

Ethical Considerations and Student Perceptions of Engagement Data in Learning Analytics

Ayaz Karimov
University of Jyväskylä
ayazkarimov3@gmail.com

Mirka Saarlea
University of Jyväskylä
mirka.saarela@jyu.fi

Samir Aliyev
University of Jyväskylä
sa@swisscyberinstitute.com

Ryan Baker
University of Pennsylvania
ryanshaunbaker@gmail.com

Abstract

The ethical use of engagement data in online education is a growing concern as institutions increasingly rely on learning analytics. This study explores students' perceptions of engagement data collection and usage by focusing on their attitudes towards privacy and data management. We conducted a survey among students (n=108) who participated in online education to understand their views on data collection practices, privacy concerns, and preferences for data handling. The results demonstrate that while many students are comfortable with their engagement data being used for personal and instructor dashboards, significant concerns remain about privacy, particularly with the collection of facial expressions and chat participation data. Students emphasized the importance of transparency and control over their data and they highlighted the need for clear communication and consent processes. These findings illustrate the necessity for ethical data practices that ensure students feel secure and informed about how their engagement data is utilized.

Keywords: data ethics, engagement data, ethical concerns, online education, student engagement

1. Introduction

The integration of learning analytics in educational environments has the potential to revolutionize the way educators understand and enhance student engagement and performance (Karimov et al., 2023). Engagement data refers to the various metrics and information collected about a student's interactions within an online learning environment. This can include data such as login times, frequency and duration of interaction with

course materials, participation in discussions, and even more sophisticated metrics like facial expressions or physiological responses. By utilizing data from various sources, including digital footprints and interaction logs, teachers can gain insights into students' learning behaviors. One of the most prominent applications of learning analytics is in the form of dashboards that provide real-time data to teachers, aiming to help them tailor their instructional strategies (Saarela & Kärkkäinen, 2017).

However, the increasing use of engagement data in online education raises significant ethical concerns, particularly regarding student privacy. The potential misuse of personal and behavioral data for performance evaluation, without clear consent or understanding from students, presents a critical challenge in ensuring ethical data practices. One of the primary ethical issues surrounding using engagement data in learning analytics concerns around student privacy. In a classroom or virtual learning environment, data on student engagement can include a wide range of information, from login times, participation in discussions, and submission of assignments to facial expression and physiological measurements (Tempelaar et al., 2020). As pointed out by Slade et al. (2019), it is unclear how aware and how comfortable students are about the scope and use of their personal and behavioral data for learning analytics, and according to the review by Viberg et al. (2018), only 18% of learning analytics studies even acknowledge "ethics" or "privacy".

This study aims to explore students' perceptions of engagement data collection and usage by focusing on their attitudes toward privacy and data management. By investigating these perceptions, the study seeks to contribute to the development of ethical data practices in online education. To achieve this aim, we conducted

a survey among students who participated in online education. The survey collected both quantitative and qualitative data on their views regarding data collection practices, privacy concerns, and preferences for data handling. This mixed-methods approach allowed for a comprehensive analysis of the students' perspectives.

The results showed that while many students are comfortable with the use of their engagement data for personal and instructor dashboards, significant concerns remain about privacy, particularly with the collection of facial expressions and chat participation data. Students emphasized the importance of transparency and control over their data. This study contributes to the ongoing discourse on ethical data practices in online education by providing empirical insights into students' privacy concerns and preferences. The findings underscore the necessity of adopting transparent and consent-based data collection practices to ensure that students feel secure and informed about how their engagement data is utilized.

2. Theoretical references

2.1. Related work

Research indicates that while students recognize the potential benefits of data-driven educational tools, they simultaneously experience concerns about privacy and data security. According to Jones et al. (2020), students often express anxiety about the extent to which their personal information is monitored and used by educational institutions (Jones et al., 2020). This anxiety stems from a lack of transparency and understanding regarding how data is collected and who has access to it, which can lead to feelings of vulnerability and discomfort among students (Roberts et al., 2016).

Furthermore, studies have shown that students' comfort levels with data collection improve when institutions prioritize transparency and offer clear explanations of data usage (Tsai et al., 2020). Transparency from institutions, coupled with clear explanations of the benefits of data-driven practices, can help build trust (Karimov et al., 2024). Furthermore, students value having a voice in the process, advocating for greater control over their data, including the ability to access, correct, or even opt-out of certain data collection practices. Ultimately, a student-centric approach that prioritizes informed consent, data security, and student agency is crucial for fostering a positive and ethical data culture in education (Brown & Klein, 2020).

Although there exists extensive literature emphasizing the significance of collecting and analyzing engagement data to construct precise

and well-performing learning analytics models (Johar et al., 2023), there are still significant gaps in our understanding of student's attitudes towards the collection and utilization of their data. This paper presents insights collected from a student survey, focusing on their stances on data collection, privacy apprehensions, and preferences concerning data management. Despite substantial research dedicated to measuring student engagement and identifying influencing factors (e.g., Wang et al., 2015; Whitehill et al., 2015), the understanding of students' comfort with their data being used for this purpose remains underexplored (Jones et al., 2020). This gap is crucial, as the ethical implementation of learning analytics hinges on student acceptance and consent.

In workplace settings, clear laws prevent employers from excessively tracking their employees. For example, in many jurisdictions, the use of keyloggers—software programs that record and save keystrokes or take screenshots—is heavily restricted (Gitte, 2024). Such measures are only permissible under specific conditions, such as when there is a suspicion of a criminal offense or a serious breach of employment contract obligations. These regulations are designed to protect employees' privacy and prevent abuse of monitoring technologies. However, how do these principles apply to students in educational settings? How much data should students be expected to share, and if given the choice, how much are they willing to share?

Young (2014) suggests that data from individual sources, while seemingly harmless on their own, can collectively form a detailed picture of a person, leading to unexpected inferences. The effectiveness of anonymity as a "placeholder for privacy" is becoming questionable (ibid.), raising concerns about the current model of consent for data collection and usage. Although consent remains important, students often lack a clear understanding of what they are consenting to. Current terms and conditions may not fully predict or prevent future uses of personal data, and there are concerns about the effectiveness of de-identification, though it does provide some reduction of risk. While most data use is benign, advancements in processing power and data capabilities could increase the risk of reidentification and lead to more significant consequences. Additionally, if students from minority groups or those with particular backgrounds are more reluctant to share data, their absence could bias models, making them less representative of all students (Li et al., 2022). Recent studies also highlight growing concerns about bias and equity in the uses of learner data (Heiser et al., 2023).

2.2. Theoretical framework

This study draws on two theoretical frameworks to analyze the ethical implications of data collection in online education. A key theory applied here is Panopticism, introduced by Michel Foucault, which explores the dynamics of surveillance and power (Foucault, 2020). In the context of online education, Panopticism helps to explain why students might feel uncomfortable or anxious about the collection of data such as facial expressions. The theory suggests that the mere possibility of being constantly observed can lead to a sense of powerlessness and self-censorship, which could negatively affect students' engagement and overall learning experience. Furthermore, Self-Determination Theory (SDT) is employed to understand how the need for autonomy influences student attitudes toward data collection practices (Deci & Ryan, 2012). SDT posits that individuals have an intrinsic need to feel in control of their actions and decisions, which is directly challenged by invasive data collection methods that do not offer clear consent or transparency.

2.3. Legal perspective

The General Data Protection Regulation (GDPR), which came into effect on May 25, 2018, is a data protection law that has influenced data collection and processing practices within the European Union (EU) and beyond (European Commission, 2023). One of the key principles of the GDPR is transparency, as outlined in §12, which requires organizations to process data in a manner that is "concise, transparent, intelligible and easily accessible" using clear and plain language (European Union, 2016).

Under §13 and §14, organizations must inform data subjects at the time of data collection about the purpose of processing, the legal basis for processing, and the rights of the data subjects, among other details. This requirement ensures that students are aware of and consent to the specific uses of their engagement data, addressing concerns about uninformed or non-consensual data collection. Data subjects have several rights under the GDPR, including the right to access §15, the right to rectification §16, the right to erasure §17, and the right to restrict processing §18 (European Union, 2016). These rights empower students to manage their engagement data actively, aligning with their expressed desire for control over their personal information.

In addition to the GDPR, the California Consumer Privacy Act (CCPA) and Switzerland's new Federal

Act on Data Protection (nFADP) provide relevant case studies from different jurisdictions. The CCPA, effective January 1, 2020, grants California residents control over their personal data, requiring businesses to inform consumers about data categories and usage purposes (§1798.100), and providing rights to access (§1798.110), delete (§1798.105), and opt-out of data sale (§1798.120) (State of California Department of Justice, 2018). Similarly, the nFADP, effective September 2023, emphasizes transparency, purpose limitation, and consent (§6), and mandates educational institutions to inform students about data collection and usage purposes (§19). It grants students rights to access (§25), rectify inaccuracies, and restrict processing (§32), and requires measures to protect data from misuse (§8), addressing security and breach concerns (Council, 2020). While CCPA is primarily a consumer protection law, its relevance to educational contexts stems from the fact that online education platforms operate within the same digital ecosystem where personal data is managed similarly to other consumer services. Therefore, the broader principles of data protection within consumer regulation, such as informed consent and the right to access and delete data are equally important in safeguarding students' data.

3. Methodology

3.1. Data collection

To collect data, we developed a survey instrument (see <https://zenodo.org/records/13629306>) which consisted of both multiple-choice and open-ended questions designed to capture detailed information about participants' demographics, educational background, frequency of online class attendance, and their attitudes towards the collection and use of engagement data. The survey included the following subscales to measure various aspects of students' attitudes: (i) Privacy Concerns Subscale: This subscale included items measuring participants' concerns about the privacy of their engagement data, such as "I am concerned about the privacy and security of my engagement data" and "How concerned would you be if your facial expressions were collected via camera during online classes?" (ii) Transparency and Understanding Subscale: This subscale assessed participants' understanding of engagement data collection and their desire for transparency. Example items include "The purpose of collecting engagement data was clearly explained to me" and "I understand what engagement metrics are being tracked and how they are calculated." (iii) Comfort with Data Usage Subscale: This subscale

measured participants' comfort levels with different uses of their engagement data, including items like "I am comfortable with my engagement data being used to create learning dashboards for me" and "I am comfortable with my engagement data being used to create learning dashboards for my instructor/teacher. (iv) Control and Consent Subscale: This subscale captured participants' preferences for control over their data and consent processes. Items included "Would you prefer more control over what engagement data is collected or how it is used?" and "What kind of control would you like?" (v) Perceived Benefits Subscale: This subscale evaluated participants' perceptions of the benefits of tracking and reporting engagement data. An example item is "Do you believe that tracking and reporting student engagement data is beneficial for online learning?"

The scales used in our survey were developed specifically for this study to address the unique context of engagement data in online learning environments. These scales were reviewed and refined through an iterative process involving four researchers with expertise in learning analytics and educational research. The team held five meetings to discuss and confirm the content, clarity, and relevance of the survey items. All authors reached a consensus on the final version of the scales.

The primary criterion for participation in the survey was having experience in online education and familiarity with the use of engagement data. To ensure that only participants with relevant backgrounds completed the survey, we primarily targeted two groups of students from specific institutions known for their extensive use of online education and engagement data. These groups were: 1) students from the Swiss Cyber Institute, which provides cybersecurity education for technology experts through online education. 2) Students from the University of Jyväskylä's Educational Technology Department, where the majority of courses are taken online. We selected these institutions because both the institutions and the students enrolled there have a high level of familiarity with engagement data. At the Swiss Cyber Institute, students frequently engage with online platforms that collect and use engagement data as part of their cybersecurity education. Similarly, at the University of Jyväskylä's Educational Technology Department, students regularly interact with digital learning environments that track engagement metrics.

This targeted sampling approach was chosen to ensure that participants had sufficient experience and understanding of engagement data practices, which was essential for providing informed responses to our survey. However, to enrich our dataset and capture a wider

range of perspectives, we also accepted responses from students at other institutions who met our participation criteria. These participants were screened to ensure they had relevant online education experience, thereby maintaining the relevance and reliability of the collected data.

The survey was conducted online using the Google Forms platform and took approximately 15-20 minutes to complete. Participants were informed about the purpose of the study and their rights as participants before beginning the survey, and consent was obtained from all participants prior to their participation. No compensation was provided to participants for completing the survey.

To mitigate common method biases and ensure the accuracy and reliability of the responses, several strategies were implemented during the survey process. First, participants were assured of the anonymity and confidentiality of their responses, which helped reduce social desirability bias and encouraged honest answers. The survey was carefully designed to include a mix of question types, including reverse-coded items, to minimize acquiescence bias. Furthermore, clear and detailed instructions were provided at the beginning of the survey, emphasizing the importance of providing thoughtful and genuine responses.

3.2. Data analysis

The collected survey data were analyzed both quantitatively and qualitatively. For the quantitative analysis, we began by cleaning the survey data to remove any incomplete or inconsistent entries. Descriptive statistics, such as frequencies and mean scores, were calculated for the multiple-choice questions to summarize the demographic information and general trends in participants' responses.

Qualitative data from open-ended questions were analyzed thematically to identify common themes and concerns raised by participants regarding the ethical implications of engagement data collection and use (Braun & Clarke, 2012). This process involved both inductive and deductive coding. Inductive codes emerged directly from participants' responses by highlighting concerns such as "fear of data misuse for marketing or illegal activities" and "apprehensions about unauthorized access and data breaches." Deductive codes were derived from existing literature and predefined categories, such as "privacy concerns," "desire for transparency," and "control over data collection and use." For instance, comments like "I want to be informed about how my data will be used" and "I should have the option to opt in or out of certain data

collection practices” were coded under the deductive themes of ”transparency” and ”control.”

3.3. Participants

The participants (n=108) are primarily 25-34 years old (n=51), with most residing in Switzerland (n=64) and Finland (n=17). A majority have postgraduate degrees (n=54), followed by those with doctoral or higher, high school, undergraduate, and other educational levels. In terms of fields of study, most participants are studying in Science, Technology, Engineering, and Mathematics (STEM) (n=67) fields, and the majority attend public institutions (n=76). Regarding online class attendance, most participants attend classes often (n=47), while others report attending always, sometimes, or rarely.

4. Results

4.1. Participant awareness, comfort levels, and privacy concerns

The participants highlighted various aspects, including participants’ comfort levels, privacy concerns, and suggestions for improving the ethical handling of engagement data. When asked if their learning platforms informed them about how their engagement data would be collected and used, 65 participants responded ”yes,” and 42 responded ”no.” Among those who were not informed, the majority (n=72.5) expressed a preference to be informed. We examined participants’ perceptions of the clarity and understanding of engagement data collection. The results show that participants generally agreed with the clarity of the explanations about why their engagement data was being collected. On a scale from 1 to 5, the average rating for this clarity was 3.04 ($\sigma=1.39$). However, the understanding of what engagement metrics are being tracked and how they are calculated was slightly lower, with an average response of 2.88.

Figure 1 visualizes participants’ answers on how comfortable they are with having their engagement data reported in (i) their personal, (ii) their teachers’, and (iii) their peers’ dashboards. The responses to the statement ”I am comfortable with my engagement data being used to create learning dashboards for me” show a moderate level of comfort, with a mean score of 3.88 out of 5 ($\sigma=1.02$). A similar trend is observed for the statement ”I am comfortable with my engagement data being used to create learning dashboards for my instructor/teacher,” which has a mean score of 3.83 ($\sigma=1.0$). These findings suggest that participants are generally comfortable with their engagement data being

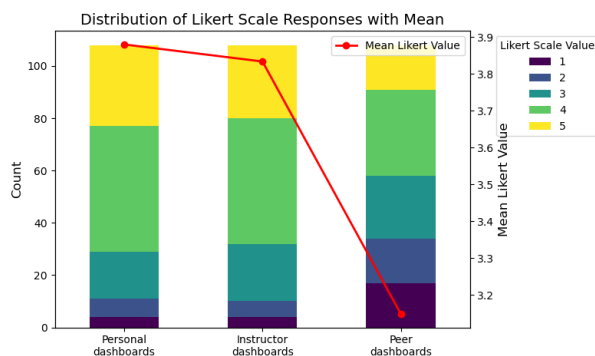


Figure 1. Students’ answers how comfortable they are with having their engagement data reported in (i) their personal, (ii) their teachers’, and (iii) their peers’ dashboards.

used for personal and instructor dashboards. However, the comfort level drops significantly when it comes to making engagement data visible to other students (see Figure 1). The statement ”I am comfortable with my engagement data being used to create learning dashboards that can be visible by other students” has a lower mean score of 3.15 ($\sigma=1.31$), indicating less comfort with this level of transparency. This is further supported by the distribution of responses, where only 9 participants rated their comfort at the highest level (5), while 23 participants rated it at the lowest level (1).

When considering the platform’s ability to store and analyze chat participation, participants again show moderate comfort, with a mean score of 3.16 ($\sigma=1.35$). The distribution of responses indicates a fairly even spread, with 18 participants expressing high comfort (rating 5) and 15 participants expressing low comfort (rating 1). Moreover, the statement ”I am concerned about the privacy and security of my engagement data” has a mean score of 2.92 ($\sigma=1.31$) which indicates fairly high concern. This is reflected in the distribution, with 38 participants rating their concern at the highest level (5).

Figure 2 illustrates that students are particularly concerned about facial expression, with a mean score of 1.91 ($\sigma=1.09$). More than half of the participants (55) rated their concern at the highest level (5), highlighting significant apprehension regarding this aspect of data collection. There appeared to be a difference between higher (postgraduate or doctoral) and lower (high school or undergraduate) education levels, but it was not statistically significant, $t(df) = 0.34, p = 0.733$. Regarding the appropriateness of tracking participation frequency, the responses show moderate agreement, with a mean score of 3.03 ($\sigma=1.24$). The distribution reveals a range of opinions, with 14 participants

rating it highly appropriate (5) and 16 participants rating it not appropriate at all (1). Finally, the comfort level with recording online activity duration for engagement analysis also shows moderate comfort, with a mean score of 2.94 ($\sigma=1.25$). The distribution of responses is similar to the other metrics, with 14 participants expressing high comfort (5) and 15 participants expressing low comfort (1).

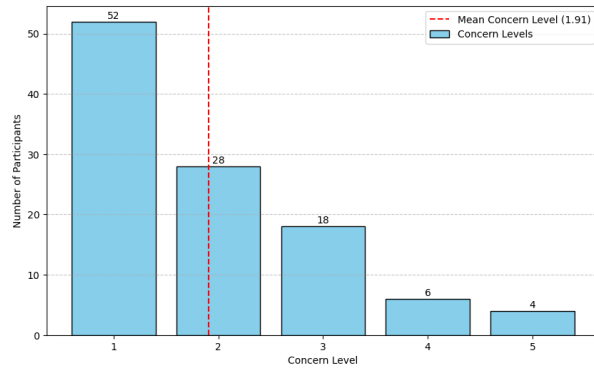


Figure 2. Distribution of Concerns about Facial Expression Data Collection. The concern levels are rated on a scale from 1 to 5, where 1 represents "Very concerned" and 5 represents "Not concerned".

4.2. Ethical considerations and participant control over data

While most of the participants (74.07%) noted that they do not have any ethical issues regarding the collection and use of their engagement data, they also expressed worries regarding the handling of engagement data. A major concern raised was the fear that data could be misused for marketing or illegal activities. One participant expressed, "not to be used for marketing purposes." They stressed the importance of transparency in managing, storing, and utilizing data, with one participant noting, "If the use purposes can be made clearly transparent, it would help." Many participants were curious about the duration of data retention, who can access it, and the specific reasons for collecting it. Issues surrounding data breaches and unauthorized access were also highlighted by indicating a lack of confidence in data security measures and concerns about ensuring deletion upon request. One participant mentioned, "I need to know how my data is handled and its retention period." Some participants also questioned the ethics of using data to influence decisions such as exam eligibility where they emphasized the necessity for communication and guidelines. Privacy concerns were common with students reporting apprehension about data being shared or sold without consent, as

potential risks associated with merging engagement data with other datasets were highlighted. One participant remarked, "Where it will be uploaded? Who will have access?"

4.3. Impact on participation, study habits, and perceived benefits

The responses to the question about the benefits of tracking and reporting student engagement data for online learning show that the majority of participants rated it positively, with many responses clustering around 3 and 4 on a scale from 1 to 5 ($\mu = 3.32$, $\sigma = 1.10$). One participant mentioned, "It helps me keep track of my progress and areas I need to improve." This suggests a general agreement among participants that tracking engagement data is beneficial. Table 2 presents the changes in class participation since students learned about their engagement scores. The most common response was "No change in participation," reported by 49 participants. 16 students stated that they participate more in discussions. One participant noted, "Knowing my engagement score motivates me to contribute more in discussions." However, 8 participants indicated that they participated less due to anxiety or discomfort, with one stating, "Seeing my engagement score makes me anxious and less likely to participate." Additionally, 7 students tried to be more visible in other ways, such as sharing resources or helping peers, and 8 participants mentioned asking more questions. The remaining respondents ($n=20$) selected the "Other" option.

Table 1. Changes in Participation Since Learning About Engagement Scores

Participation Change	Count
No change in participation	49
I participate more in discussions	16
I participate less due to anxiety or discomfort	8
I try to be more visible in other ways (e.g., sharing resources, helping peers)	7
I ask more questions	8
Other	20

Table 3 illustrates the changes in study habits following awareness of engagement scores. The majority, 58 participants, reported "No change in study habits." 14 participants stated that they study more hours per week, while 10 have started using more diverse learning resources like videos and articles. 9 participants mentioned studying less but more efficiently, and 6 indicated focusing on different subjects or topics more

relevant to their interests. Students also had the option of selecting 'Other'; 11 of them did so.

33.3% of participants mentioned that they would prefer more control over what engagement data is collected or how it is used. From this perspective, they emphasized the need for transparency and access to their data by expressing a desire to review and audit the data collected about them. Another major theme was the desire for control over data collection and usage. Participants wanted the ability to opt in or out of certain data collection practices and to specify preferences for how their data is used for certain data collection types. Additionally, they sought permission before data collection or usage and the option to choose which data metrics are collected. One participant stated, *"I want to be able to delete all data related to me. I want full control over my data."* Privacy and security concerns were also one of the aspects that they mentioned. Examples of these include requests to remove personally identifiable information, control the number of people with access to the data, and ensure that data collection practices are clearly communicated and consented to.

Table 2. Changes in Study Habits Since Learning About Engagement Scores

Study Habits Change	Count
No change in study habits	58
I study more hours per week	14
I use more diverse learning resources (e.g., videos, articles)	10
I study less but more efficiently	9
I focus on different subjects or topics more relevant to my interests	6
Other	11

The responses identified several key themes on how to make the process of collecting and using engagement data more ethical and comfortable. Transparency and information were highly valued, with participants requesting clear explanations of how data is analyzed and used, as well as being informed about data collection at the beginning of the course and before each class. One participant mentioned, *"I want to keep my collected data transparent and provide explanations."* Ensuring the privacy and anonymity of collected data was a significant concern and participants suggested keeping data anonymized and removing personally identifiable information. One participant noted, *"I need to be assured about anonymization to feel comfortable with the data collection process."* Access to their data and feedback on its usage were important and within this framework, participants mentioned that they would like to have regular engagement reports

and the ability to see the same data as teachers and administrators. Lastly, participants emphasized the importance of demonstrating the benefits and relevance of data collection to ensure that students see a real benefit and that data collection methods accurately reflect their engagement and knowledge levels.

5. Discussion

Our study underscores the complex attitudes students hold toward the collection and use of their engagement data in online learning environments. As emphasized by Jones et al. (2020), while learning analytics literature has evolved to address multiple facets of student privacy (e.g., Ifenthaler & Schumacher, 2016; Jones, 2019), the perceptions and voices of students themselves have not been comprehensively addressed. Our research fills this gap by presenting findings that reveal a nuanced perspective: students recognize the potential benefits of engagement data but express significant concerns about privacy, transparency, and control over their data.

More specifically, significant privacy concerns persist, particularly with the collection of facial expressions and chat participation data. This is consistent with findings by Slade et al. (2019), who highlighted the ambivalence students feel towards extensive data collection. Facial expression data, in particular, raises serious ethical questions due to its intrusive nature and the potential for misuse. The collection of facial expression data, while potentially useful for detecting engagement and emotional states, poses risks that must be carefully managed. Monkaresi et al. (2016) concluded their article with the "hope that improved automatic detection of engagement in computerized education environments will lead to more effective learning". While improved detection of engagement has the potential to enhance learning experiences by allowing teachers to tailor their courses more effectively, it also raises concerns about student privacy. It is important to consider the balance between the benefits of using engagement data and the extent of information students should be expected to share. In our research, one of the participants also mentioned, *"I need to be assured about anonymization to feel comfortable with the data collection process,"* highlighting the need for robust anonymization techniques to protect student privacy, particularly in biological data collection.

The positive reception of engagement data for personal and instructor dashboards aligns with existing literature, which indicates the value of learning analytics in enhancing educational outcomes (Matcha et al., 2020). Students in our study recognized that

engagement data could help tailor learning experiences and provide meaningful feedback. Furthermore, our study indicates that while students see the benefits of engagement data, these benefits must be balanced with ethical considerations. The call for regular engagement reports and the ability to access the same data as teachers and administrators suggest a move towards more inclusive and transparent data practices (Heiser et al., 2023). This approach not only respects student autonomy but also builds trust in the data collection process.

The desire for control over data collection and usage reflects findings from previous studies, where students expressed a strong preference for consent and control mechanisms (Li et al., 2022). Participants in our study wanted the ability to review and audit their data, opt in or out of certain data collection practices, and ensure that data collection is consented to and clearly communicated. This supports Young (2014)'s argument that informed consent and transparency are essential in educational data practices.

The legal perspective is essential in shaping ethical practices for using engagement data in online education, as it aligns with the concerns and preferences expressed by students in our study. Regulations like the GDPR in the EU, the CCPA in the US, and the nFADP in Switzerland offer clear guidelines for transparency and privacy protection. Our findings show that students highly value transparency and control over their data. This is in line with the GDPR's requirements for clear communication about data use and the CCPA's rules for accessing, deleting, and opting out of data collection. The GDPR gives students rights such as access, correction, deletion, and restriction of their data, which matches the students' desire for more control over their engagement data found in our study. Similarly, the nFADP's focus on the right to access and delete personal data mirrors our participants' wish to review and manage their data. By following these legal guidelines, educational institutions can build trust and promote the ethical use of learning analytics, addressing the privacy and security concerns raised by our respondents.

The significant concern students express regarding the collection of facial expression data can be further understood through the lens of psychological and behavioral theories. For instance, Panopticism, a concept introduced by Michel Foucault, suggests that the possibility of being observed constantly can create a sense of powerlessness and anxiety, as individuals alter their behavior due to the perceived surveillance (Foucault, 2020). This could explain why students are particularly uneasy about facial expression tracking, which they may perceive as a form of

constant monitoring. Additionally, the SDT poses that individuals have an inherent need for autonomy and control over their lives, including their personal data (Deci & Ryan, 2012). The intrusion of collecting facial expression data without clear consent or understanding may be seen as a violation of this autonomy which leads to discomfort and resistance.

5.1. Limitations and future work

Of the related works, the studies by Li et al. (2022) and Jones et al. (2020) are the most similar to our study in terms of content, as they explored similar themes regarding student perceptions of engagement data. However, these studies focused on students from universities in the United States. In contrast, our study primarily gathered perspectives from students in European countries. Unlike their studies, we did not collect any data regarding gender and ethnic origin. For example, Li et al. (2022) investigated students' general propensity to consent to learning analytics stratified by ethnicity and gender. They found that students identifying as Black were significantly less likely to respond and self-reported lower levels of institutional trust, while female students expressed concerns about data collection but were more comfortable with instructors using their data for learning engagement purposes. In a future study, it would be interesting to determine if similar differences can be found in other educational institutions across different geographical and cultural contexts.

Another limitation relates to the sample size and generalizability. While our study provides useful insights, the sample size of 108 students may not be large enough to generalize the findings to all online education contexts. Additionally, the demographic composition and specific characteristics of our participants, who were mainly from Finland and Switzerland—European countries with high GDP—could have affected the generalizability of the results. Repeating this study with a larger and more diverse sample from other institutions would help to increase the generalizability of the findings. Finally, while our study focused on ethical concerns related to facial expressions and chat participation, it is important to note that other forms of engagement data, such as interaction with course materials and clickstream data, also warrant consideration. These data types present additional ethical challenges related to privacy and consent, and future research should explore these aspects in more depth.

5.2. Practical implications

The findings of this study have several practical implications for educators, administrators, and policymakers involved in online education. First, the study highlights the importance of transparency and clear communication when collecting and using student engagement data. Educational institutions should implement clear guidelines and consent processes that inform students about what data is being collected, how it will be used, and who will have access to it. This will help build trust and ensure that students feel secure and respected in their learning environments.

Second, the study highlights the need for institutions to consider students' comfort levels with different types of engagement data. For example, while students may be generally comfortable with their data being used for personal or instructor dashboards, they may be less comfortable with data being visible to their peers. Institutions should carefully design their learning analytics tools to align with these preferences, potentially allowing students to control the visibility of their data. Additionally, these insights can be invaluable for policy developers seeking to understand students' perspectives on data privacy and usage. By incorporating these findings into policy frameworks, developers can create regulations and guidelines that better align with student expectations and concerns, thereby fostering more ethical and effective educational practices.

6. Conclusion

This study explored students' perceptions of engagement data collection and usage in online education by focusing on their attitudes towards privacy, transparency, and control over their data. The findings indicate that while students generally see the benefits of using engagement data for personal and instructor dashboards, significant concerns remain regarding privacy, particularly with the collection of facial expressions and chat participation. The majority of participants emphasized the need for clear communication about data collection practices and expressed a strong desire for greater control over their data. This includes being informed about how their data is used by having the option to opt in or out of certain data collection practices and ensuring data security.

Moreover, students' comfort levels varied depending on who had access to their engagement data. They were more comfortable with their data being used for personal and instructor dashboards than for dashboards visible to other students. This highlights the need

for institutions to consider these comfort levels when implementing learning analytics tools. The results also demonstrate the importance of ethical data practices in learning analytics by suggesting that educational institutions should prioritize transparency, consent, and data security to address students' concerns. By doing so, institutions can foster a more trustworthy and secure learning environment that respects students' privacy and encourages their engagement.

In conclusion, the ethical use of engagement data in online education requires careful consideration of students' perceptions and concerns. Institutions must adopt practices that enhance transparency, provide control over data usage, and ensure robust data security measures. These steps are crucial in building students' trust and promoting the ethical integration of learning analytics in education. Future research should continue to explore these issues, particularly in diverse educational settings, to develop comprehensive guidelines for ethical data practices in online learning environments.

Acknowledgements

The work was supported by the Swiss Cyber Institute and Academy of Finland (grant no. 356314).

References

- Braun, V., & Clarke, V. (2012). *Thematic analysis*. American Psychological Association.
- Brown, M., & Klein, C. (2020). Whose data? which rights? whose power? a policy discourse analysis of student privacy policy documents. *The Journal of Higher Education*, 91(7), 1149–1178.
- Council, S. F. (2020). New federal act on data protection (nfadp). *Swiss Federal Council*. Available online: <https://www.kmu.admin.ch/kmu/en/home/facts-and-trends/digitization/data-protection/new-federal-act-on-data-protection/nfadp.html> (accessed on 13 December 2023).
- Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. *Handbook of theories of social psychology*, 1(20), 416–436.
- European Commission. (2023). Data protection in the eu [Available at: European Commission].
- European Union. (2016). General data protection regulation (gdpr) [Articles 12, 13, 14, 15, 16, 17, 18, 32].
- Foucault, M. (2020). Panopticism. In *The information society reader* (pp. 302–312). Routledge.

- Gitte, H. (2024). Pc-überwachung am arbeitsplatz: Was ist erlaubt? [Accessed: 2024-06-06]. <https://www.arbeitsrechte.de/pc-ueberwachung-am-arbeitsplatz/>
- Heiser, R., Dello Stritto, M. E., Brown, A., & Croft, B. (2023). Amplifying student and administrator perspectives on equity and bias in learning analytics: Alone together in higher education. *Journal of Learning Analytics*, 10(1), 8–23.
- Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development*, 64, 923–938.
- Johar, N. A., Kew, S. N., Tasir, Z., & Koh, E. (2023). Learning analytics on student engagement to enhance students' learning performance: A systematic review. *Sustainability*, 15(10), 7849.
- Jones, K. M. (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), 1–22.
- Jones, K. M., Asher, A., Goben, A., Perry, M. R., Salo, D., Briney, K. A., & Robertshaw, M. B. (2020). "we're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology*, 71(9), 1044–1059.
- Karimov, A., Saarela, M., & Kärkkäinen, T. (2023). The impact of online educational platform on students' motivation and grades: The case of khan academy in the under-resourced communities. *International conference on educational data mining*, 466–473.
- Karimov, A., Saarela, M., & Kärkkäinen, T. (2024). Ethical educational data processing differences of students with special needs in post-soviet countries. *International conference on educational data mining*.
- Li, W., Sun, K., Schaub, F., et al. (2022). Disparities in students' propensity to consent to learning analytics. *International Journal of Artificial Intelligence in Education*, 32, 564–608.
- Matcha, W., Uzir, N. A., Gašević, D., & Pardo, A. (2020). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies*, 13(2), 226–245.
- Monkaresi, H., Bosch, N., Calvo, R. A., & D'Mello, S. K. (2016). Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Transactions on Affective Computing*, 8(1), 15–28.
- 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, 1959.
- Saarela, M., & Kärkkäinen, T. (2017). Knowledge discovery from the programme for international student assessment. *Learning Analytics: Fundamentals, Applications, and Trends: A View of the Current State of the Art to Enhance e-Learning*, 229–267.
- Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 235–244.
- State of California Department of Justice. (2018). California consumer privacy act (ccpa) [Sections 1798.100, 1798.105, 1798.110, 1798.120].
- Tempelaar, D., Nguyen, Q., & Rienties, B. (2020). Learning analytics and the measurement of learning engagement. *Adoption of data analytics in higher education learning and teaching*, 159–176.
- Tsai, Y.-S., Whitelock-Wainwright, A., & Gašević, D. (2020). The privacy paradox and its implications for learning analytics. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 230–239.
- Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in human behavior*, 89, 98–110.
- Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating how student's cognitive behavior in mooc discussion forums affect learning gains. *International Educational Data Mining Society*.
- Whitehill, J., Williams, J., Lopez, G., Coleman, C., & Reich, J. (2015). Beyond prediction: First steps toward automatic intervention in mooc student stopout. *Available at SSRN 2611750*.
- Young, E. (2014). Educational privacy in the online classroom: FERPA, MOOCs, and the big data conundrum. *Harv. JL & Tech.*, 28, 549.