JOURNAL OF LEARNING ANALYTICS

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

Volume 12(3), 34–46. https://doi.org/10.18608/jla.2025.8713

Towards More Transparency in Learning Analytics: Sharing Information with University Students Increases their Awareness of Data Collection Practices

Stoo Sepp[1]

Abstract

As learning analytics practices become more commonplace in educational settings, student knowledge about the collection and use of their data becomes more of an interest. How students perceive the collection and use of their data has been researched for many years, with legitimate privacy and ethical concerns raised. While various guidelines, models, and frameworks have been proposed to address these concerns, the ways educational institutions practically address them by providing more information for increased transparency has yet to be widely investigated. The present study provides an initial investigation into the effectiveness of three different formats of data disclosure statements in a higher education setting. Participants were presented with one of three different formats from a fictional university: 1) a generic text, 2) a detailed text, or 3) an icon-based "nutrition label." Participants then completed a survey to assess their understanding and perceptions of data collection practices. Results suggest that regardless of format, participants demonstrated an increased understanding of these practices when the disclosure was not generic. Additionally, student acceptance of data collection and beliefs about sharing were unaffected by disclosure of any kind. This study provides initial evidence to inform learning analytics practices to address identified concerns from students around a lack of transparency. Universities and other institutions in the higher education sector may revisit their data disclosure methods and language to ensure that they are both accurate and transparent, so that students better understand data collection within the scope of their studies.

Notes for Practice

- Providing students with details on how their data is used for learning analytics purposes promotes their understanding and builds transparency.
- While the format of data disclosure statements (e.g., text or visual) may not play an important role, simply providing more detailed information about what data is collected can increase student understanding of LA practices.
- Including not only what is collected, but which parties have access to these data should also be included in disclosure statements.
- To support efforts of decolonization in educational research and practice, increased transparency and agency can serve to support informed consent and increased acceptance of LA in implementation.

Keywords: Analytics practice, transparency, ethics, privacy, disclosure

Submitted: 10/10/2024 — **Accepted:** 27/06/2025 — **Published:** 24/11/2025

Corresponding author [1]Email: stoo.sepp@une.edu.au Address: School of Education, University of New England, Elm Avenue, Armidale NSW 2351, Australia. ORCID iD: https://orcid.org/0000-0002-0509-4603

1. Introduction

As data-informed decision-making becomes more commonplace in educational settings, the need to understand how associated methods and approaches can inform teaching and learning practices becomes more important than ever. As part of this effort, critically examining how students are informed of and subsequently understand common practices around educational data collection and use is of the utmost importance. As Roberts et al. (2016) found, university student attitudes towards learning analytics suggested a general lack of knowledge about learning analytics itself, and identified student concerns regarding who might be accessing their information. This study pointed to a potential lack of transparency around who has access to student data and how it may be used in a higher educational setting. Additionally, Viberg et al. (2022) found that when students were asked to rate how they perceived their institution's implementation of learning analytics, their ideal practices in the area of privacy and ethics fell well below the perceived practices in reality. This aligns with Wang's (2018) discussion of big data in education, in which privacy, ethics, and by extension transparency, are noted as big concerns. Taken together, this points to both a desire from students for more transparency and institutional opportunities for more robust ethical practices in the implementation of learning analytics. The present study builds upon this work to investigate data disclosure statements as a means to operationalize transparency, and to assess its effects on student understanding of data collection practices. This paper first provides an overview of the learning analytics field, along with related theories of trust and ethical frameworks. This is followed by a report on the intervention, a discussion of findings, and implications for practice.

1.1. Learning Analytics and Data Collection

Learning analytics (LA) is a well-established area of research and practice in educational settings. It is generally defined as the measurement, collection, analysis, and reporting of data about learners and their contexts for purposes of understanding and optimizing learning and the environments in which it occurs (Long & Siemens, 2011).

In their review of LA methods and practices, Avella et al. (2016) noted that schools (and other educational organizations) should recognize the importance of using data-driven approaches in teaching and learning. This is especially true given the increase in the ubiquity of online teaching and learning since the COVID-19 pandemic and the increased use of technologies that provide data on student learning. When combined with traditional metrics such as grades, demographics, and other data, these insights can provide a more holistic picture of learning to better support students through their educational experiences. With these factors in mind, research on LA has grown to explore a number of areas including student retention and tracking of learning outcomes (de Freitas et al., 2015; Jayaprakash et al., 2014), in-class collaboration (Worsley & Ochoa, 2020), social presence in online learning (Norz et al., 2023), and many others.

Outside of the scope of research, however, data collection is pervasive and widespread in education, from early childhood settings through university. Aside from demographic and academic data (including grades), large amounts of data are collected as students engage with learning technologies, including learning management systems (LMSs), video systems, and other platforms. As outlined in the seminal paper by Lockyer et al. (2013) these data can be categorized into types of analytics such as *checkpoint analytics* (e.g., when students log in or access materials) and *process analytics* (e.g., how students engage with technology as they work towards a goal). In their review of data used in LA research focused on online learning, Kew and Tasir (2022) also point to different types of data that align with these types of analytics — behavioural, network, and level data. Common examples of such data can be derived from an analysis of log data from an LMS (see Nistor & Hernández-García, 2018), include mouse click (or finger tap) events when navigating an online course (behaviour), discussion forum posts (network), assignment submissions, and exam and quiz responses (level) — all timestamped and linked to every student in every class. While the scale of such data collection is commonplace and largely accepted outside of educational settings, throughout the lifespan of research in LA, the ethical and privacy considerations around the collection and use of student data in particular has been an important part of scholarly examination.

1.2. Ethics, Privacy, and Trust

Early discussions of LA included the need to consider privacy and ethics in this work, with Campbell et al. (2007) suggesting that privacy and data governance needed to be considered when engaging in LA work. In their discussion of surveillance and power in online learning environments, Land and Bayne (2005) noted that for the most part, staff — including learning designers, administrators, and instructors — have access to such data, whereas students do not. This raised an important aspect of LA practice — that student awareness and equity in the collection and use of educational data needed more attention, especially when considered in the larger context of ethics and privacy.

Slade and Prinsloo (2013) provided a meaningful discussion of student privacy in the context of LA, noting that when data is collected in other fields, such as health care, individuals must provide consent whereas in higher education this is not always the case. This raises concerns about the commodification of students and their data, and with universities and other organizations engaging with many third party, for-profit educational technology companies to fulfill their growing needs, what these companies do with their data becomes even more of a concern.

As Roberts et al. (2016) found in their exploration of student attitudes towards LA and big data, many students were simply not aware that their data was collected, and had different understandings of what data would even need to be collected. They also noted specific concerns around who would have access to their data as well as implications for bias towards them based on data collected, thus potentially "driving inequality" (p. 5). Further, in a study that explored the disconnect between university student expectations and realities of learning analytics, findings for 674 students enrolled in a university in the UK suggested that student expectations of how their data was collected, protected, and shared, did not match the reality of what their institution was actually doing (Tsai et al., 2020). It was also found that students expressed distrust towards third-party vendors that their university engaged with, citing concern over control over their data and the potential for it to be used for commercial purposes (Tsai et al., 2021).

Hoel et al. (2017) discuss this issue from a broader policy perspective, arguing that existing privacy frameworks such as the EU's General Data Protection Regulation (GDPR), should inform learning analytics practice. Appropriately, the GDPR includes a so-called right to be informed when personal data is collected, which supports the call of Joksimović et al. (2021) for "privacy-driven" learning analytics. Such an approach would focus on practices that limit data collection to specific purposes, protect private information, provide robust consent mechanisms, and ensure transparency to all students.

While most educational institutions are bound by policy, legislation, and other external factors to ensure transparency around information-related privacy for their students, it is not always easy to implement for administrators. Through this lens, Rubel and Jones (2016) discuss the delicate balancing act that many institutions engage in when weighing the net benefits of student data collection to support learning and continuous improvement with the potential loss of privacy, lack of informed consent, and compliance with legislation. To this end, researchers have proposed several ideas on how to address these issues.

1.3. Addressing Privacy Issues

Both Slade and Prinsloo (2013) and Pardo and Siemens (2014) provided guidance in the ethical use of learning analytics, with the former proposing principles around the moral use of data, student agency, and transparency, and the latter outlining transparency, student agency, access rights, accountability, and assessment of data use. At the centre of this guidance is transparency around what data is collected and for what purposes. Building upon this work, Drachsler and Greller (2016) proposed the DELICATE framework, which provides guidelines for the ethical development of learning analytics practices including recommendations around engagement, transparency, consent, and other areas. Mutimukwe et al. (2022) extended this to develop a model for investigating the privacy concerns of students regarding learning analytics. This model is based on the antecedents–privacy concerns–outcomes (APCO) model by Smith et al. (2011) and is rooted in relationships between perceived privacy risks, perceived privacy control, and prior trusting and non-self-disclosure behaviours. While previous theorizing and model development provide meaningful ways of conceptualizing privacy and ethical approachs to learning analytics for practitioners and researchers, the methods by which any guidelines are implemented remains largely underexplored. In their systematic review of privacy and data protection issues in learning analytics research, Liu and Khalil (2023) note a specific lack of empirical research into how to address such issues from a legal framework or technical perspective. While one aspect of many such solutions is through informed consent, this first requires providing students with information about what data is being collected.

Indeed, how students are informed of data collection is of key importance. As Whitman (2021) asserts, if students are to consent to their data being collected, they must first be accurately informed as to what is collected; however, the degree to which this happens in practice is not always clear or consistent. Prinsloo and Slade (2018) assert that in addressing consent, higher educational institutions "have tended to do so in a fairly generic and often opaque way" (p. 12). This again suggests that transparency is a key first step towards ensuring the ethical use of educational data by ensuring that students are aware of the details of data collection before they consent to it, thus supporting institutional efforts to protect their privacy and support their learning. Indeed Hoel and Chen (2018) propose that "openness and transparency are essential and should be an integral part of institutional policies" (p. 12). They also suggest that students should be consulted as to how data is collected and used. Based on this line of theorizing, it is clear that transparency in practice should be the goal of any robust LA implementation in higher education, with a capital "T" for transparency being the focus in Draschler and Greller's (2016) DELICATE framework.

1.4. Transparency and Data Disclosure Statements

Previous studies have indicated that university student expectations around the use of learning analytics data should be inclusive of permissions and transparency (Mutimukwe et al., 2022; Tsai et al., 2020) with Rubel and Jones (2016) advocating for transparency and clarity for students around what is collected, how it is collected, and the benefits that such collection would bring. To support this aim, Jones (2019) advocated for an educational version of the Platform for Privacy Preferences Project (P3P), developed by the World Wide Web Consortium (W3C), which would provide students with a detailed and transparent mechanism for informed consent as well as opt in/opt out mechanisms for educational data collection. When students were asked about privacy principles for three different learning analytics systems in a German university, results indicated that their willingness to share personal information differed based on the system used; when they had more agency over the data they provided, they were more accepting of the system itself (Ifenthaler & Schumacher, 2016). While students may have an opinion or a preference about the detailed data they choose to share in their private lives, in many instances this is not the case for universities and other educational organizations — students typically do not have a choice of what data is collected, and it varies greatly in terms of how the institution discloses which data is collected and for what purpose. While many studies explore student preferences and comfort with sharing their data, there is a need for more research investigating how universities disclose the data they collect to students, as well as the effectiveness of such efforts.

One avenue to support this research is through the use of data disclosure statements, which provide details of how personal information is collected, used, and stored. Such statements can be part of privacy policies or terms of reference in a commercial product, but mainly aim to provide concise and clear information to users about data collection practices. One aspect of how such disclosures could potentially be shared with students has to do with the format of the statement. Drawing upon research in cognitive and educational psychology, Cognitive Load Theory (Sweller et al., 2011) has identified a modality effect, which occurs when information presented across modalities — such as text, diagrams, video, and audio — can increase learning and retention of information that students are exposed to (for a review, see Ginns, 2005). For example, diagrams that are not self-explanatory should include short text descriptions; if they don't, students won't have the background information to derive meaning from the diagram alone (Sweller, 1994). Additionally, Moore and Fitz (1993) discussed how learning materials based on Gestalt visual design principles can support effective learning, specifically how grouping and integrating visual information with text and diagrams can support understanding. While the present study's focus is not on the design of learning materials per se, an investigation of multiple formats of data disclosure statements (namely the integration and grouping of visual information) can inform the design of potential solutions that concretely increase transparency in LA practice.

1.5. The Present Study

The present study provides an initial investigation into the effectiveness of data disclosure statements at an imaginary university in Australia. Given that transparency has consistently been identified as a priority for ensuring ethical practices within LA, participants were presented with one of three data disclosure formats: 1) a general format with vague statements about data collection (control condition), 2) a detailed text format (text condition) that outlined specific data collected, or 3) a "nutrition-label" format (label condition) that provided the same information using grouped icons and simple text labels. The study used a between-subjects design, with participants grouped into separate conditions based on the format of their disclosure presentation.

The primary research question was as follows:

How does the format of LA data disclosure statements affect university students' understanding and acceptance of data collection practices, motivations, and uses?

This question was designed to test the following hypothesis:

H1: Participants who are exposed to the detailed text and nutrition label formats will demonstrate increased confidence that specific data is collected compared to those in the control condition.

H2: Participants who are exposed to the nutrition label format will demonstrate increased confidence that specific data is collected compared to those in the text condition.

H3: Participants in the detailed text and nutrition label formats will demonstrate higher acceptance of data collection by their university compared to those in the control condition.

H4: Participants in the detailed text and nutrition label formats will report more conservative data collection practices with regards to who should have access to their data due to an increased awareness of what is being collected.

2. Methods

2.1. Participants

G*Power version 3.1.9.6 (Faul et al., 2007) was used to determine the minimum sample size needed to test the outlined hypotheses through an *a priori* power analysis. The required sample size to achieve 80% power for detecting a medium effect, at a significance of $\alpha = .05$, was $N = 159$. Participants thus numbered 187 university students, who were recruited through convenience sampling in online courses from two first-year psychology classes (through for-credit participation in research studies, thus reducing self-selection bias), and seven introductory educational technology classes (through voluntary participation) at a regional university in New South Wales, Australia. Inclusion criteria required fluency in English, and access to a laptop computer, smartphone, or tablet. While convenience sampling may lead to some inherent biases, participants were recruited from a diverse sample ranging in age from 16 years old to retirement age, located in all corners of the country, and representing the exact target population with a vested interest in the topic — students whose data is collected for online learning purposes. Participants who volunteered to participate were randomly assigned to one of three conditions: control, text, or label.

2.2. Materials and Procedure

Students completed an online survey using the Qualtrics survey software. Upon starting the survey, students were asked for basic demographic data, including gender and age range. After entering basic demographic data, each participant was exposed to a data disclosure statement presented in one of three formats, depending on their assigned condition. Data disclosure statements were developed by the researcher and presented to participants as coming from the fictional "University of Australia." Figure 1 shows a generic, vague disclosure statement (control condition) that indicates the participant would be required to provide personal information, without providing much detail. Figure 2 shows a more specific text-based disclosure statement (text condition) that provides clearer details of exactly what data is collected and why. Figure 3 shows the same information as Figure 2, though presented in a “nutrition label” format (label condition), with categorized and icon-based information designed to mimic data collection disclosure statements from mobile application stores.

To progress through your studies, you will be required to provide personal information to the University of Australia. “Personal information” is as defined in the Privacy and Personal Information Act 1998 (NSW), and includes name and contact information and details of previous study or experience. This information will be used to for:

- admission, enrolment and study progression (as applicable);
- sending you information regarding the University of Australia;
- various other administrative, security, academic and statistical purposes that assist the University to achieve its objects and functions.

Figure 1. Generic data collection disclosure statement.

Data Used to Improve your Learning Experience

The following data is collected by educational systems and may be used by instructional and support staff to improve your learning experiences and the experiences of other students.

The university collects a number of data-points to improve teaching and learning. These include but are not limited to: Your name, location and IP Address; the times you access your online class space; how many times and when you access learning materials and resources; how many times and when you access links; how many times you view and post to forums, including specific word counts; when and what you manually mark as complete; your assignment marks; how many times you submitted assignments and; how many times you viewed feedback.

Figure 2. Text-based data collection disclosure statement.

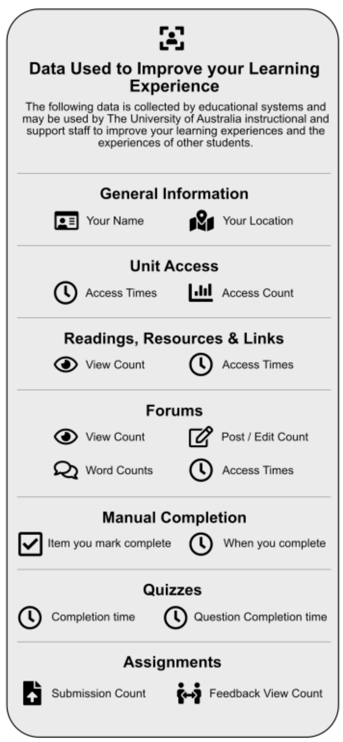

Figure 3. Data collection disclosure statements in the form of a “nutrition label.”

2.3. Questions About Data Collection and Perceptions

After participants were exposed to one of the three data collection disclosure statements, depending on their assigned condition, they were asked a series of questions related to their understanding of which personal data points are collected, focusing on the context of online learning. These were presented in a Likert format ranging from “I’m sure they don’t collect it” to “I’m sure they collect it.” They were then asked how their peers might understand these same practices. As Heath (2014) noted, drivers for LA adoption are rooted in the acceptance, understanding, and expected benefits of such systems. To investigate these elements as they relate to data disclosure statements, participants were asked questions relating to these issues. Finally, building upon previous investigations of student perceptions of data collection for LA work (Tsai et al., 2020) participants were then asked questions relating to their understanding of who *does* have access to their data, and who *should* have access to their data. Again, these were formatted as Likert questions, ranging from "strongly disagree" to "strongly agree." Lastly, participants were asked *when* they believe students should be informed of the details of data collected about them.

2.4. Data Analysis

All data was screened for normality (using Shapiro-Wilk tests) and homogeneity of variance (using Levene's test). Data related to each hypothesis was then analyzed and grouped by condition, with aggregated means examined through appropriate analysis of variance tests, with aligned post-hoc tests conducted as needed.

3. Results and Discussion

A summary of results for each data collection disclosure condition is presented in Table 1.

Table 1. Confidence in which Data Points Collected by University

Dependent Variable	Disclosure Format Condition		
	Control (n=48)	Text (n=66)	Label (n=73)
My name	4.65 (0.86)	4.32 (1.10)	4.73 (0.63)
Details of device I'm using	2.63 (1.33)	3.36 (1.25)	3.03 (1.24)
My location	3.50 (1.34)	3.74 (1.22)	3.70 (1.43)
My IP Address	3.33 (1.40)	4.26 (1.13)	3.18 (1.34)
When I access my online class	3.29 (1.37)	4.59 (0.74)	4.38 (0.89)
How often I access my online class	3.48 (1.20)	4.47 (0.85)	4.14 (0.99)
When I view learning materials	3.02 (1.33)	4.21 (0.90)	4.00 (1.14)
How often I view learning materials	3.15 (1.25)	4.14 (0.97)	3.77 (1.16)
How often I post to forums	3.40 (1.25)	3.95 (1.01)	3.73 (1.11)
When I view forum posts	2.81 (1.27)	3.89 (1.14)	3.75 (1.15)
How often I view forum posts	2.96 (1.18)	3.79 (1.07)	3.37 (1.24)
The word count of my forum posts	2.88 (1.28)	3.88 (1.27)	3.55 (1.04)
When I mark an item complete	3.21 (1.29)	3.98 (1.16)	3.95 (1.01)
Items I mark complete	2.94 (1.31)	3.82 (1.30)	4.10 (1.02)
When I submit assignments	4.19 (1.18)	4.56 (0.90)	4.81 (0.46)
Assignment submission counts	3.94 (1.21)	4.30 (0.99)	4.56 (0.69)
How long it takes me to complete quizzes	3.56 (1.34)	4.05 (1.21)	4.14 (1.03)
How long it takes me to complete quiz questions	3.25 (1.26)	3.53 (1.19)	3.63 (1.14)

Note: This table presents means and standard deviations from Likert responses ranging from "I'm sure they don't collect it" to "I'm sure they collect it" represented by a numerical scale of 1–5.

3.1. Confidence That Data Was Collected

Table 1 shows results for each data point presented to participants. The mean for participant confidence that data was collected was calculated across all data points to serve as an aggregated indicator of overall understanding of data collection. This resulted in values on a scale of one ("I'm sure they don't collect it") to five ("I'm sure they collect it") of 3.34 for the control condition (SD = 0.51), 4.05 for the text condition (SD = 0.58) and 3.91 for the label condition (SD = 0.56). Given that the responses for confidence in data points being collected were collected using Likert-style questions, these data are considered intervals for the purpose of analysis. To determine whether the format of presentation resulted in differences between each group, testing the assumptions of an ANOVA were undertaken based on the individual data points collected. Given the small sample sizes, a Shapiro-Wilk test for normality was conducted for each group, which indicated that the means were normally distributed for the control (W = 0.97, p = 0.28), text (W = 0.97, p = 0.12), and label (W = 0.96, p = 0.04) conditions. A Levene's test was then conducted and the homogeneity of variance was found to be equal between groups ($F(2,184) = 7.231$, p <.001). Given that the assumptions were satisfied, an ANOVA was conducted to compare the effect of presentation format on mean confidence that data was collected. There was a significant difference between groups, $F(2,184) = 16.38$, p <.001. A Tukey's HSD post-hoc test revealed that the text and label conditions reported significantly higher confidence that specific data points were collected at $p < 0.001$ when compared to the control condition; the text and label conditions were not significantly different from each other (p = 0.492).

Given the unequal sample sizes across conditions, an additional non-parametric ANOVA was conducted using the Kruskal-Wallis H test to ensure consistent reults. The test indicated a significant difference between conditions ($\chi^2(2) = 21.458$, $p <.001$). A pairwise comparison conducted using a Wilcoxon signed-rank test again revealed that participants in both text ($p <.001$) and label ($p <.001$) conditions reported significantly higher confidence that specific data points were collected when compared to the control condition. The same test revealed non-significant findings when comparing text and label conditions ($p =.114$). *H1* tested whether participants exposed to the detailed text and nutrition label style disclosure statements would demonstrate higher confidence that specific data points were collected as compared to the generic disclosure (control). In this case *H1* was supported. *H2* tested whether participants exposed to the nutrition labels would report higher confidence that specific data points were collected compared to the detailed text disclosure, which was not supported.

3.2. Acceptance of Data Collection (H3)

A summary of results for how participants understood data collection, its benefits, and their acceptance of such practices is presented in Table 2. Questions specifically focus on their subjective understanding of why data is collected, the benefits for students, and their comfort and concern with university practices.

Table 2. Acceptance of Data Collection Procedures

Dependent Variable	Disclosure Format Condition		
	Control (n=48)	Text (n=66)	Label (n=73)
I have a clear understanding of what educational data related to me is collected.	2.94 (1.29)	3.32 (0.95)	3.15 (1.10)
I have a clear understanding of why this educational data is collected.	3.10 (1.37)	3.18 (1.16)	3.18 (1.12)
I understand the benefits that the collection of these data can bring to myself as an individual.	3.13 (1.23)	3.11 (1.07)	3.14 (1.06)
I understand the benefits that the collection of these data can bring to my class.	3.27 (1.30)	3.18 (1.08)	3.33 (1.09)
I understand the benefits that the collection of these data can bring to all students at the university.	3.29 (1.32)	3.33 (1.11)	3.37 (1.06)
I trust the university based on my understanding of educational data collection.	3.50 (1.07)	3.29 (1.29)	3.37 (1.06)
I understand how the University of Australia has ensured the ethical use of my educational data.	2.90 (1.31)	2.74 (1.18)	2.96 (1.14)
I would feel more comfortable if the University of Australia was clearer about how my educational data is used in an ethical manner.	4.10 (0.95)	4.09 (0.89)	3.92 (0.98)
I would feel more comfortable if staff had ethical use statements.	3.81 (1.18)	3.58 (1.08)	3.41 (1.14)
I am concerned that too much education data related to me is being collected.	3.21 (1.25)	3.26 (1.06)	2.93 (1.11)
I would like the choice to opt in or out of the collection of these data.	3.77 (1.19)	3.73 (1.18)	3.66 (1.15)

Note: This table presents means and standard deviations from Likert responses ranging from "strongly disagree" to "strongly agree" represented by a numerical scale of 1–5.

As with the previous analysis, the mean for participant acceptance of data collection procedures was calculated across all data points, again serving as an aggregated indicator of overall acceptance of data collection procedures. This resulted in values on a scale of one ("strongly disagree") to five ("strongly agree") of 4.12 for the control condition ($SD = 1.14$), 4.27 for the text condition ($SD = 1.15$), and 4.27 for the label condition ($SD = 0.98$). Given that the responses for acceptance of data collection were collected using Likert-style questions, these data are considered intervals for the purpose of analysis. To determine whether the format of presentation resulted in differences between each group, testing of the assumptions of an ANOVA were undertaken based on the individual data points collected. A Shapiro-Wilk test for normality was conducted for each group, which indicated that the means were not normally distributed for the control ($W = 0.74$, $p <.001$), text ($W = 0.73$, $p <.001$), or label ($W = 0.81$, $p <.001$) conditions. A Levene's test was then conducted and the homogeneity of variance was found to be unequal between groups ($F(2,184) = 0.742$, $p =.478$). Given that the assumptions for an ANOVA were not satisfied, a non-parametric Kruskal-Wallis H test revealed no significant differences between groups ($\chi^2(2) = 2.28$, $p =.321$). H3, which tested whether participants exposed to text and label disclosure formats demonstrated higher acceptance of data collection practices, was not supported.

Aside from hypothesis testing, Figure 4 presents a visual examination of results, which provide additional insights for each dependent variable. A visual inspection of this plot suggests strong support from participants for the ability to access and download their own data, for more clarity in ethical data collection practices, and for the ability to opt in or out of educational data collection.

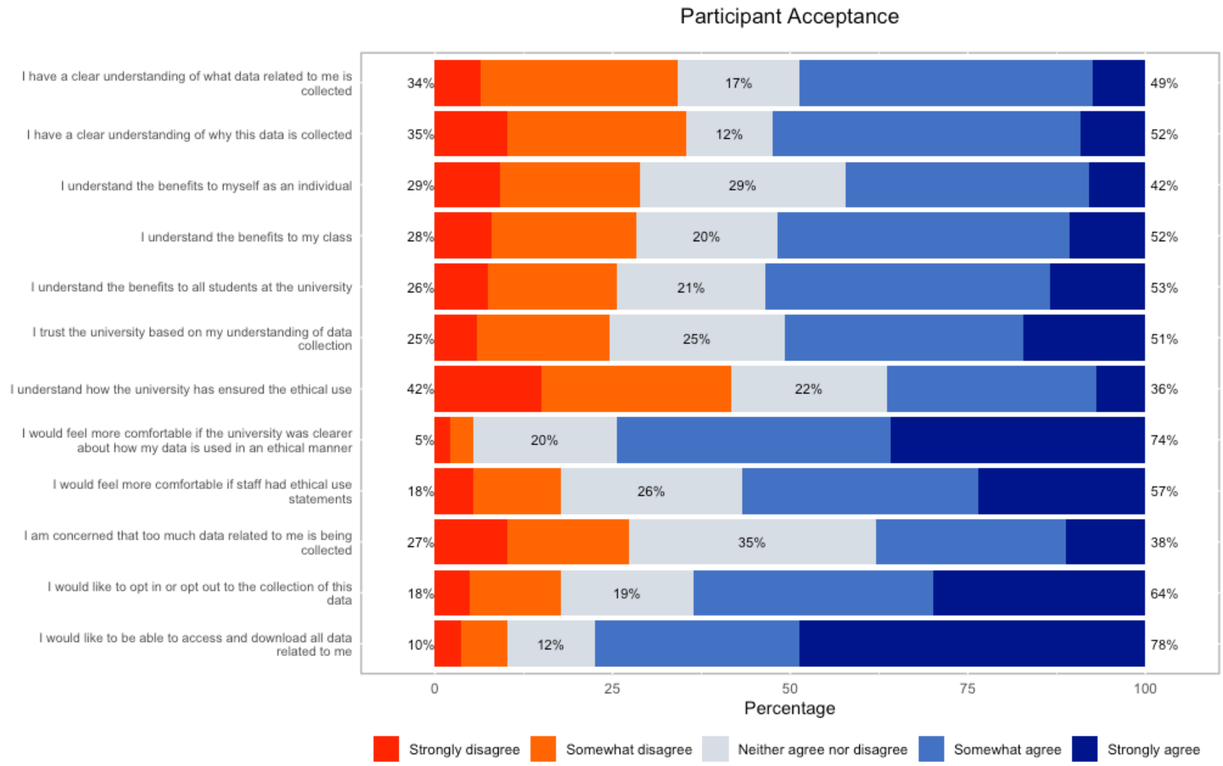

Figure 4. Likert plot showing overall participant understanding and acceptance of data collection.

3.3. Access to LA Data (H4)

A summary of results indicating participant opinions about who should have access to their educational data is presented in Table 3.

Table 3. Opinions on Data Access

Party	Disclosure Format Condition		
	Control (n=48)	Text (n=66)	Label (n=73)
My instructor	3.46 (0.62)	3.26 (0.81)	3.27 (0.80)
Instructional support staff	3.02 (0.70)	2.82 (1.04)	2.84 (0.90)
Marking/grading staff	2.85 (0.92)	2.52 (1.10)	2.60 (0.98)
IT staff	3.15 (0.77)	2.92 (0.95)	2.74 (0.96)
Student support staff	2.65 (0.93)	2.53 (1.00)	2.66 (1.03)
Program or degree coordinator	3.35 (0.70)	3.15 (0.81)	3.04 (0.81)
Academic leadership staff	2.83 (0.93)	2.76 (0.96)	2.71 (0.90)
Marketing staff	2.31 (1.17)	1.95 (1.01)	1.89 (0.97)
Government organizations	2.04 (1.03)	1.82 (0.98)	1.78 (0.99)
Private companies	1.25 (0.56)	1.30 (0.66)	1.26 (0.58)

Note: This table presents means and standard deviations from Likert responses including "Definitely shouldn't have access," "I don't care if they have access," "Should have limited access as needed" and "Definitely should have access" represented by a numerical scale of 1–4.

Table 3 shows results for each party that may be privy to participants' educational data, and the opinions of participants regarding who should have access. The mean was calculated for all data points collected as an aggregated indicator for overall opinions on data access on a scale of one ("Definitely shouldn't have access") to four ("Definitely should have access"). Results were 2.69 for the control condition (SD = 0.52), 2.50 for the text condition (SD = 0.60), and 2.48 for the label condition (SD = 0.48). Again, for the purposes of this analysis, the means of these Likert responses are considered intervals. A test of normality using a Shaprio-Wilk test indicated normally distributed data for the control (W=0.98, p <.011), text (W = 0.97,

p =.323), and label (W = 0.95, p =.011) conditions, while a Levene's test resulted in non-equal variances between groups ($F(2,184) = 2.705$, $p = 0.069$). With assumptions for a traditional ANOVA not met, a resulting Krusal-Wallis test suggested non-significant differences between groups in terms of who should have access to the participants' educational data (χ 2(2) = 5.578, $p = 0.061$). H4 tested if participants exposed to more detailed data disclosures in the text and label conditions were more conservative in their thoughts around who should have access to their data, compared to the control condition. Overall, H4 was not supported.

Beyond hypothesis testing, Figure 5 shows a plot presenting Likert responses about who participants thought had access to their data (which was not covered in data disclosure conditions). Figure 6 then presents Likert responses regarding who participants thought *should* have access to their data.

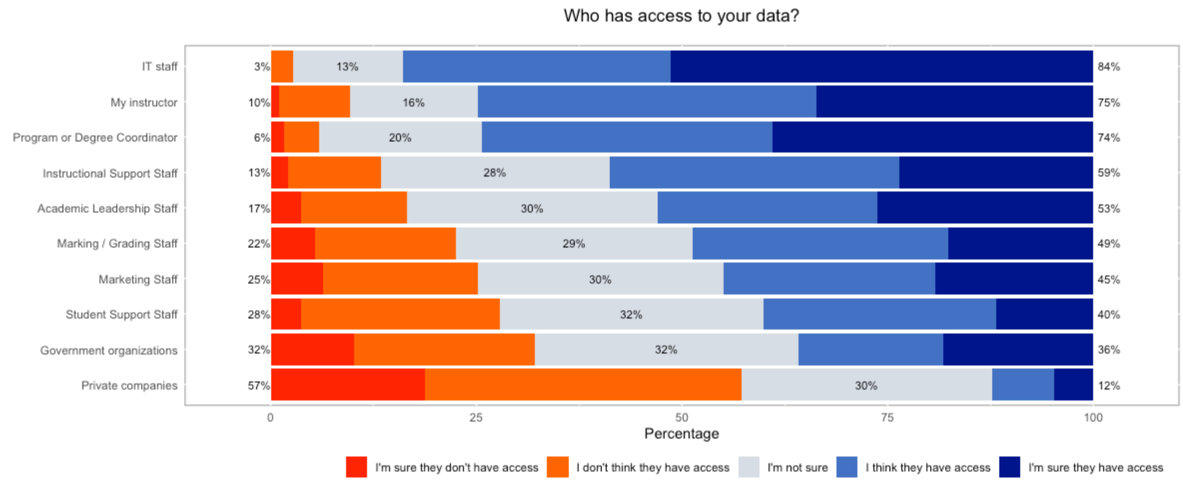

Figure 5. Likert plot showing overall participant perception of who has access to their data.

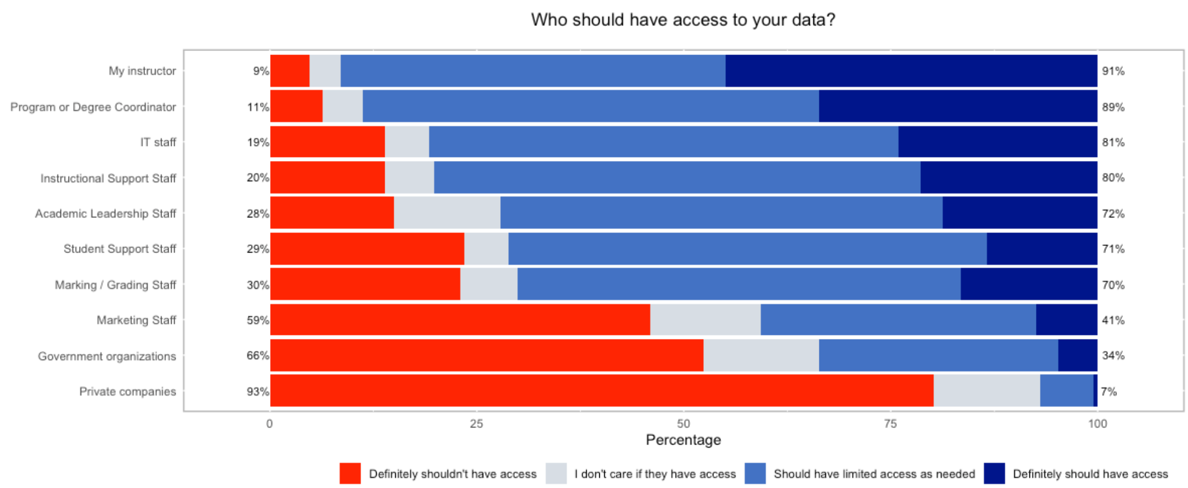

Figure 6. Likert plot showing overall participant perception of who should have access to their data.

Interestingly, participants reported with very high confidence that IT staff should have access to educational data, with academic and other support staff rounding out the results, however participants in both cases thought that private companies did not have, nor should have access to their educational data.

3.4. Supplemental Data

While not directly linked to research questions or hypotheses, participants were asked as part of this study about when they believe students should be informed of the details of data collection, with the option to check multiple answers. Indeed, Ferguson (2012) asserted that when discussing the challenges of learning analytics research there "is no agreed method for researchers to obtain informed and ongoing consent to the use of data." This also applies in practice, given that students may not be asked to provide consent when new technologies are adopted in the middle of their university degree, or when existing vendors change their data collection practices or processes. Figure 7 presents the results, showing that over 50% of respondents thought that students should be informed at multiple "contact points," ranging from initial admission to the university to multiple times throughout a semester or term.

ISSN 1929-7750 (online).

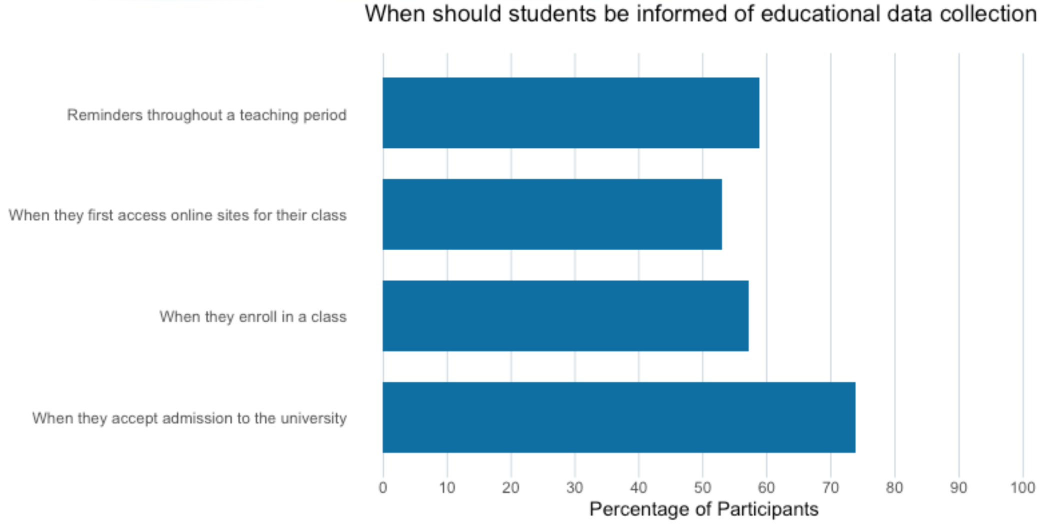

Figure 7. Plot showing participant opinions of when students should be informed of data collection.

3.5. Discussion

Building on previous studies in the area of ethics and data collection in LA, this study provided initial insights into how increased transparency in LA practice and desired student understanding of data collection in higher education may be implemented. This study extends findings investigating student awareness and trust in educational settings (Slade et al., 2019; Tsai et al., 2020) and provides initial contributions to our understanding of how increased transparency affects student perceptions of data collection for learning analytics purposes.

Overall, when participants were informed of what specific data is collected about them, they demonstrated a higher confidence that such data was collected, regardless of the format (*H1*). While this result seems incredibly obvious, the technology-based assumptions discussed by Prinsloo et al. (2022) — that students are aware and have agency over data collection — quickly fall apart, providing evidence that students may not be aware of what data is collected about them, or who has access to it. Interestingly, the various formats of disclosure used in this study did not have any significant differential effects (*H2*), suggesting that simply presenting a detailed disclosure in either text or visual formats provided students with the information they needed to increase their awareness of data collection practices. When considering this result through the lens of educational psychology research, this may be attributed to the lack of inherent meaning in the icons used such that short text passages or visual icons with text are equally effective in providing necessary information to students. Further, participants' level of acceptance for data collection, and their support for different parties accessing their data was not affected by how disclosures were presented (*H3 and H4*). Given the ubiquity of data collection that occurs in the private lives of students for activities ranging from social media to online shopping, data collection may simply be an accepted "feature" of online life; however, the details and awareness of what data is collected and who has access to it is nevertheless important when working towards more ethical LA practices. Educational organizations should thus work to provide more transparency around the details of what data is collected about students as well as who has access to it. While these practices are normally articulated through data governance efforts, the sharing of this information may not be consistently or clearly communicated to students.

As Prinsloo et al. (2022) importantly raise, considering consent in the era of such ubiquitous data collection is of the utmost importance, especially when it comes to the power dynamics at play and the historical effects of colonialism that continue to affect historically marginalized populations and Indigenous students. Though LA practices are predominantly justified to support the improvement of teaching and learning, practitioners and researchers should be mindful of the history behind uninformed consent and the commodification of individuals' information without their consent, particularly in marginalized and Indigenous populations (Held, 2019; Tuck & Yang, 2014). For this reason, ethical practices around disclosure and consent serve not only to support decolonization in research, but also in any practice involving the collection of personal information.

In service of these efforts, the role that private companies play in education thus presents a unique challenge. Results showed that participants didn't think that private companies did, nor should have access to their data, when in most educational settings this is usually a given under the guise of "product improvement." As Tsai et al. (2020) found, students have high expectations for their university to obtain consent and secure data when sharing it with third parties, especially given that students expressed distrust towards these parties (Tsai et al., 2021). Whether it be LMSs, web conferencing and video-chat platforms, or even generative AI technologies, these vendors definitely do have access and will in the foreseeable future due to their relationships with educational institutions. If students are not aware of this reality, educational institutions should be mindful of how students are informed about who has access to their data both from within the organization, and from without, including listing the names and locations of any third-party vendors and private companies that work with the organization. Doing so will move closer to a transparent and more trusting relationship between students and their schools and universities, while also aligning with decolonization efforts within research and practice.

While this study provides initial insights into the effects of exposing students to different means of disclosing data collection procedures in higher education, it did include some limitations. The sample was primary recruited from a single rural university in Australia, and was limited to voluntary and self-reported responses. To extend this work, future studies may wish to further validate findings across different populations and contexts, with more participants across international universities. Additionally, further validating and measuring the lasting effectiveness of data disclosure statements and formats may also inform data literacy across contexts. By integrating previous measures such as the control over data (COD) scale and sharing of data (SOD) questionnaire employed by Ifenthaler and Schumacher (2016), more details on transparency and willingness to share could further inform the design of effective data disclosure strategies. The first step, however, is simply for institutions to take action. As Nichols (2024) highlights in a case study on developing an ethical position for LA implementation, research, policy, data governance, and previous frameworks and guidelines can all assist in ensuring data collection and use is done in an ethically robust manner. Along with the findings of this study, future research that focuses on practical methods to increase transparency and student agency can also contribute to actionable ethical frameworks, moving us towards more open and transparent LA policies and practices.

4. Conclusion

This paper reported on an initial investigation into a practical implementation of increased transparency for LA practice in higher education. Results indicated that simply providing more information to students increases their confidence in and understanding of what specific data points are collected about them, with the format — whether text or visual — not having an effect. Further results indicated that participants may have an inaccurate understanding of who has access to their data outside of their immediate instructional support staff, thus supporting the need for more transparency within each educational context. As supported by previous research, it continues to be clear that the privacy and ethical concerns of students should always be respected and made an integral part of the conversation when considering the adoption and implementation of LA systems for educational use. It is clear that increased transparency can be achieved; however, further research is needed to provide a more holistic means of addressing privacy and ethical concerns so that students have more agency over their data, trust their education providers, and have a clear understanding of who has access to their data and for what purposes.

Declaration of Conflicting Interest

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author declared no financial support for the research, authorship, and/or publication of this article.

References

Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning, 20*(2). https://doi.org/10.24059/olj.v20i2.790

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics. *EDUCAUSE Review, 42*(4), 40–57. https://er.educause.edu/-/media/files/article-downloads/erm0742.pdf

de Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., Dunwell, I., & Arnab, S. (2015). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology, 46*(6), 1175–1188. https://doi.org/10.1111/bjet.12212

Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue: A checklist for trusted learning analytics. In *LAK '16: Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 89–98). ACM Press. https://doi.org/10.1145/2883851.2883893

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304–317. https://doi.org/10.1504/ijtel.2012.051816

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*(4), 313–331. https://doi.org/10.1016/j.learninstruc.2005.07.001

Heath, J. (2014). Contemporary privacy theory contributions to learning analytics. *Journal of Learning Analytics, 1*(1), 140–149. https://doi.org/10.18608/jla.2014.11.8

Held, M. B. E. (2019). Decolonizing research paradigms in the context of settler colonialism: An unsettling, mutual, and collaborative effort. *International Journal of Qualitative Methods, 18*. https://doi.org/10.1177/1609406918821574

Hoel, T., & Chen, W. (2018). Privacy and data protection in learning analytics should be motivated by an educational maxim: Towards a proposal. *Research and Practice in Technology Enhanced Learning, 13*(1), Article 20. https://doi.org/10.1186/s41039-018-0086-8

Hoel, T., Griffiths, D., & Chen, W. (2017). The influence of data protection and privacy frameworks on the design of learning analytics systems. In *LAK '17: Proceedings of the seventh international learning analytics & knowledge conference* (pp. 243–252). ACM Press. https://doi.org/10.1145/3027385.3027414

Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development, 64*(5), 923–938. https://doi.org/10.1007/s11423-016-9477-y

Jayaprakash, S. M., Moody, E. W., Lauría, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics, 1*(1), 6–47. https://doi.org/10.18608/jla.2014.11.3

Joksimović, S., Marshall, R., Rakotoarivelo, T., Ladjal, D., Zhan, C., & Pardo, A. (2021). Privacy-driven learning analytics. In E. McKay (Ed.), *Manage your own learning analytics: Implement a Rasch modelling approach* (pp. 1–22). Springer Cham. https://doi.org/10.1007/978-3-030-86316-6_1

Jones, K. M. L. (2019). Learning analytics and higher education: A proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. *International Journal of Educational Technology in Higher Education, 16*(1), Article 24. https://doi.org/10.1186/s41239-019-0155-0

Kew, S. N., & Tasir, Z. (2022). Learning analytics in online learning environment: A systematic review on the focuses and the types of student-related analytics data. *Technology, Knowledge and Learning, 27*(2), 405–427. https://doi.org/10.1007/s10758-021-09541-2

Land, R., & Bayne, S. (2005). Screen or monitor? Issues of surveillance and disciplinary power in online learning environments. In R. Land & S. Bayne (Eds.), *Education in cyberspace* (pp. 164–177). Routledge. https://doi.org/10.4324/9780203391068-21

Liu, Q., & Khalil, M. (2023). Understanding privacy and data protection issues in learning analytics using a systematic review. *British Journal of Educational Technology, 54*(6), 1715–1747. https://doi.org/10.1111/bjet.13388

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist, 57*(10), 1439–1459. https://doi.org/10.1177/0002764213479367

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 30–40. https://er.educause.edu/-/media/files/article-downloads/erm1151.pdf

Moore, P., & Fitz, C. (1993). Gestalt theory and instructional design. *Journal of Technical Writing and Communication, 23*(2), 137–157. https://doi.org/10.2190/g748-by68-l83t-x02j

Mutimukwe, C., Viberg, O., Oberg, L.-M., & Cerratto-Pargman, T. (2022). Students' privacy concerns in learning analytics: Model development. *British Journal of Educational Technology, 53*(4), 932–951. https://doi.org/10.1111/bjet.13234

Nichols, M. (2024). Development of an approved learning analytics ethics position. *Open Learning: The Journal of Open, Distance and e-Learning, 39*(3), 212–225. https://doi.org/10.1080/02680513.2021.1986376

Nistor, N., & Hernández-García, Á. (2018). What types of data are used in learning analytics? An overview of six cases. *Computers in Human Behavior, 89*, 335–338. https://doi.org/10.1016/j.chb.2018.07.038

Norz, L.-M., Dornauer, V., Hackl, W. O., & Ammenwerth, E. (2023). Measuring social presence in online-based learning: An exploratory path analysis using log data and social network analysis. *The Internet and Higher Education, 56*, Article 100894. https://doi.org/10.1016/j.iheduc.2022.100894

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology, 45*(3), 438–450. https://doi.org/10.1111/bjet.12152

Prinsloo, P., & Slade, S. (2018). Student consent in learning analytics: The devil in the details? In J. Lester, C. Klein, A. Johri, & H. Rangwala (Eds.), *Learning analytics in higher education* (pp. 118–139). Routledge. https://doi.org/10.4324/9780203731864

Prinsloo, P., Slade, S., & Khalil, M. (2022). The answer is (not only) technological: Considering student data privacy in learning analytics. *British Journal of Educational Technology, 53*(4), 876–893. https://doi.org/10.1111/bjet.13216

Roberts, L. D., Howell, J. A., Seaman, K., & Gibson, D. C. (2016). Student attitudes toward learning analytics in higher education: "The Fitbit version of the learning world." *Frontiers in Psychology, 7*, Article 1959. https://doi.org/10.3389/fpsyg.2016.01959

Rubel, A., & Jones, K. M. L. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society, 32*(2), 143–159. https://doi.org/10.1080/01972243.2016.1130502

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. In *LAK19: Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 235–244). ACM Press. https://doi.org/10.1145/3303772.3303796

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly*, *35*(4), 989–1015. https://doi.org/10.2307/41409970

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Tsai, Y.-S., Whitelock-Wainwright, A., & Gašević, D. (2020). The privacy paradox and its implications for learning analytics. In *LAK '20: Proceedings of the tenth international conference on learning analytics & knowledge* (pp. 230–239). ACM Press. https://doi.org/10.1145/3375462.3375536

Tsai, Y.-S., Whitelock-Wainwright, A., & Gašević, D. (2021). More than figures on your laptop: (Dis)trustful implementation of learning analytics. *Journal of Learning Analytics, 8*(3), 81–100. https://doi.org/10.18608/jla.2021.7379

Tuck, E., & Yang, K. W. (2014). R-words: Refusing research. In D. Paris & M. T. Winn (Eds.), *Humanizing research: Decolonizing qualitative inquiry with youth and communities* (pp. 223–248). SAGE Publications. https://doi.org/10.4135/9781544329611.n12

Viberg, O., Engström, L., Saqr, M., & Hrastinski, S. (2022). Exploring students' expectations of learning analytics: A person-centered approach. *Education and Information Technologies*, *27*(6), 8561–8581. https://doi.org/10.1007/s10639-022-10980-2

Wang, F., Li, W., Mayer, R. E., & Liu, H. (2018). Animated pedagogical agents as aids in multimedia learning: Effects on eye-fixations during learning and learning outcomes. *Journal of Educational Psychology*, *110*(2), 250–268. https://doi.org/10.1037/edu0000221

Whitman, M. (2021). Modeling ethics: Approaches to data creep in higher education. *Science and Engineering Ethics*, *27*(6), Article 71. https://doi.org/10.1007/s11948-021-00346-1

Worsley, M., & Ochoa, X. (2020). Towards collaboration literacy development through multimodal learning analytics. In *Companion proceedings of the tenth international conference on learning analytics & knowledge (LAK20)* (pp. 585–595). Society for Learning Analytics Research.