analytics tools or technologies do not need to do so under the banner of scholarly research.

There are many reasons why data may be collected or processed without explicit consent by the data subject. The GDPR, for instance, recognizes six legal bases of data processing of which consent is but one (Article 6). At the same time, this article also creates a potential data processing exemption for scientific research or statistical purposes, thereby leaving situations that may be up to an institution's discretion.

The question we explore with this study is what the effects would be if institutions *did* engage with students for their consent to use the data. While good consent statements are transparent and meant to inform users of particular data-use practices, we made a deliberate decision to simplify the consent form, inspired by what today's consent dialogues are, to study the general act of consenting. This is similar to a click-through End-User License Agreement (EULA). There are strong arguments for requiring explicit consent or opt-in especially when data processing is likely to be unexpected or surprising for the data subject (Rao et al. 2016; Schaub and Cranor 2020), such as a use of data that is not readily apparent from transaction context (e.g. an LMS used to determine participation grades or what assignments to offer), but how such policies might influence trust and consent behaviors are unclear and left for future experiments.

## Context Dependency and its Importance in Learning Analytics

Even with extensions to uncover detailed patterns of reasoning through consent decisions, it is important to keep in mind that the dataset used in this study consists solely of records from a single major university in the United States with many initiatives to promote ethical innovations in learning analytics. Data collected at other institutions may have different underlying distributions and lead to distinct results with different conclusions; this is why we have provided contextual information in “[Recruitment & Participants](#)”. Conducting similar analyses at community colleges or other educational settings may help generalize the results of this paper, especially as Li et al. (2019) demonstrated a need to reevaluate predictive models when the training set for predictive models is altered. A cross-cultural survey administered at institutions around the world might allow for privacy expectations to be better understood.

With that said, it is not always possible to share demographic information across institutions due to data privacy concerns, and so we might want to ensure the reliability of self-reported demographics. Yet, of the people who did not omit their response, there was a perfect agreement between the self-reports and university records for ethnicity, and a 98% agreement for gender; we expect the gender discrepancy as we supported a non-binary option in our survey, while gender in our institution's records are still being reported dichotomously. The high levels of agreement also show that it may be possible to obtain accurate demographic information by only asking students to self report demographics without resulting in significant loss of data. Not only does this give students more agency over what they choose to share, it also suggests that one may be able to conduct similar studies requiring demographic information at other institutions, even if those details are not available to researchers or centrally stored.

Eventually, inter-institutional datasets compiled from various types of educational institutions around the world may be joined in order to enrich our understanding of consent factors as it relates to diverse sociocultural backgrounds. On the algorithmic front, one may obtain more specific estimates of data sharing ranges, and demographics can be tied with students' responses to calculate various opt-out ranges since different subgroups have different privacy perspectives and rates of participation, as we have shown. The performance of predictive models may therefore incur greater differential effects, which will continue to necessitate further research to ensure there is a balance between maintaining trust and personal privacy while advancing education and ensuring fairness across diverse student populations. We may also obtain a greater understanding of what aspects of institutional structure and policy are most effective at fostering a culture that encourages broad participation and minimizes paternalistic policies of data collection. For instance, many institutions are hierarchical and rely on audit culture, and so differences in measures of success as well as flatter administrative structures with shared duties and codified values may have an impact on student and staff behavior as well as baseline participation rates.

## Conclusion

In this study, we have addressed two critical questions regarding the use of students' educational data in learning analytics. RQ1 asked about students' perspectives on their educational data being used for learning analytics systems, and RQ2 sought to find the population and participation characteristics of students who indicated a preference to allow or deny such usage.

For RQ1, we have identified three primary factors: trust in the institution, concern with individual data collection, and comfort with instructor data use for student engagement that influence a student's willingness to consent to their data being used in learning analytics. Higher levels of institutional trust and greater comfort with the idea of instructors using educational data for instructional improvement correlate with much greater probabilities that a student will allow their data to be used for learning analytics. By contrast, apprehension to personal data collection leads to a lower chance of consent.

Given these factors, we then explored RQ2 and found that students who identified as female were more likely to trust the institution and instructor data use than male students, but were also more generally concerned about data collection practices. Meanwhile, Black students indicated lower levels of institutional trust. We also note that female students had a higher response rate and that White students were over-represented while Black students were underrepresented among people who made a consent decision.

While there are some limitations that come with survey instruments and the fact that this study was conducted at a single university, our findings surface important implications for institutions to consider when collecting data for learning analytics and we layout additional routes for confirming and generalizing the results presented here. We demonstrate that varying engagement rates reflect existing educational disparities between minority students and that instructors can influence students' consent decisions.

The findings for RQ1 and RQ2 illustrate that it is insufficient to only provide students' with consent prompts and expect unbiased data without a more concerted effort to involve instructors and institutional decision-makers. We support student agency and agree that collecting or using data without student consent violates that, though we also emphasize that consent is only a part of the puzzle in allowing for greater informational self-determination; it may take a combination of many other forms of agency such as transparency, data access, or opportunities to object. Balancing student agency is therefore not a straightforward matter and requires careful consideration and design.

The differential response rates we identify show that perspectives from those who are underrepresented are still not properly accounted for even when stratifying equally across groups. Additionally, the difference lies not only with consent rate, but also in the perspectives held by respective subpopulations; there are unique underlying perspectives that guide individuals' actions from each of these groups. Therefore, relying solely on such an approach for data collection is likely to continue producing biased predictions based on biased data, influencing the efficacy of educational technology and its potential to treat students fairly. In order to ensure the ethical use of AI in education, the method of data collection is important, but it is imperative not to fixate solely on this aspect. Taking note of students' major concerns and steps to strengthen trust in their institution's numerous stakeholders should be addressed via tangible actions such as implementing transparent data practices; allowing inequity to continue will only serve to increase mistrust, thereby lowering engagement and affecting institutions' ability to support all its students. Instead, by understanding the key factors that influence consent and their relation to students' personal background, institutions will be better equipped with the knowledge to enable technology-supported education while maintaining ethical data use and public trust.

## **Appendix A: Debrief, Consent, and Survey Instrument**

(After the student clicks on yes or no in the email, a new page pops up, which contains more survey questions)

Thank you for your participation in the study!

### **Debrief on Study Purpose**

In the previous email you were asked whether you would be willing to allow your data to be used or not for learning analytics systems. The goal of this study is to understand willingness and attitudes to share educational data by different subpopulations of students in order to ensure that future learning analytics systems are equitable and fair for all groups of students. For this reason, we are studying students' opt-in and opt-out propensity in combination with demographic information retrieved from their student records. This purpose was concealed in the email message to prevent biased responses. Please note that your response to the email and your demographic information obtained from your institutional demographic records are de-identified to protect your anonymity and privacy and cannot be linked back to you. As a result, it is also not possible to withdraw from the study (we would not be able to identify your response).

We'd like to ask you a few questions regarding your response, and this should take no more than 15 min to complete. This survey is anonymous. Responses will be used for academic research and may appear in publications about our findings. If you have further questions or concerns about the study, please contact our research team. Upon completion of the survey, you will be given a \$5 Mastercard gift card as a thank you for your time and participation.

To participate in the survey, please proceed to the next page...

### **Name of the Institution Consent for Student Participation in a Research Study**

#### 1. Study Overview

We invite you to a study that explores student perceptions regarding the use of students' educational data in learning analytics systems by filling out this survey which asks your perspectives on the use of student data in learning analytics and some demographic information. Taking part in this research project is voluntary. You do not have to participate, and you can exit the survey at any time. Please take time to read this entire form before deciding whether to take part in this research project.

#### 2. Purpose of This Study

The objective of this study is to understand what factors might affect students' willingness to opt-out of having their educational data being used in learning analytics systems.

#### 3. Who can participate in the study

Participation in the study is through the email invitation you have received.

#### 4. Information about study participation

If you complete the survey, your answers will be recorded. This takes about 20 min to complete.

#### 5. Information about Study Risks and benefits

There are no known risks associated with this study. However, because this study collects information about you, there is potential for a breach of confidentiality (see below for protection mechanisms). This study will help us and other [Name of the Institution] educators better understand how to use educational data to build services for students like you.

6. Ending the Study

You are free to exit the survey at any time without penalty or cost.

7. Financial Information

For your completion of the study, you will receive \$5 gift card by mail. All information collected on the remuneration form will not be connected to your survey responses.

8. Protecting and sharing research information

Only authorized study team members will have access to collected data using approved university systems. After linking responses to institutional demographic records all identifiable information will be removed. Anonymized data may be retained for future studies.

9. Contact Information

Please contact the researchers listed below to obtain more information about the study or express any concerns you may have.

If you have questions about your rights as a research participant, or wish to obtain information, ask questions or discuss any concerns about this study with someone other than the researcher(s), please contact the following: [Name of the Institution] Health Sciences and Behavioral Sciences Institutional Review Board (IRB-HSBS)

Beginning of the Survey Question:

Learning analytics are systems that measure, collect, analyze, and report data of learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.

- In your response to the email, you indicated that you choose to opt into (or opt out of) having your student data being used by learning analytics systems, please describe the important factors that affect your decision. [blank for students to fill]
- In your response to the email, you indicated that you choose to opt into (or opt out of) having your student data being used by learning analytics systems, please describe the important factors that affect your decision. [blank for students to fill]
- What do you see as the benefits of having your student data being used by learning analytics systems? [blank for students to fill]
- What do you see as the concerns of having your student data being used by learning analytics systems? [blank for students to fill]
- There are different stakeholders at the university when it comes to how students' data are used. Consider how comfortable or uncomfortable you are with the following scenarios: [Extremely Uncomfortable (1), (2), (3), Neutral (4), (5), (6), (7) Extremely Comfortable] My educational data is used in learning analytics systems to...



help me monitor my course activities.  
 help other students monitor their course activities.  
 help instructors gain insights about students' engagement.  
 help academic advisors monitor individual student performance.  
 help administrators understand student enrollment and academic achievement.

- We're interested in learning more about how you interact with other students. Please indicate to what extent the following statements describe you? [Completely false = 1, 2, 3, 4, 5, 6, Completely true = 7]

I like to help other students learn.  
 I like to share my ideas and materials with other students.  
 I like to cooperate with other students.  
 I can learn important things from other students.  
 I try to share my ideas and materials with other students when I think it will help them.  
 Students learn lots of important things from each other.  
 It is a good idea for students to help each other learn.  
 I like to do better work than other students.  
 I work to get better grades than other students do.  
 I like to be the best student in the class.  
 I don't like to be second.  
 I like to compete with other students to see who can do the best work.  
 I am happiest when I am competing with other students.  
 I like the challenge of seeing who is best.  
 Competing with other students is a good way to work.  
 I don't like working with other students in school.  
 I like to work with other students. (reverse)  
 It bothers me when I have to work with other students.  
 I do better work when I work alone.  
 I like work better when I do it all myself.  
 I would rather work on school work alone than with other students.  
 Working in small groups is better than working alone. (reverse)

Next we'd like to learn more about you.

- Please indicate to what extent the following statements describe you? [1 = never or definitely no, 2, 3, 4, 5, 6, 7, 8, and 9 = always or definitely yes]

I'd rather depend on myself than others.  
 I rely on myself most of the time; I rarely rely on others.  
 I often do "my own thing."  
 My personal identity, independent of others, is very important to me.  
 It is important that I do my job better than others.  
 Winning is everything.  
 Competition is the law of nature.  
 When another person does better than I do, I get tense and worked up.  
 If a peer gets a prize, I would feel proud.

The well-being of my peers is important to me.  
 To me, pleasure is spending time with others.  
 I feel good when I cooperate with others.  
 Parents and children must stay together as much as possible.  
 It is my duty to take care of my family, even when I have to sacrifice what I want.  
 Family members should stick together, no matter what sacrifices are required.  
 It is important to me that I respect the decisions made by my groups.

(Note: for the statement “When another person does better than I do, I get tense and aroused.” We changed the wording “aroused” to “worked up” based on feedback from survey pilot testing. We also revised the wording of the question prompt to be clearer.)

Next, we are interested in learning about your relationship with the university.

- In the following questions, “employees” may refer to any faculty, instructors, staff, or administrators at [Institution Name]. Please indicate to what extent do you agree or disagree with the following statements? (1 = Strongly Disagree, 2, 3, 4, 5, 6, 7 = Strongly Agree)

Since I am unable to personally monitor all of [Institution Name]’s activities, I rely on the employees of the college to get the job done right.  
 I have faith in college employees to do those things that relate to my education at [Institution Name] that I cannot do myself.  
 I am confident that college employees do those things that relate to my education at [Institution Name] that my family cannot do for me.  
 (R)In general, I do not have confidence in [Institution Name].  
 I believe [Institution Name] is a credible organization.  
 I feel that I can rely on [Institution Name].  
 I believe [Institution Name] is capable of designing academic programs that meet student needs.  
 I believe [Institution Name] employees are experts in the positions that they hold.  
 (R)Generally speaking, [Institution Name] employees are untrained.’  
 People with relevant work experience are employed in all areas at [Institution Name].  
 [Institution Name] does things competently.  
 (R)Unfortunately, [Institution Name] does things poorly.’  
 [Institution Name] employees perform their tasks with skill.  
 [Institution Name] does things in a capable manner.  
 (R)[Institution Name] is not sympathetic to my needs.’  
 I believe that [Institution Name] is a friendly college.  
 In general, I like the attitudes of [Institution Name] administrators, faculty, and staff.  
 [Institution Name] treats me in a friendly manner.  
 (R)[Institution Name] employees treat others better than they treat me.’  
 [Institution Name] will always deal with me in a friendly manner.

[Institution Name] attempts to develop a friendship with me.  
 [Institution Name] uses its knowledge and experience to clear the confusion of the educational process.  
 [Institution Name] employees tell me what they are thinking.  
 [Institution Name] employees tell me what is on their mind.  
 Employees at [Institution Name] share their thoughts with me.  
 (R)[Institution Name] keeps information from me.’  
 [Institution Name] is sincere in what it promises to students.  
 I believe that [Institution Name] is honest when dealing with me.  
 I believe that [Institution Name] will always be honest in its associations with me.  
 [Institution Name] follows through on promises made to me.  
 (R)Keeping promises is a problem for [Institution Name].’  
 If [Institution Name] promises something to me, they will stick to it.  
 Academically, [Institution Name] does things that they promise to do for me.  
 [Institution Name] does not hold academic integrity as a standard by which to live.’  
 [Institution Name] strives to be a perfect academic organization.  
 [Institution Name] always tells me the truth.  
 [Institution Name] employees would not lie to me.  
 [Institution Name] deals honestly with me.  
 (R)Sometimes [Institution Name] does dishonest things.’

(Note: we added a preface to explain the purpose of the questions, and we added a definition for “Institution Name employees” based on feedback from survey pilot testing)

Next we’d like to learn more about you.

- Please evaluate to what extent you agree or disagree with the following statements [(1 = Strongly Disagree, 2, 3, 4, 5, 6, 7 = Strongly Agree)]

Consumer online privacy is really a matter of consumer’s right to exercise control and autonomy over decisions about how their information is collected, used, and shared.

Consumer control of personal data lies at the heart of consumer privacy.

I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.

Companies seeking information online should disclose the way the data are collected, processed, and used.

A good consumer online privacy policy should have a clear and conspicuous disclosure.

It is very important to me that I am aware and knowledgeable about how my personal information will be used.

It usually bothers me when online companies ask me for personal information.

When online companies ask me for personal information, I sometimes think twice before providing it.

It bothers me to give personal information to so many online companies.

I'm concerned that the online companies are collecting too much personal information about me.

- What's your gender:

Woman

Man

Non-binary

Prefer not to disclose

Prefer to self-describe

- What is your year in your current university program?

Freshmen/1st Year

Sophomore / 2nd Year

Junior / 3rd Year

Senior/ 4th Year

Other

Prefer not to disclose

- Are you a first-generation college student? (Definition: No parent or guardian in the household has a bachelor's degree)

Yes

No

Prefer not to disclose

Are you of Hispanic, Latino, or Spanish origin? Yes(skip Q13) No Prefer not to disclose

- Which of the following racial or ethnic group(s) do you identify yourself as? [Mark all that apply]

American Indian or Alaska Native

White

Asian

Middle Eastern or North African

Black or African American

Native Hawaiian or Other Pacific Islander

Prefer not to disclose

Other [blank to fill]

- I am a

Domestic student [skip Q15]

International student [to Q15]

Prefer not to disclose

- I am from

(list of countries)

Prefer not to disclose

Thank you for your time and input! Your response has been recorded!

## Appendix B: Codebook for Open-Ended Questions

### Layer 1 Code: Value Data Contribution

Layer 2 Code	Definition
For data completeness and accuracy	Mention that people should contribute their data, or just for the sake of completeness. Has to have an emphasis between a single point or partial set of data vs a total/whole dataset in a broader context. When students say without their data the learning analytics/analysis won't be complete or accurate, or that having more data would lead to improvements.
For new understanding/ research insights (education, learning, teaching, student behavior)	When students say the use of student data can help understand or gain insights or research about anything related to learning, education, students' behaviors, backgrounds, situations, performances etc. It is about understanding of the current situation or issues without specifying the goal is to improve something. If students mention "improving" any aspect, code under "improve education/learning/teaching environment/experiences"
For supporting general research (no specification)	Code here when students don't specify the goal of the research. When students say they want to help research/researcher, they value research, they want to be a part of the research by contributing their data, or they think research is important.
For improving or supporting learning analytics systems/tools	When students think positively of learning analytics/tools, and want to help improve learning tools. The "system" here is not the invisible education system, but an actual digital system.
For improvement of education or learning/teaching	When students say help/improve (students) learning, (teachers) teaching, education experiences or environment. Referring to make things better.
For generic support (good causes/purpose, importance)	When students generically say "purpose", "(good) cause", "it's beneficial", "it's important", etc. without other explanation.

**Layer 1 Code: Trust**

Layer 2 Code	Definition
Trust in Institution	When students mentions they have trust in the university or the institution. When you see the word “trust” or other ways to show beliefs and faith in such concepts
Trust in Researchers	When students mentions they have trust in the researchers. When you see the word “trust” or other ways to show beliefs and faith in such concepts

**Layer 1 Code: Distrust**

Layer 2 Code	Definition
Distrust in Institution	When students mention they don’t trust the university or the institution. When you see “don’t trust” or “distrust” or other ways to show disbelief and lack of faith in such concepts
Distrust in Researchers	When students mention they don’t trust the researcher or research. When you see “don’t trust” or “distrust” or other ways to show disbelief and lack of faith in such concepts

**Layer 1 Code: Positive Impact/Benefits on Other Parties**

Layer 2 Code	Definition
Other Students	Positive impact/benefits on other students. Also include “future students” here.
Faculty/Staff	Positive impact/benefits on faculty/staff
University	Positive impact/benefits on the university
Society	Positive impact/benefits on society
Others	When students don’t specify the stakeholders but generally says “benefit others or someone”.

**Layer 1 Code: Negative Impact on Other Parties**

**Definition** Negative impact on other students

### Layer 1 Code: Impact on Students Themselves

Layer 2 Code	Definition
Positive impact on student themselves	This includes only positive impacts on student themselves. Students have to specify “themselves” in the statement. Include statements like “it will benefit me”. If they refer to students in general, don’t code here, but code under impact on other parties.
Negative impact on student themselves	This includes only negative impacts on student themselves. Students have to specify “themselves” in the statement. Include statements like “it will be used against me”. If they refer to students in general, don’t code here, but code under impact on other parties.
Neutral/no impact on student themselves	This includes only neutral impacts on student themselves. Students have to specify “themselves” in the statement. Include statements like “it won’t affect me” or “no impact on me”. If they refer to students in general, don’t code here, but code under impact on other parties.

### Layer 1 Code: Privacy

Layer 2 Code	Definition
Data Privacy & Security Concerns or Considerations	This includes statements have the word “privacy”, “private”, “personal”, or any other synonymous, and show concerning perspectives. This include concerns about inappropriate data access, use, storage and tracking. Code here when data safety is mentioned, or whether they feel safe about their data or not. It could be either students’ personal data safety or the general data safety. It is not about students’ physical safety. If student mention anonymity, code here.
Generic Mentioning Privacy (not concerns)	Student might mention their perspectives or attitude toward privacy. They might only say “I value privacy”. Code such statements here. If student says “I have privacy concerns”, don’t code here and code under “specific privacy concerns”

**Layer 1 Code: Confidentiality**

**Definition** Code only when you see students mention concepts like “confidentiality” or “confidential”. May be that the students think data/information confidentiality is important or that they have data/information confidentiality concerns.

**Layer 1 Code: Data Stolen/Breaches/Leakage**

**Definition** When students mention data stolen, breaches, leakage concepts specifically.

**Layer 1 Code: Lack Transparency (on Data Collection, Access, Use, Storage, Flow, Tracking, Analysis, Interpretation, Algorithm)**

**Definition** Any time the students say, “I don’t know/it’s unclear how data is collected, accessed, used, stored, analyzed, and interpreted by who.” If they are worried about selling of the data, code here as well.

**Layer 1 Code: Inaccurate Interpretation/Representation from the Data and Analytics**

**Definition** Students worry the data analysis provides an inaccurate representation of the students. Code when students say somethings like “the system can’t represent my grade, or I may not be representative of other students.” This is different from interpretation. Code as negative impact as needed.

**Layer 1 Code: Lack Agency/Control/Consent**

**Definition** When students worry about consent; state there is no control of the data, no permission asked, or that they are not involved in the data process

**Layer 1 Code: Data Value/Compensation**

**Definition** When students think their data has value; they want some compensation/benefit

**Layer 1 Code: No Concerns**

**Definition** When students say “I don’t have any issue”, “no concerns/risks,” etc. The difference between “no impact” and “neutral impact on student themselves” is that no concern is very general, without saying anything about the students themselves. Also includes those who assume that this information is public or not personal, or “don’t care”.



## Layer 1 Code: Miscellaneous

**Definition** Other things that do not fit into the other categories. If only part of the statement seems fit in MISC, other parts of the statements can be labeled with other codes. Also, code the whole statement here.

**Acknowledgements** The research reported in this article was made possible in part by a grant from the Spencer Foundation (#201900093). The views expressed are those of the authors and do not necessarily reflect the views of the Spencer Foundation.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alexander, L. (1996). The moral magic of consent (ii). *LEG*, 2, 165.
- Alhadad, S., Arnold, K., Baron, J., Bayer, I., Brooks, C., Little, R., Rocchio, R., Shehata, S., Whitmer, J. (2015). The predictive learning analytics revolution: leveraging learning data for student success. EDUCAUSE Working Group.
- Bailey, S., Burbidge, L., Campbell, P.B., Jackson, B., Marx, F., McIntosh, P. (1992). *The aauw report: how schools shortchange girls*. Washington, DC: National Education Association.
- Bainbridge, W. L., & Lasley, T. J. (2002). Demographics, diversity, and k-12 accountability: the challenge of closing the achievement gap. *Education and Urban Society*, 34(4), 422–437.
- Barocas, S., & Nissenbaum, H. (2009). On notice: the trouble with notice and consent. In *Proceedings of the engaging data forum: the first international forum on the application and management of personal electronic information*.
- Barocas, S., & Nissenbaum, H. (2014). Big data's end run around anonymity and consent. *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, 1, 44–75.
- Baumann, C., & Harvey, M. (2018). Competitiveness vis-à-vis motivation and personality as drivers of academic performance. *International Journal of Educational Management*, 32(1), 185–202.
- Beattie, S., Woodley, C., analytics, S.outter.K., rights, learner.d.ata. (2014). Creepy rhetoric and reality: critical perspectives on educational technology. In *Proceedings Ascilite* (pp. 421–425).
- Benfante, R., Reed, D., MacLean, C., Kagan, A. (1989). Response bias in the Honolulu heart program. *American Journal of Epidemiology*, 130(6), 1088–1100.
- Bensimon, E. M. (2005). Closing the achievement gap in higher education: an organizational learning perspective. *New Directions for Higher Education*, 2005(131), 99–111.
- Boshuizen, H. C., Viet, A., Picavet, H. S. J., Botterweck, A., Van Loon, A. (2006). Non-response in a survey of cardiovascular risk factors in the dutch population: determinants and resulting biases. *Public Health*, 120(4), 297–308.
- Buchanan, E. A. (2011). Internet research ethics: past, present, and future. In *The handbook of internet studies*, (Vol. 11, p. 83).
- Buchmann, C., DiPrete, T. A., McDaniel, A. (2008). Gender inequalities in education. *The Annual Review of Sociology*, 34, 319–337.
- Carnegie Classification IHE (2017). Carnegie classification. Carnegie classification of institutions of higher education. <https://carnegieclassifications.iu.edu/>.
- Carson, L.R. (2009). "I am because we are:" collectivism as a foundational characteristic of African American college student identity and academic achievement. *Social Psychology of Education*, 12(3), 327–344.

- Cho, H., Rivera-Sánchez, M., Lim, S. S. (2009). A multinational study on online privacy: global concerns and local responses. *New Media & Society*, 11(3), 395–416.
- Clarke, J., & Nelson, K. (2013). Perspectives on learning analytics: issues and challenges. observations from Shane Dawson and phil long. *The International Journal of the First Year in Higher Education*, 4(1), 1–8.
- Cofone, I. N. (2018). Algorithmic discrimination is an information problem. *Hastings Law Journal*, 70, 1389.
- Cook, B. J., & Córdova, D. I. (2007). Minorities in higher education twenty-second annual status report: 2007 supplement. Tech. rep., American Council on Education.
- Corbett, C., Hill, C., St Rose, A. (2008). Where the girls are: the facts about gender equity in education. ERIC.
- Coventry, L. M., Jeske, D., Blythe, J. M., Turland, J., Briggs, P. (2016). Personality and social framing in privacy decision-making: a study on cookie acceptance. *Frontiers in Psychology*, 7, 1341.
- Cozma, I. (2011). How are individualism and collectivism measured. *Romanian Journal of Applied Psychology*, 13(1), 11–17.
- Crenshaw, K., Gotanda, N., Peller, G., Thomas, K. (1995). Critical race theory. The key writings that formed the movement. New York.
- Deakin Crick, R., & Goldspink, C. (2014). Learner dispositions, self-theories and student engagement. *British Journal of Educational Studies*, 62(1), 19–35.
- DePass, A., & Chubin, D. (2008). *Understanding interventions that encourage minorities to pursue research careers*. Bethesda: American Council for Cell Biology.
- Deutsch, M. (1949). A theory of co-operation and competition. *Human Relations*, 2(2), 129–152.
- Diaz, V., & Brown, M. (2012). Learning analytics: a report on the eli focus session. EDUCAUSE Review Online Retrieved from <http://net.educause.edu/ir/library/PDF/ELI3027.pdf>.
- Dixon, A. D., & Rousseau, C. K. (2005). And we are still not saved: critical race theory in education ten years later. *Race Ethnicity and Education*, 8(1), 7–27.
- Drachler, H., & Greller, W. (2016). Privacy and analytics: it's a delicate issue a checklist for trusted learning analytics. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 89–98).
- Draper, N. A., & Turow, J. (2019). The corporate cultivation of digital resignation. *New Media & Society*, 21(8), 1824–1839.
- Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? In *The future of children* (pp. 35–54).
- Dwork, C., Immorlica, N., Kalai, A. T., Leiserson, M. (2018). Decoupled classifiers for group-fair and efficient machine learning. In *Conference on fairness, accountability and transparency* (pp. 119–133).
- Ellison, G., & Ellison, S. F. (2009). Search, obfuscation, and price elasticities on the internet. *Econometrica*, 77(2), 427–452.
- Friedler, S. A., Scheidegger, C., Venkatasubramanian, S., Choudhary, S., Hamilton, E. P., Roth, D. (2019). A comparative study of fairness-enhancing interventions in machine learning. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 329–338).
- Gardner, J., & Brooks, C. (2018). Student success prediction in moocs. *User Modeling and User-Adapted Interaction*, 28(2), 127–203.
- Gardner, J., Brooks, C., Baker, R. (2019). Evaluating the fairness of predictive student models through slicing analysis. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 225–234).
- Ghosh, A. K., Whipple, T. W., Bryan, G. A. (2001). Student trust and its antecedents in higher education. *The Journal of Higher Education*, 72(3), 322–340.
- Gordon, T., Moore, F. E., Shurtleff, D., Dawber, T. R. (1959). Some methodologic problems in the long-term study of cardiovascular disease: observations on the framingham study. *Journal of Chronic Diseases*, 10(3), 186–206.
- Grossman, H., & Grossman, S.H. (1994). Gender issues in education. ERIC.
- Hagenauer, G., & Volet, S. E. (2014). Teacher–student relationship at university: an important yet under-researched field. *Oxford Review of Education*, 40(3), 370–388.
- Hanson, S. (2008). *Swimming against the tide: African American girls and science education*. Philadelphia: Temple University Press.
- Hofstede, G. (1984). *Culture's consequences: international differences in work-related values* (Vol. 5). London: Sage.

- Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development*, 64(5), 923–938.
- Jacobsen, S. J., Mahoney, D. W., Redfield, M. M., Bailey, K. R., Burnett, J. J. C., Rodeheffer, R. J. (2004). Participation bias in a population-based echocardiography study. *Annals of Epidemiology*, 14(8), 579–584.
- Johnson, C., & Engelhard, G. (1992). Gender, academic achievement, and preferences for cooperative, competitive, and individualistic learning among AFRICAN-American adolescents. *The Journal of Psychology*, 126(4), 385–392.
- Johnson, D. W., & Johnson, F. P. (1991). *Joining together: group theory and group skills*. Englewood Cliffs: Prentice-Hall, Inc.
- Johnson, D. W., & Norem-Hebeisen, A. A. (1979). A measure of cooperative, competitive, and individualistic attitudes. *The Journal of Social Psychology*, 109(2), 253–261.
- Johnson, D. W., Johnson, R. T., Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works? *Change: The Magazine of Higher Learning*, 30(4), 26–35.
- Kim, K. K., Sankar, P., Wilson, M. D., Haynes, S. C. (2017). Factors affecting willingness to share electronic health data among California consumers. *BMC Medical Ethics*, 18(1), 25.
- Kruse, A., & Pongsajapan, R. (2012). Student-centered learning analytics. *CNDLS Thought Papers*, 1(9).
- Ladson-Billings, G. (1998). Just what is critical race theory and what's it doing in a nice field like education? *International Journal of Qualitative Studies in Education*, 11(1), 7–24.
- Land, R., & Bayne, S. (2005). Disciplinary power in online learning environments. In *Education in cyberspace* (p. 165).
- Li, W., Brooks, C., Schaub, F. (2019). The impact of student opt-out on educational predictive models. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 411–420).
- Liu, L.T., Dean, S., Rolf, E., Simchowitz, M., Hardt, M. (2018). Delayed impact of fair machine learning. arXiv:180304383.
- Long, T., Cummins, J., Waugh, M. (2017). Use of the flipped classroom instructional model in higher education: instructors' perspectives. *Journal of Computing in Higher Education*, 29(2), 179–200.
- Macfadyen, L. P., & Dawson, S. (2010). Mining lms data to develop an "early warning system" for educators: a proof of concept. *Computers & Education*, 54(2), 588–599.
- Madsen, M. C. (1967). Cooperative and competitive motivation of children in three Mexican sub-cultures. *Psychological Reports*, 20(3\_suppl), 1307–1320.
- Malhotra, N. K., Kim, S. S., Agarwal, J. (2004). Internet users' information privacy concerns (iupc): the construct, the scale, and a causal model. *Information Systems Research*, 15(4), 336–355.
- Matsuda, M.J. (2018). *Words that wound: critical race theory, assaultive speech, and the first amendment*. New York: Routledge.
- Matz, R. L., Koester, B. P., Fiorini, S., Grom, G., Shepard, L., Stangor, C. G., Weiner, B., McKay, T. A. (2017). Patterns of gendered performance differences in large introductory courses at five research universities. *AERA Open*, 3(4), 2332858417743754.
- McWhirter, E. H. (1997). Perceived barriers to education and career: ethnic and gender differences. *Journal of Vocational Behavior*, 50(1), 124–140.
- University of Michigan AA (2017). LA guiding principles. Center for Academic Innovation. <https://ai.umich.edu/learning-analytics-guiding-principles/>.
- University of Michigan AA (2020). CAI mission and principles. Center for Academic Innovation. <https://ai.umich.edu/our-mission-and-principles/>.
- University of Michigan AA (n.d.) Defining DEI. Office of Diversity, Equity, and Inclusion. <https://diversity.umich.edu/about/defining-dei/>.
- Moorman, P., Newman, B., Millikan, R., Tse, C. K., Sandler, D. (1999). Participation rates in a case-control study: the impact of age, race, and race of interviewer. *Annals of Epidemiology*, 9(3), 188–195.
- Moorman, P. G., Skinner, C. S., Evans, J. P., Newman, B., Sorenson, J. R., Calingaert, B., Susswein, L., Crankshaw, T. S., Hoyo, C., Schildkraut, J. M. (2004). Racial differences in enrolment in a cancer genetics registry. *Cancer Epidemiology and Prevention Biomarkers*, 13(8), 1349–1354.
- Niall, S. (2017). Consent and the GDPR: what approaches are universities taking? <https://analytics.jiscinvolve.org/wp/2017/06/30/consent-and-the-gdpr-what-approaches-are-universities-taking/>.
- Niles, F. S. (1995). Cultural differences in learning motivation and learning strategies: a comparison of overseas and Australian students at an Australian university. *International Journal of Intercultural Relations*, 19(3), 369–385.

- Nobles, W.W. (2006). *Seeking the Sakhu: foundational writings for an African psychology*. Chicago: Third World Press.
- Noguera, P.A. (2016). Race, education, and the pursuit of equity in the twenty-first century. In *Race, equity, and education* (pp. 3–23). Springer.
- Ocuppaugh, 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.
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450.
- Petersen, R. J. (2012). Policy dimensions of analytics in higher education. *Educause Review*, 47(4), 44–46.
- Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, 16(3), 9–20.
- Prizada, A., Yan, L. L., Garside, D. B., Schiffer, L., Dyer, A. R., Daviglius, M. L. (2004). Response rates to a questionnaire 26 years after baseline examination with minimal interim participant contact and baseline differences between respondents and nonrespondents. *American Journal of Epidemiology*, 159(1), 94–101.
- Polonetsky, J., & Jerome, J. (2014). Student data: trust, transparency, and the role of consent. Transparency, and the role of consent.
- Prinsloo, P., & Slade, S. (2014a). Educational triage in open distance learning: walking a moral tightrope. *International Review of Research in Open and Distributed Learning*, 15(4), 306–331.
- Prinsloo, P., & Slade, S. (2014b). Student data privacy and institutional accountability in an age of surveillance. In *Using data to improve higher education* (pp. 195–214). Brill Sense.
- Prinsloo, P., & Slade, S. (2017). Ethics and learning analytics: charting the (un) charted. In *Handbook of learning analytics*. SoLAR.
- Ramos, E., Lopes, C., Barros, H. (2004). Investigating the effect of nonparticipation using a population-based case-control study on myocardial infarction. *Annals of Epidemiology*, 14(6), 437–441.
- Rao, A., Schaub, F., Sadeh, N., Acquisti, A., Kang, R. (2016). Expecting the unexpected: understanding mismatched privacy expectations online. In *Twelfth symposium on usable privacy and security (SOUP) 2016* (pp. 77–96).
- Reidenberg, J. R., & Schaub, F. (2018). Achieving big data privacy in education. *Theory and Research in Education*, 16(3), 263–279.
- Richardson, R., Schultz, J. M., Crawford, K. (2019). Dirty data, bad predictions: how civil rights violations impact police data, predictive policing systems, and justice. *NYUL Review Online*, 94, 15.
- Roberts, L. D., Chang, V., Gibson, D. (2017). Ethical considerations in adopting a university-and system-wide approach to data and learning analytics. In *Big data and learning analytics in higher education* (pp. 89–108). Springer.
- Rubel, A., & Jones, K. M. (2016). Student privacy in learning analytics: an information ethics perspective. *The Information Society*, 32(2), 143–159.
- Saldaña, J. (2015). *The coding manual for qualitative researchers*. London: Sage.
- Schaub, F., & Cranor, L. (2020). Usable and useful privacy interfaces. In *An Introduction to Privacy for Technology Professionals* (pp. 175–229). International Association of Privacy Professionals.
- Schwartz, S. H. (1990). Individualism-collectivism: critique and proposed refinements. *Journal of Cross-Cultural Psychology*, 21(2), 139–157.
- Sedenberg, E., & Hoffmann, A.L. (2016). Recovering the history of informed consent for data science and internet industry research ethics. arXiv:160903266.
- Seifert, T. (2004). Understanding student motivation. *Educational Research*, 46(2), 137–149.
- Shore, C. (2008). Audit culture and illiberal governance: universities and the politics of accountability. *Anthropological Theory*, 8(3), 278–298.
- Shore, C., & Wright, S. (2003). Coercive accountability: the rise of audit culture in higher education. In *Audit cultures* (pp. 69–101). Routledge.
- Simon, F., Małgorzata, K., Beatriz, P. (2007). *Education and training policy no more failures ten steps to equity in education: ten steps to equity in education*. Paris: OECD Publishing.
- Slade, S., & Prinsloo, P. (2013). Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529.

- Slade, S., Prinsloo, P., Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 235–244).
- Sun, K., Mhaidli, A. H., Watel, S., Brooks, C. A., Schaub, F. (2019). It's my data! tensions among stakeholders of a learning analytics dashboard. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (p. 594). ACM.
- Taylor, D.M., & Moghaddam, F.M. (1994). *Theories of intergroup relations: international social psychological perspectives*. Westport: Greenwood Publishing Group.
- Tondeur, J., Pareja Roblin, N., van Braak, J., Voogt, J., Prestridge, S. (2017). Preparing beginning teachers for technology integration in education: ready for take-off? *Technology, Pedagogy and Education*, 26(2), 157–177.
- Triandis, H.C. (2018). *Individualism and collectivism*. New York: Routledge.
- Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology*, 74(1), 118.
- Triandis, H. C., Bontempo, R., Villareal, M. J., Asai, M., Lucca, N. (1988). Individualism and collectivism: cross-cultural perspectives on self-ingroup relationships. *Journal of Personality and Social Psychology*, 54(2), 323.
- Utz, C., Degeling, M., Fahl, S., Schaub, F., Holz, T. (2019). (un) informed consent: studying gdpr consent notices in the field. In *Proceedings of the 2019 ACM sigsac conference on computer and communications security* (pp. 973–990).
- Van Maele, D., Forsyth, P.B., Van Houtte, M. (2014). *Trust and school life*. Rotterdam: Springer.
- Yamaguchi, S. (1994). Collectivism among the Japanese: a perspective from the self. In *Cross-cultural research and methodology*.
- Zepke, N., & Leach, L. (2010). Beyond hard outcomes: 'soft' outcomes and engagement as student success. *Teaching in Higher Education*, 15(6), 661–673.
- Zhou, M. (2015). Moderating effect of self-determination in the relationship between big five personality and academic performance. *Personality and Individual Differences*, 86, 385–389.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Warren Li<sup>1</sup>  · Kaiwen Sun<sup>1</sup>  · Florian Schaub<sup>1</sup>  · Christopher Brooks<sup>1</sup>

Kaiwen Sun  
kwsun@umich.edu

Florian Schaub  
fschaub@umich.edu

Christopher Brooks  
brooksch@umich.edu

<sup>1</sup> School of Information, University of Michigan, 105 S. State St., Ann Arbor, MI 48104, USA