International Review of Research in Open and Distributed Learning

On the Ethical Issues Posed by the Exploitation of Users' Data in MOOC Platforms: Capturing Learners' Perspectives

Matthieu Tenzing Cisel ⓘD

Cite this article

Article abstract

Due notably to the emergence of massive open online courses (MOOCs), stakeholders in online education have amassed extensive databases on learners throughout the past decade. Administrators of online course platforms, for instance, possess a broad spectrum of information about their users. This information spans from users' areas of interest to their learning habits, all of which is deduced from diverse analytics. Such circumstances have sparked intense discussions over the ethical implications and potential risks that databases present. In this article, we delve into an analysis of a survey distributed across three MOOCs with the intention to gain a deeper understanding of learners' viewpoints on the use of their data. We first explore the perception of features and mechanisms of recommendation systems. Subsequently, we examine the issue of data transmission to third parties, particularly potential recruiters interested in applicants' performance records on course platforms. Our findings reveal that younger generations demonstrate less resistance towards the exploitation of their data.

On the Ethical Issues Posed by the Exploitation of Users' Data in MOOC Platforms: Capturing Learners' Perspectives

Matthieu Tenzing Cisel
Institut Des Humanités Numériques, CY Cergy Paris Université

Abstract

Due notably to the emergence of massive open online courses (MOOCs), stakeholders in online education have amassed extensive databases on learners throughout the past decade. Administrators of online course platforms, for instance, possess a broad spectrum of information about their users. This information spans from users' areas of interest to their learning habits, all of which is deduced from diverse analytics. Such circumstances have sparked intense discussions over the ethical implications and potential risks that databases present. In this article, we delve into an analysis of a survey distributed across three MOOCs with the intention to gain a deeper understanding of learners' viewpoints on the use of their data. We first explore the perception of features and mechanisms of recommendation systems. Subsequently, we examine the issue of data transmission to third parties, particularly potential recruiters interested in applicants' performance records on course platforms. Our findings reveal that younger generations demonstrate less resistance towards the exploitation of their data.

Keywords: learning analytics, massive open online course, MOOC, ethics, recommender systems, data privacy

On the Ethical Issues Posed by the Exploitation of Users' Data in MOOC Platforms: Capturing Learners' Perspectives

Over the past decade, virtual learning environments have been increasingly used in educational settings and, along with them, dashboards whose aim is notably to monitor users' activity. This evolution has triggered vivid debates worldwide over the ethical questions raised by the use of learning analytics (LA), whose potential applications have been widely documented (Ifenthaler, 2017; Ifenthaler & Tracey, 2016). Pardo and Siemens (2014) listed the privacy issues that needed to be considered during the design of LA-based research projects if learners were to "make rational, informed choices regarding consent to having their data collected, analysed, and used" (p. 438). Prinsloo and Slade (2015, 2017, 2018) highlighted the importance of students' self-management of analytics.

To address these issues, Jisc, a charitable organization that champions the use of digital technologies in the UK, developed a *Code of Practice for Learning Analytics* (Jisc, 2015; Sclater, 2016), which discusses a taxonomy of issues and necessary governance structures. An Australian study by West et al. (2016) collected views from key academic stakeholders, from professors to institutional leaders, to develop a decision-making framework aimed at fostering transparent institutional policies and broader ethical literacy.

The rise of massive open online courses (MOOCs) in the early 2010s led to a sharp increase in publications on the use of LA on an unprecedented scale. Early studies dealt with hundreds of thousands of learners (Breslow, 2013; Ho et al., 2014), and there continues to be a stream of research focusing, in some cases, on the behaviour of millions of registrants worldwide (Kizilcec et al., 2017; Wintermute et al., 2021), raising numerous ethical issues (Marshall, 2014).

As MOOC designers, we took the opportunity to investigate some of these issues through a survey consisting of closed questions, which was broadcast in three different online courses that started in 2021. We employed a self-selection sampling strategy. The large audiences of these courses enabled us to gather a significant number of responses, and respondents were likely prone to consider the issue of data privacy because all of their actions were being logged by the hosting platform. We investigated two emblematic data-related use cases in MOOCs: recommender systems and the transmission of data to third parties. The first represents a typical service aimed at improving user experience, while the latter involves the monetization of user data. In the next paragraphs, we expose some of the reasons why these problems are significant in the context of MOOCs.

The need for user oversight of LA has become increasingly pressing with the rise of artificial intelligence in online learning (Hsu et al., 2022), and the increasing use of adaptive systems with personalization (Hwang et al., 2022). As underscored in our previous publication focused on MOOC catalogues, the diversity of the offer has increased sharply during the 2010s (Cisel, 2019), increasing the need for dedicated recommender systems. In some online course platforms such as DataCamp (Baumer et al., 2020), users are even provided with a list of course and resource suggestions after an automated assessment in order to improve their score over time.

A few publications have delved into the issue of recommender systems in the context of MOOCs: "With MOOCs['] proliferation, learners will be exposed to various challenges and the traditional problem in TEL [technology enhanced learning], finding the best learning resources, is more than ever up to date" (Bousbahi & Chorfi, 2015, p. 1813).

As the audience of MOOC platforms grows, associated recommendation algorithms may influence at large scales the topics that users choose to pursue. This could lead to gradual homogenization of learning projects (Tough, 1971) across the landscape of online education, and one should not underestimate the potential long-term impacts of these recommender systems. While these long-term dynamics should not be overlooked, learners may be more immediately concerned with what happens to the data fed into these algorithms, particularly with regard to potential data leaks. Due to privacy concerns, researchers working with MOOC data have developed frameworks to ensure the protection of users' anonymity (Hutt et al., 2022). It is worth exploring learners' perspectives on the exploitation of information even for services like recommender systems, which were designed to improve the user experience. This becomes even more important if this information is monetized through interactions with third parties (Ferguson, 2019).

Selwyn (2019, p.11) highlighted the risk that profit-led projects posed to public education in the "burgeoning data economy." He advocates that learners should have a greater oversight on their LA, a topic that gained increasing importance with the implementation of the General Data Protection Regulation (Presthus & Sørum, 2019). With regard to MOOCs, employers may express skepticism about the value of MOOC certificates (Radford et al., 2014), particularly because not all classes are career relevant (Kizilcec & Schneider, 2015). Employers might therefore show more interest in other types of data generated by these courses (Allal-Chérif et al., 2021).

If we adopt a strict definition of MOOCs—online classes with free registration and open access to most course material (Daniel, 2012)—concerns about the monetization of learners' data are heightened by the complexity of MOOCs' business models (Porter, 2015). During the early stages of the MOOC movement, authors demonstrated that these courses were costly to design and broadcast (Hollands & Tirthali, 2014) and that income from selling certificates rarely covered design costs (Burd et al., 2015). While it was argued that these classes did not necessarily need to be economically sustainable as they were used in part as advertisement tools by academic institutions (Kolowich, 2013), it remains likely that various stakeholders, especially hosting platforms like Coursera, will seek to diversify their income sources (Burd et al., 2015). It is therefore relevant to explore learners' reactions to the monetization of their data, particularly if it involves sharing it with third parties like recruiters, and to understand how different groups of learners may respond to varying business models. This latter issue drove us to incorporate sociodemographic variables in our analysis.

Sociologists have repeatedly proven the importance of age when it came to individuals' perspectives on information privacy (Regan et al., 2013). Miltgen and Peyrat-Guillard (2014, p. 107), for instance, underscored the influence of age in an influential empirical research study: "In addition to culture, age dictates how people relate to IT, a phenomenon that influences their privacy concerns." A generational shift in the concern over data privacy could trigger long-term changes in the acceptability of a diverse range of uses of LA; while the respondents' ages were collected in most of the aforementioned quantitative studies, this information has not been exploited to understand the data. These diverse considerations led us to investigate a set of research questions, three of which are presented in this article:

RQ1: What are learners' perspectives on the exploitation of MOOC data to feed recommender systems, and how much do their opinions depend upon the goal of the recommendation?

RQ2: What are learners' perspectives on the transmission of MOOC data to third parties (recruiters, etc.)?

RQ3: To what extent does the age of learners affect their perspective on the use of their data?

Based on the findings of Miltgen and Peyrat-Guillard (2014), we hypothesized that younger users were more open to the use of their data.

Privacy Concerns and the Calculus Framework

The scholarly exploration across the field of management information systems (MIS) and other related disciplines inspired Dinev et al. (2013) to develop a comprehensive research model that can prove useful to the interpretation of privacy concerns regarding LA. We relied on this author's work, including literature reviews, as a theoretical framework to interpret our data. Central to this model is the "calculus framework" of privacy, which encapsulates the dynamic interplay between risk and control elements (as per the studies of Culnan & Armstrong, 1999; Dinev & Hart, 2006). This calculus perspective on privacy has been acknowledged for its significant utility in dissecting the pressing privacy issues faced by the contemporary consumer (Culnan & Bies, 2003).

Underpinning this perspective is the concept that privacy is not an absolute term. It implies that there is a certain fluidity to it, shaped by a cost–benefit analysis or, in other words, a "calculus of behavior" (Laufer & Wolfe, 1977, p.22). Such a notion draws a nuanced portrait of privacy, showcasing it as a concept that extends beyond the simplistic binary of private or public. To strengthen this core concept of privacy, Dinev et al. (2013) has interwoven it with modern theories of control and risk. These theories strive to elucidate the process by which individuals formulate their perceptions of risk and control regarding their personal information.

The extent of privacy concerns that an individual harbours has been demonstrated to influence their decisions about revealing personal information. Personal data and the concept of privacy risk play vital roles in this process. Additionally, the calculus framework of overall privacy (as highlighted by Dinev & Hart, 2004, 2006) further emphasizes the risk–control interplay.

Interestingly, both risk and control are seen to function as privacy-driven beliefs linked to the possible repercussions of data disclosure, raising questions about the perceived risks of a disclosure of LA data. Therefore, based on this extensive literary exploration, Dinev and Hart pinpoint the two primary factors that contribute to the perception of privacy: the control perceived over one's information and the perceived risk associated with it.

This calculus perspective on privacy, hailed for its utility in analysing modern consumer privacy concerns (Culnan & Bies, 2003), reiterates the understanding that privacy is not a fixed construct. As advocated by Klopfer & Rubenstein (1977), it is a fluid concept, subject to a cost–benefit analysis and, thus, forms a part of the "calculus of behavior" (Laufer & Wolfe, 1977).

These identified components—perceived control and perceived risk—are then integrated into Dinev et al.'s (2013) model. The author proposes that the risk an individual perceives is influenced by a combination of factors. These include the expected outcomes from sharing their information, the sensitivity of the information, the importance placed on transparency about the information, and the legal expectations tied to regulations.

Delving deeper into the model, the privacy calculus concept suggests that individuals weigh costs against benefits in their privacy decisions. This trade-off scenario is assumed to be a vital aspect when an individual is considering sharing personal information with service providers or businesses. This is because, according to the calculus perspective of privacy, consumers carry out a cost–benefit analysis when asked to provide personal data (Culnan, 1993; Dinev & Hart, 2006; Milne & Gordon, 1993).

This decision-making behaviour is driven by an individual's aim to achieve the most positive net outcome, as highlighted by Stone and Stone (1990). Overall, Dinev et al. (2013) extends the calculus perspective on privacy by incorporating it with an understanding of how contextual (information sensitivity), organizational (transparency importance), and legal (regulatory expectations) influences shape an individual's perceived risk. They further propose that this blend of influences collectively determines an individual's decision on information disclosure, whether it be LA for a recommender system or grades in a MOOC for a potential recruiter.

Methods

In this section, we initially present different authors' approaches to the ethics of data use in online education, and then position our work within this expanding literature. We then delve into the data collection tool—MOOCs—and discuss Likert scale–based items. We finally describe the respondents' characteristics.

Information Ethics: An Empirical Approach

Although many ethical issues raised by the increasing use of LA have been mapped through literature reviews (Cerratto Pargman & McGrath, 2021; Tzimas & Demetriadis, 2021) and specific articles, the majority of contributions are conceptual, especially from the perspective of information ethics (Rubel & Jones, 2016). In a literature review, Hakimi et al. (2021) listed only 14 empirical papers or case studies out of the 77 publications about the ethical concerns raised by LA that they considered, therefore highlighting the need for more data collection on the topic. The articles published by Schumacher and Ifenthaler (2018) and Jones et al. (2020) well illustrate a qualitative approach to exploring learners' perspectives. Both articles feature qualitative studies on students' expectations about how their analytics would be used.

Our current work belongs to the category of survey-based assessment of learners' perspectives on their analytics, with the aim of understanding how their attitudes vary depending on the proposed purpose of these analytics (Arnold & Sclater, 2017; Viberg et al., 2022; West et al., 2020; Whitelock-Wainwright et al., 2019; Whitelock-Wainwright et al., 2020; Whitelock-Wainwright et al., 2021). Whitelock-Wainwright and colleagues proposed and validated the Student Expectations of Learning Analytics Questionnaire (SELAQ), a comprehensive tool that encompasses common issues and applications of LA in higher education. SELAQ considers three dimensions: expectations of service features,

communalities, and ethical and privacy concerns. "Will the university ask for my consent before my educational data are outsourced for analysis by third-party companies?" is a typical question drawn from the latter category. A significant proportion of items are relevant only in the context of a university. As the survey is targeted towards students, a significant proportion of items are relevant only in a university context; therefore, this research instrument is not applicable to our context. Unlike these authors, our study focuses on a narrower range of possible uses but explores them in greater depth.

Using MOOCs to Circulate Surveys

As argued by Zimmermann et al. (2016), MOOCs are great tools to collect data, notably through surveys. Due to their ability to quickly reach thousands, if not tens of thousands of potential respondents, these courses have been used extensively over the years to carry out survey-based research projects. In some of our previously published research, thanks to collaboration with dozens of MOOC instructors, we gathered more than 8,000 responses, enabling us to obtain statistically robust results (Cisel, 2018). Stephens-Martinez et al. (2014) used MOOCs to gather thousands of responses for a survey aimed at evaluating the mock-up of a dashboard. Learners were required to assess the utility of a set of indicators meant to monitor course activity.

The use of an online course to broadcast a survey on LA is made all the more relevant by the fact that respondents are themselves directly concerned with the question. The mere registration and navigation to the page where the survey is embedded (we located the form within the course itself) generates logs that can thereafter be analysed. Many respondents typically engage in proposed graded activities, which makes questions about the processing of their grades and analytics relevant to them. Self-selection bias is inherent in online surveys (Bethlehem, 2010), but it did not impede comparisons within items or across age categories, as self-selection affected all groups equally.

Data Collection Tool

The MOOC platform France Université Numérique (FUN) was established by the French Ministry of Higher Education in 2014 and has gathered millions of registrations over the years (Cisel, 2019; Wintermute et al., 2021); it requires all instructors, at the beginning of each course, to incorporate a survey of 26 items focusing on users' sociodemographic data and motivations to follow the class. This survey also allows instructors to include a few of their own questions.

We incorporated three items based on the use of LA in the form of 4-point Likert items in three MOOCs launched in 2022: Data and Critical Thinking, An Introduction to Data Science, and Blended Learning in Higher Education. The use of Likert items is a departure from the binary yes/no approaches that were adopted by Arnold & Sclater (2017), for instance. The authors designed items that limited respondents to a binary "yes" or "no" response—for example: "Would you be happy for data on your learning activities to be used if it kept you from dropping out or helped you get personalized interventions?" Likert items introduce more nuance in the responses (Joshi et al., 2015).

Likert Items

The first item laid emphasis on the type of data used by recommendation systems: "If you use a learning platform enough (like FUN MOOC or Coursera), it has a certain amount of data about you: your interests, skills, etc. Technically, this data can be transmitted to a number of stakeholders outside the platform (however, this is not the case for FUN). Out of the following data, what do you think about the fact that they could be transmitted to third parties (potential recruiters, etc.)?". The items listed were:

courses completed, key strengths, grades, areas of interest, learning habits, registrations, and dropped courses.

The second item was on the features of recommender systems: "Imagine you are browsing a MOOC catalogue, like FUN MOOC, and following classes. Within the platform, recommendation algorithms suggest personalized elements (courses, activities, learning partners, etc.) to improve your learning experience. What is your position on the fact that your data are mobilized to provide you with the following recommendations?" Five possibilities were then listed: grades, key strengths, courses completed, areas of interest, and learning habits.

The third item was about the data used by recommender systems: "The more you agree to share data about yourself, the more it is possible to improve the performance of recommendation algorithms (for courses, educational content, etc.). What is your position on sharing the following personal data in order to improve a recommender system?" Seven different types of data were then presented, ranging from courses completed to learning habits.

At the beginning of each question, we reminded the learners that the MOOC platform that hosted the courses did not use their data in any of the ways listed in the items. We wanted the respondents to understand that they were dealing with theoretical situations that did not apply directly to them in their MOOC.

Sample

The three courses were open for an entire year, from November 2021 to November 2022, allowing for continuous collection of surveys as users kept registering. Out of the 3,018 responses received throughout the year, 404 were incomplete, resulting in a total of 2,614 valid responses. Most of them came from the Data and Critical Thinking MOOC (62.5%) (Table 1). The sex ratio was close to 1:1, and 69.8% of respondents lived in France. For further analyses, we segmented age into 10-year intervals, beginning at 16. Learners over 66 years of age were grouped into a single category, representing approximately 10% of respondents. They were, however, excluded for further analyses since most were retired as per French law, and items on transmission of data to third parties like recruiters therefore proved irrelevant.

Table 1

Descriptive Statistics for Relevant Variables Among Respondents (N = 2,614)

Variable	Level	% of respondents
Age (years)	16–25	11.8
	26–35	18.6
	36–45	19.5
	46–55	22.3
	56–65	18.0
	66+	9.9
Gender	Male	51.1
	Female	48.2
	Other	0.6
Country of residence	France	69.8
	Other	30.2
MOOC (origin of responses)	Data and Critical Thinking	62.5
	Data Science and Its Stakes	6.6
	Blended Learning in Higher Education	30.8

Data Analysis

Various strategies exist for designing Likert items. One key dichotomy lies in the choice between even and odd numbers of levels. Researchers commonly favour a 5-point or 7-point scale (Dawes, 2008) when determining their agreement or disagreement with a given option. This approach, however, has been criticized for enabling undecided respondents to default to the median option (Dawes, 2008). An even number of points urges respondents to "pick a side"; hence, it was our chosen method. All three items regarding learners' perspectives on the use of their data were presented in 4-point Likert items.

Statistical analysis of Likert data has followed two distinct approaches over the years. The first involves converting the different scale levels into numbers. For example, the levels *somewhat favourable* and *favourable* might be represented by the numbers 3 and 4, respectively. This method facilitates the use of comprehensive models like linear regression, enabling comparison of the relative influence of independent variables such as gender and age. Conversely, some researchers prefer to treat the Likert item as categorical data. The approach of treating numeric data as if it were categorical has faced criticism for violating the fundamental assumptions of parametric models. Since our main independent variable was age, reducing the necessity for comprehensive models, we opted for the traditional Chi-square approach to Likert data (Croasmun & Ostrom, 2011). Age was categorized to facilitate this analysis. To visualize our results, we used the Likert function from the HH package in R 4.0.

Results

Below, we examine learners' views on recommender systems, concentrating on their attitudes towards various system features and the kind of data used to power them. Next, we present their views on data sharing from online course platforms with third parties, such as recruiters. In both scenarios, we demonstrate the correlation between the learners' age group and their perspectives.
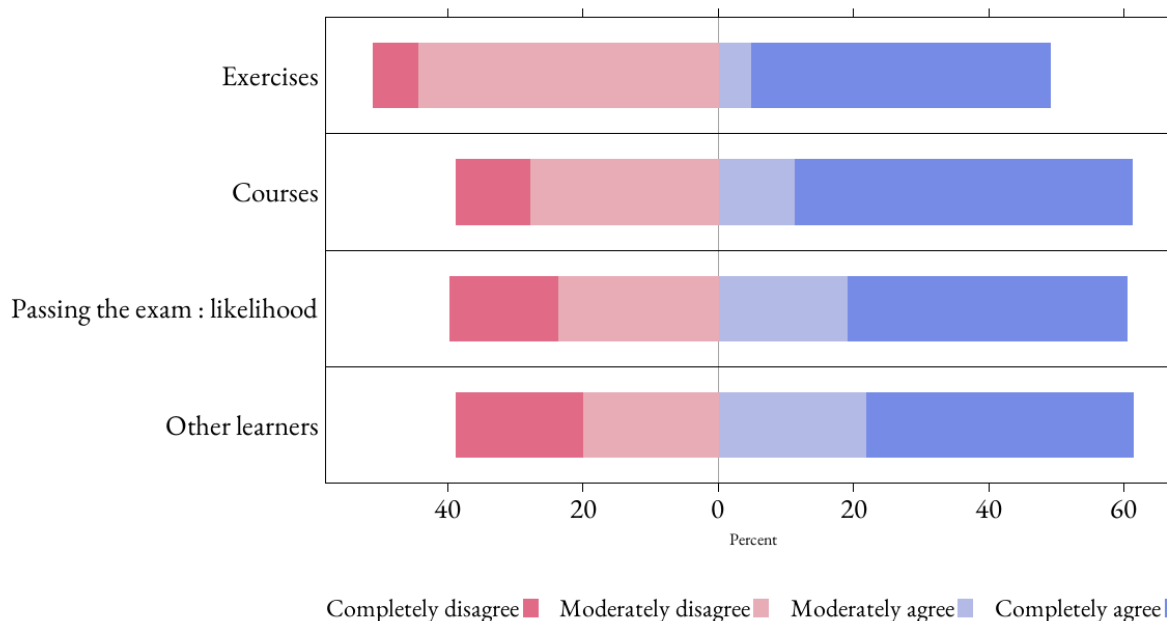
Recommender Systems

Even if there were statistically significant differences across the four items of the survey (X^2 [9, N = 2,614] = 846.03, p < 2.210^{-16}), respondents overall had favourable opinions about the use of a recommendation system for the different use cases that we listed. Exercise and course recommendation scored the highest (49.0% and 60.1% agreed, respectively) (Figure 1).

While respondents seemingly had a positive view of the existence of these algorithms in different situations, they "frowned upon" the fact that these algorithms used different types of data to make recommendations (Figure 2). Their position differed significantly depending on the type of data (X^2 [9, N = 2,614] = 130.8, p < 2.210^{-16}).

Figure 1

Posture Towards Different Features of the Recommender System (N = 2,614)

The use, in recommender systems, of the IDs of the videos that were viewed was the least controversial (55.6% agreed, and among them 40.7% completely agreed), while at the other extremity of the spectrum, the exploitation of learners' activity in course forums by recommender systems was rejected by the majority of respondents (50.6% disagreed overall, among which 37.6% completely disagreed) (Figure 2). Recommendation algorithms' mobilization of learning habits also appeared to generate stronger rejection than data such as registration data.

Figure 2

Posture Towards the Use of Different Types of Learner Data to Enable the Use of Recommendation Systems (N = 2,614)
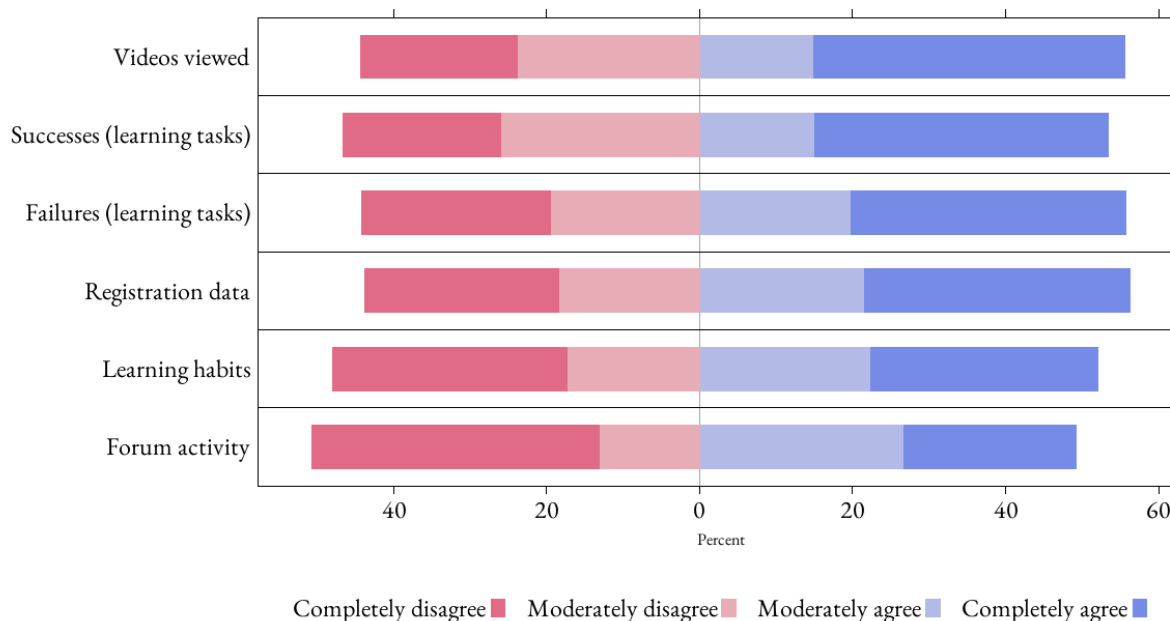
Respondents' age was strongly correlated with their perspective on the use of different data in the recommender systems, as illustrated by algorithms' use of data about learning habits (X^2 [12, $N = 2,614$] = 129.7, $p < 2.10^{-16}$). Younger learners overall appeared less reluctant to have their data used. For instance, 17.2% of learners in the 16−25 age category completely opposed the use of data on learning habits for any type of recommendation, while this proportion reached 36.4% of older users in the 56−65 age category (Figure 3). This pattern appeared to be consistent across all types of data used by recommendation systems, and it also held when it came to the transmission of learners' data to third parties, as we will see in the next section.

Figure 3

Participants' Age and Posture on the Use of Learning Habits Data in Recommendation Algorithms (N = 2,614)
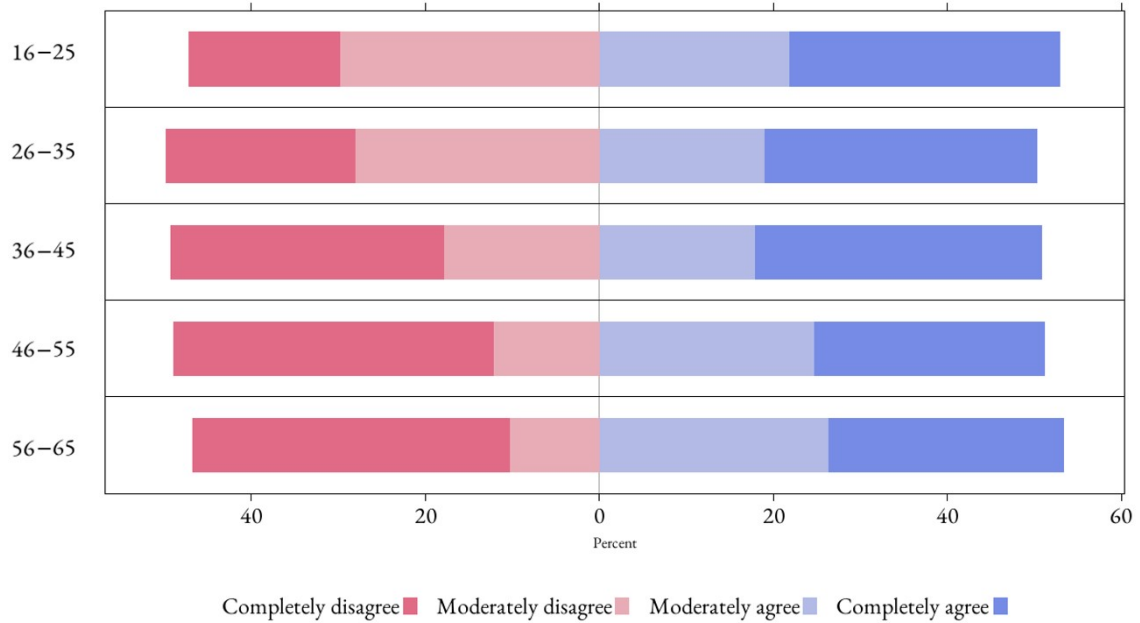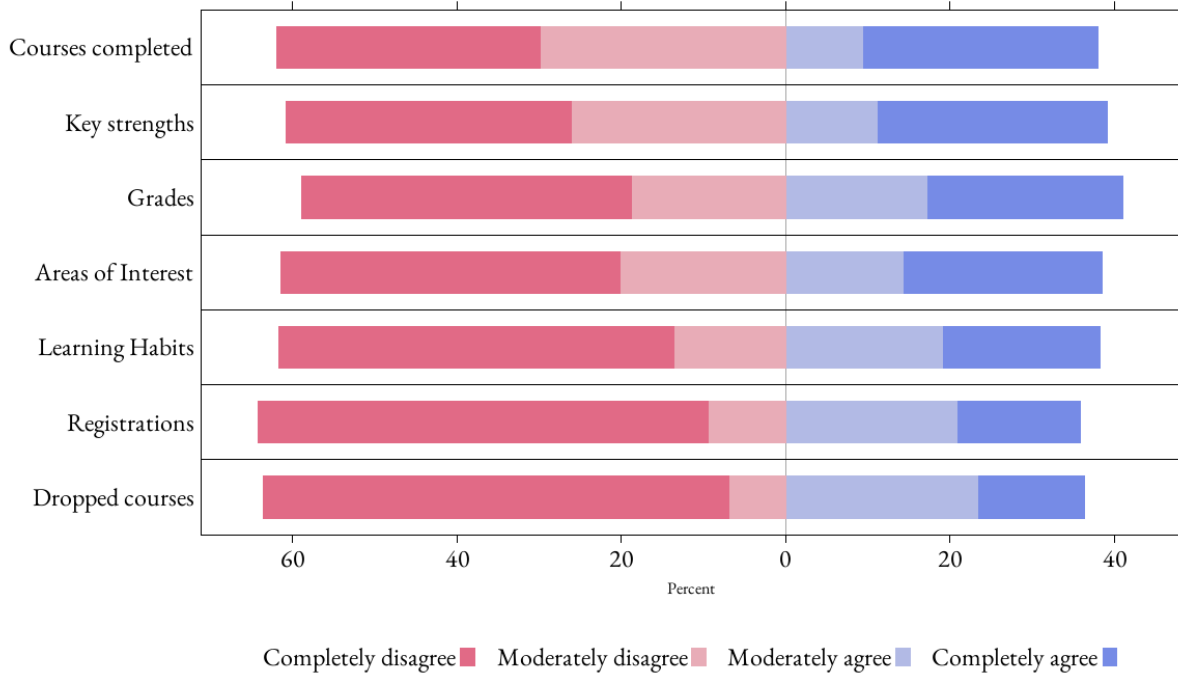

Transmission of Data to Third Parties and Generational Divides

Respondents overall opposed the possible transmission of their data to third parties, with a statistically significant difference across items (X^2 [9, N = 2,614] = 217.5, p < 2.2·10^{-16}) (Figure 4). This opposition was weaker for the list of courses that they had completed (only 33% completely disagreed) but appeared strong when it came to learning habits (49% completely disagreed) or to the different classes that they had registered to (55% completely disagreed).

Figure 4

*Learners' Stance on Data Transmission to Potential Recruiters (N = 2,614)*

Learners' posture on data transmission to third parties depended upon their age. We found statistically significant relationships between age and all types of data, with stronger effects for data such as grades (X^2 [12, $N = 2,614$] = 279.2, $p < 2.10^{-16}$; Figure 5) and learning habits (X^2 [12, $N = 2,614$] = 252.8, $p < 2.10^{-16}$; Figure 6).

Figure 5

Relationship Between Age and Posture on Grade Transmission Data to Potential Recruiters (N = 2,614)
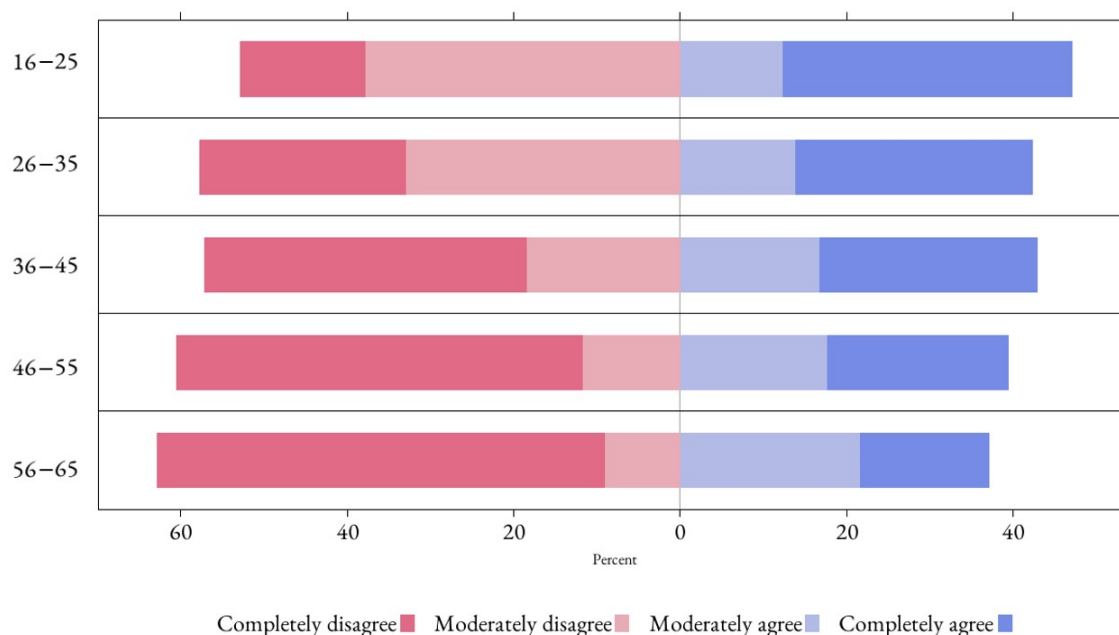
Younger generations (16–25 years) were much more open to the transmission of their grades (35% completely agreed; only 15% completely disagreed). Most respondents over 35 years old opposed the transmission of their grades; the ratio of respondents who completely disagreed to completely agreed reached 54% to 15% for those aged 56–65 (Figure 5). A similar trend was observed with regard to the transmission of data on learning habits. Among the youngest learners (Figure 6), 28% completely agreed while 26 % completely disagreed. For individuals aged 56–65, the former and the latter categories accounted for 12% and 61% of responses, respectively.

Figure 6

Relationship Between Age and Posture on Transmission of Learning Habits Data to Third Parties (Potential Recruiters, etc.; N = 2,614)
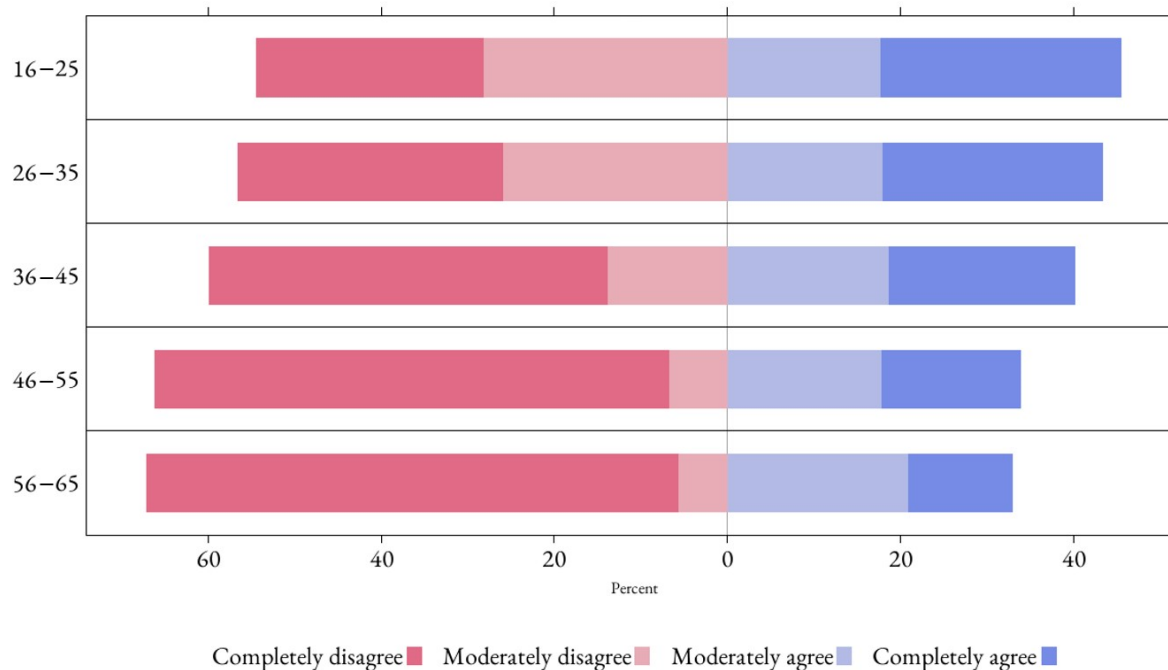
Discussion

We here discuss the questions of recommender systems and of transmission of data to third parties, which represent two typical uses of learners' data, on the one hand to improve the user experience and on the other hand to generate revenues. We then highlight the existence of a potential generational divide on data privacy, especially with regard to the transmission of data.

Recommender Systems

Like Arnold and Sclater (2017), we observed significant contrasts with regard to how respondents would react to a set of proposed use cases. The authors noticed that students were on average reluctant to the idea of their data being used to enable performance comparisons within the classroom, while, unsurprisingly, a minority would raise objections to these data being used to improve grades. Our study mirrored these findings. Participants generally exhibited favourable attitudes towards the different functionalities proposed, with one exception—exercise recommendation. A notable portion of the respondents expressed moderate disagreement with this case of data use, highlighting privacy concerns similar to those noted by Culnan and Armstrong (1999).

Drawing parallels with the concept of privacy calculus by Dinev and Hart (2006), we argue that students perform a cost–benefit analysis when determining the acceptability of data use. They weigh the potential benefits—such as improved grades—against the perceived privacy risks—such as the discomfort associated with performance comparisons if learning data were to be used outside of the context of recommender systems.

To address these variances in attitudes, MOOC platforms could consider offering a customizable user experience with regard to recommender systems. This approach aligns with Dinev's integrated model, in which perceived control over personal information is a key factor in privacy perceptions. As such, these platforms could allow learners to toggle specific features of the system. Some developers, such as Bousbahi and Chorfi (2015), have already taken steps in this direction by giving users the ability to select which data are fed into the recommendation algorithms. This approach affords a degree of personalization and imbues learners with a sense of control over their data, which, according to Culnan and Bies (2003), can effectively mitigate privacy concerns.

Consider a platform like Facebook as an example. Users on this social networking site are provided with the flexibility to determine their own privacy parameters. They can set rules on a variety of aspects such as who is permitted to view their profiles and personal details, the individuals who can locate them through the platform's search function, the methods through which others can contact them, and even the kind of personal narratives that get published on their profiles. Adjustable privacy settings seem to significantly influence users' comfort level regarding privacy. It has been posited that when individuals perceive that they have an adequate degree of control over their personal information, their concerns surrounding privacy typically diminish (Culnan & Armstrong, 1999).

Broadening user control over what is fed to the algorithms, as suggested by Bousbahi and Chorfi (2015), can significantly reduce resistance to adopting recommender systems in online learning platforms such as Coursera or FUN MOOC. When learners are given more autonomy over their data, they are likely to trust the system more, leading to higher engagement and more effective use of the recommender features. While it may be improbable for users to have a say in the decisions made by online learning stakeholders with regard to the use of these algorithms, providing them with a degree of control could prevent feelings of distrust. If users feel that their privacy is being violated, they may not necessarily stop using the platform, but they might choose to disregard the recommendations made by the system, rendering it less effective. Therefore, the sense of control over their data could not only ease privacy concerns but also boost the utility and relevance of the recommender systems.

Transmission of Data to Third Parties

Sharing data with third parties, such as potential recruiters, was once regarded as a promising business model for MOOCs and is now a practice that has been adopted by course platforms like Udacity (Allal-Chérif et al., 2021). Unsurprisingly, younger learners showed less reluctance than senior workers with regard to the transmission of their LA. This result is consistent with the generational divide that Miltgen and Peyrat-Guillard (2014) observed regarding data privacy in their qualitative study. Respondents entering the job market, in contrast to those with a well-established professional trajectory, might view data sharing as an opportunity to increase their visibility. However, third parties such as recruiters might be interested in data of which the transmission may not necessarily benefit learners.

Learning habits and details on uncompleted courses, whether due to lack of engagement or difficulties, could potentially offer employers deeper insights into an individual's profile. However, the transmission of this type of data would mean much weaker control for learners over the display of their activity on the course platform; it would represent a significant shift from the current situation, where individuals mostly show what classes they have engaged in (Allal-Chérif et al., 2021) or what certificates they have obtained (Radford et al., 2014), but not their grades, transmission of which could be informative for employers but harmful for potential applicants. Respondents' may be conscious of this misalignment,

which could explain why they appeared much more reluctant to share data on their learning habits than on their key strengths, for instance.

Arguably, in contrast with online platforms that focus exclusively on upskilling the workforce, generalist platforms like Coursera and FUN MOOC cover a diversity of topics that are arguably unrelated to learners' professional activity (poetry, etc.; Cisel, 2019). This means that these platforms potentially have detailed information on users' hobbies, as long as we consider that registering to a course signals someone's area of interest. Our results suggest that learners may be reluctant to the transmission of this type of data.

This is likely one of the unique aspects of generalist MOOC platforms: a learner might have different objectives depending on the type of course they enrol in (Kizilcec & Schneider, 2015). This is in contrast with career-oriented platforms—DataCamp, for example, which is focused on data science (Baumer et al., 2020). As Miltgen and Peyrat-Guillard (2014, p. 113) showed through focus groups, for similar types of data, individuals have a distinct stance on privacy depending on the context in which they could be used. Expectations regarding the management of their LA could vary significantly across different types of courses (e.g., career-oriented or hobby-oriented), adding complexity to the task of platforms striving to customize data management according to user preferences.

A Generational Divide in Terms of Data Privacy?

Concerning the correlation between learners' age and acceptance of the use of their data, our findings align with prior studies on users' viewpoints regarding data privacy (Miltgen & Peyrat-Guillard, 2014; Regan et al., 2013). While this generalized explanation may suffice for recommender systems, it seems less comprehensive when considering the question of data transmission to third parties, as it leaves essential queries unaddressed: How much do respondents' views hinge on their generational affiliation? How much do they rely on their age? Currently, it is challenging to determine whether the existing response structure will remain consistent over time.

The perceived value of visibility to potential recruiters—as understood in the Dinev et al.'s (2013) calculus framework—could largely hinge on the professional career stage that learners are in. For example, younger respondents who presently view data transmission favourably due to their quest for career opportunities may exhibit more reluctance once they have established themselves professionally. However, we might be witnessing a more lenient attitude towards data privacy from individuals born in the 1990s and 2000s, as they have been granting access to their data to various entities via their mobile devices or social networking accounts since their adolescence.

This debate echoes conversations around the cultural-historical activity theory (Roth & Lee, 2007), which asserts that an individual's cultural backdrop and the era in which they were born should be factored into the examination of their psychological traits and opinions. Distinguishing between these two competing explanations is crucial for predicting the evolution of citizens' perceptions of data privacy and its potential impact on future political decisions.

Conclusion

Limitations

This case study is not exempt from conventional scrutiny concerning the practical applicability of our results to the technological design and enhancement of the systems under consideration. Additionally, due to the specificity of our queries, our insights cannot be effortlessly generalized beyond the MOOC context. However, given that these courses now constitute a significant portion of the online education environment, we assert that despite the challenges related to the external validity of our findings, dedicated studies on ethical usage of online learners' analytics bear inherent value. A principal constraint of this research lies in our inability to comprehensively understand how learners perceive information privacy and retention at large, leaving us unable to determine if their views are unique to the educational context or not.

Perspectives

As MOOCs often host classes related to learners' hobbies, users may view data use on these platforms differently when compared with platforms that focus primarily on professional skills. Notably, platforms like DataCamp (Baumer et al., 2020) offer both courses and job opportunities, suggesting that learners' perceptions of recommender systems and data sharing with third parties like recruiters might differ when they view these platforms as both skill-development tools and career-advancement avenues.

Finally, contrasting respondents' viewpoints about their data use across different platforms (e.g., Facebook, Netflix, etc.), with the aim to determine if they adopt a more permissive attitude concerning their data usage in the context of MOOCs, would be enlightening. How would individuals who showed minimal hesitation about the use of their learning data react if other aspects of their lives were exposed? Would the generational divide be as pronounced on social media platforms?

References

Allal-Chérif, O., Aránega, A. Y., & Sánchez, R. C. (2021). Intelligent recruitment: How to identify, select, and retain talents from around the world using artificial intelligence. *Technological Forecasting and Social Change*, *169*, Article 120822. https://doi.org/10.1016/j.techfore.2021.120822

Arnold, K. E., & Sclater, N. (2017). Student perceptions of their privacy in leaning analytics applications. In *LAK '17: Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 66-69). Association for Computing Machinery. https://doi.org/10.1145/3027385.3027392

Baumer, B. S., Bray, A. P., Çetinkaya-Rundel, M., & Hardin, J. S. (2020). Teaching introductory statistics with Datacamp. *Journal of Statistics Education*, *28*(1), 89–97. https://doi.org/10.1080/10691898.2020.1730734

Bethlehem, J. (2010). Selection bias in Web surveys. *International Statistical Review*, *78*(2), 161–188. https://doi.org/10.1111/j.1751-5823.2010.00112.x

Bousbahi, F., & Chorfi, H. (2015). MOOC-Rec: A case based recommender system for MOOCs. *Procedia—Social and Behavioral Sciences*, *195*, 1813–1822. https://doi.org/10.1016/j.sbspro.2015.06.395

Burd, E. L., Smith, S. P., & Reisman, S. (2015). Exploring business models for MOOCs in higher education. *Innovative Higher Education*, *40*, 37–49. https://doi.org/10.1007/s10755-014-9297-0

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom research into edX's first MOOC. *Research & Practice in Assessment*, *8*, 13-25. https://eric.ed.gov/?id=ej1062850

Cerratto Pargman, T., & McGrath, C. (2021). Mapping the ethics of learning analytics in higher education: A systematic literature review of empirical research. *Journal of Learning Analytics*, *8*(2), 123–139. https://doi.org/10.18608/jla.2021.1

Cisel, M. T. (2018). Interactions in MOOCs: The hidden part of the iceberg. *International Review of Research in Open and Distributed Learning*, *19*(5). https://doi.org/10.19173/irrodl.v19i5.3459

Cisel, M. (2019). The structure of the MOOC ecosystem as revealed by course aggregators. *American Journal of Distance Education*, *33*(3), 212-227. https://doi.org/10.1080/08923647.2019.1610285

Croasmun, J. T., & Ostrom, L. (2011). Using Likert-type scales in the social sciences. *Journal of Adult Education*, *40*(1), 19–22. https://files.eric.ed.gov/fulltext/EJ961998.pdf

Culnan, M. J. (1993). "How did they get my name?" An exploratory investigation of consumer attitudes toward secondary information use. *MIS Quarterly*, *17*(3), 341–363. https://doi.org/10.2307/249775

Culnan, M. J., & Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science*, *10*(1), 104–115. https://www.jstor.org/stable/2640390

Culnan, M. J., & Bies, R. J. (2003). Consumer privacy: Balancing economic and justice considerations. *Journal of Social Issues*, *59*(2), 323–342. https://doi.org/10.1111/1540-4560.00067

Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in Education*, *2012*(3), Article 18. https://jime.open.ac.uk/articles/10.5334/2012-18

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, *50*(1), 61–104. https://doi.org/10.1177/14707853080500010

Dinev, T., & Hart, P. (2004). Internet privacy concerns and their antecedents-measurement validity and a regression model. *Behaviour & Information Technology*, *23*(6), 413–422. https://doi.org/10.1080/01449290410001715723

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, *17*(1), 61–80. https://www.jstor.org/stable/23015781

Dinev, T., Xu, H., Smith, J. H., & Hart, P. (2013). Information privacy and correlates: An empirical attempt to bridge and distinguish privacy-related concepts. *European Journal of Information Systems*, *22*(3), 295–316. https://doi.org/10.1057/ejis.2012.23

Ferguson, R. (2019). Ethical challenges for learning analytics. *Journal of Learning Analytics*, *6*(3), 25-30. https://doi.org/10.18608/jla.2019.63.5

Hakimi, L., Eynon, R., & Murphy, V. A. (2021). The ethics of using digital trace data in education: A thematic review of the research landscape. *Review of Educational Research*, *91*(5), 671-717. https://doi.org/10.3102/00346543211020116

Ho, A., Chuang, I., Reich, J., Coleman, C., Whitehill, J., Northcutt, C., Williams, J., Hansen, J., Lopez, G., & Petersen, R. (2015). HarvardX and MITx: Two years of open online courses fall 2012-summer 2014. *Available at SSRN 2586847*. http://dx.doi.org/10.2139/ssrn.2586847

Hollands, F. M., & Tirthali, D. (2014). Resource requirements and costs of developing and delivering MOOCs. *The International Review of Research in Open and Distributed Learning*, *15*(5). https://doi.org/10.19173/irrodl.v15i5.1901

Hsu, T.-C., Abelson, H., & Lao, N. (2022). Editorial—Volume 23, issue 1. *The International Review of Research in Open and Distributed Learning*, *23*(1), i–ii. https://doi.org/10.19173/irrodl.v22i4.5949

Hutt, S., Baker, R. S., Ashenafi, M. M., Andres-Bray, J. M., & Brooks, C. (2022). Controlled outputs, full data: A privacy-protecting infrastructure for MOOC data. *British Journal of Educational Technology*, *53*(4), 756–775. https://doi.org/10.1111/bjet.13231

Hwang, G.-J., Tu, Y.-F., & Tang, K.-Y. (2022). AI in online-learning research: Visualizing and interpreting the journal publications from 1997 to 2019. *The International Review of Research in Open and Distributed Learning*, *23*(1), 104–130. https://doi.org/10.19173/irrodl.v23i1.6319

Ifenthaler, D. (2017). Are higher education institutions prepared for learning analytics? *TechTrends*, *61*(4), 366-371. https://doi.org/10.1007/s11528-016-0154-0

Ifenthaler, D., & Tracey, M. W. (2016). Exploring the relationship of ethics and privacy in learning analytics and design: Implications for the field of educational technology. *Educational Technology Research and Development*, *64*(5), 877–880. https://doi.org/10.1007/s11423-016-9480-3

Jisc. (2015, June). *Code of practice for learning analytics*. https://www.jisc.ac.uk/sites/default/files/jd0040_code_of_practice_for_learning_analytics_190515_v1.pdf

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. https://doi.org/10.1002/asi.24358

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, *7*(4), 396–403. https://doi.org/10.9734/BJAST/2015/14975

Kizilcec, R. F., Saltarelli, A. J., Reich, J., & Cohen, G. L. (2017). Closing global achievement gaps in MOOCs. *Science*, *355*(6322), 251–252. https://doi.org/10.1126/science.aag2063

Kizilcec, R. F., & Schneider, E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Transactions on Computer–Human Interaction*, *22*(2), Article 6. https://doi.org/10.1145/2699735

Klopfer, P. H., & Rubenstein, D. I. (1977). The concept *privacy* and its biological basis. *Journal of Social Issues*, *33*(3), 52–65. https://doi.org/10.1111/j.1540-4560.1977.tb01882.x

Kolowich, S. (2013). The professors who make the MOOCs. *The Chronicle of Higher Education*, *18*, 1–12. https://www.chronicle.com/article/the-professors-behind-the-mooc-hype/

Laufer, R. S., & Wolfe, M. (1977). Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of Social Issues*, *33*(3), 22–42. https://doi.org/10.1111/j.1540-4560.1977.tb01880.x

Marshall, S. (2014). Exploring the ethical implications of MOOCs. *Distance Education*, *35*(2), 250–262. https://doi.org/10.1080/01587919.2014.917706

Milne, G. R., & Gordon, M. E. (1993). Direct mail privacy-efficiency trade-offs within an implied social contract framework. *Journal of Public Policy & Marketing*, *12*(2), 206–215. https://www.jstor.org/stable/30000091

Miltgen, C. L., & Peyrat-Guillard, D. (2014). Cultural and generational influences on privacy concerns: A qualitative study in seven European countries. *European Journal of Information Systems*, *23*(2), 103-125. https://doi.org/10.1057/ejis.2013.17

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

Porter, S. (2015). The economics of MOOCs: A sustainable future? *The Bottom Line*, *28*(1/2), 52–62. https://doi.org/10.1108/BL-12-2014-0035

Presthus, W., & Sørum, H. (2019). Consumer perspectives on information privacy following the implementation of the GDPR. *International Journal of Information Systems and Project Management*, *7*(3), 19–34. https://doi.org/10.12821/ijispm070302

Prinsloo, P., & Slade, S. (2015). Student privacy self-management: Implications for learning analytics. In *LAK '15: Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 83-92). Association for Computing Machinery. https://doi.org/10.1145/2723576.2723585

Prinsloo, P., & Slade, S. (2017). Ethics and learning analytics: Charting the (un)charted. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (pp. 49-57). Society for Learning Analytics Research. https://doi.org/10.18608/hla17.004

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/10.4324/9780203731864-6

Radford, A. W., Robles, J., Cataylo, S., Horn, L., Thornton, J., & Whitfield, K. (2014). The employer potential of MOOCs: A mixed-methods study of human resource professionals' thinking on MOOCs. *The International Review of Research in Open and Distributed Learning*, *15*(5). https://doi.org/10.19173/irrodl.v15i5.1842

Regan, P., Fitzgerald, G., & Balint, P. (2013). Generational views of information privacy? *Innovation: The European Journal of Social Sciences*, *26*(1–2), 81–99. https://doi.org/10.1080/13511610.2013.747650

Roth, W. M., & Lee, Y. J. (2007). "Vygotsky's neglected legacy": Cultural-historical activity theory. *Review of Educational Research*, *77*(2), 186–232. https://doi.org/10.3102/0034654306298273

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/10.1080/01972243.2016.1126142

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, *78*, 397-407. https://doi.org/10.1016/j.chb.2017.06.030

Sclater, N. (2016). Developing a code of practice for learning analytics. *Journal of Learning Analytics*, *3*(1), 16-42. https://doi.org/10.18608/jla.2016.31.3

Selwyn, N. (2019). What's the problem with learning analytics? *Journal of Learning Analytics*, *6*(3), 11–19. https://doi.org/10.18608/jla.2019.63.3

Stephens-Martinez, K., Hearst, M. A., & Fox, A. (2014). Monitoring MOOCs: Which information sources do instructors value? In *L@S '14: Proceedings of the First ACM Conference on Learning @ Scale Conference* (pp. 79–88). Association for Computing Machinery. https://doi.org/10.1145/2556325.2566246

Stone, E. F., & Stone, D. L. (1990). Privacy in organizations: Theoretical issues, research findings, and protection mechanisms. In K. M. Rowland & G. R. Ferris, *Research in personnel and human resources management* (Vol. 8, pp. 349–411). JAI Press Inc.

Tough, A. (1971) *The adult's learning project*. Ontario Institute for Studies in Education.

Tzimas, D., & Demetriadis, S. (2021). Ethical issues in learning analytics: A review of the field. *Educational Technology Research and Development*, *69*(2), 1101–1133. https://doi.org/10.1007/s11423-021-09977-4

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

West, D., Huijser, H., & Heath, D. (2016). Putting an ethical lens on learning analytics. *Educational Technology Research and Development*, *64*(5), 903–922. https://doi.org/10.1007/s11423-016-9464-3

West, D., Luzeckyj, A., Searle, B., Toohey, D., Vanderlelie, J., & Bell, K. R. (2020). Perspectives from the stakeholder: Students' views regarding learning analytics and data collection. *Australasian Journal of Educational Technology*, *36*(6), 72–88. https://doi.org/10.14742/ajet.5957

Whitelock-Wainwright, A., Gašević, D., Tejeiro, R., Tsai, Y. S., & Bennett, K. (2019). The student expectations of learning analytics questionnaire. *Journal of Computer Assisted Learning*, *35*(5), 633–666. https://doi.org/10.1111/jcal.12366

Whitelock-Wainwright, A., Gašević, D., Tsai, Y. S., Drachsler, H., Scheffel, M., Muñoz-Merino, P. J., Tammets, K., & Delgado Kloos, C. (2020). Assessing the validity of a learning analytics expectation instrument: A multinational study. *Journal of Computer Assisted Learning*, *36*(2), 209–240. https://doi.org/10.1111/jcal.12401

Whitelock-Wainwright, A., Tsai, Y. S., Drachsler, H., Scheffel, M., & Gašević, D. (2021). An exploratory latent class analysis of student expectations towards learning analytics services. *The Internet and Higher Education*, *51*, Article 100818. https://doi.org/10.1016/j.iheduc.2021.100818

Wintermute, E. H., Cisel, M., & Lindner, A. B. (2021). A survival model for course-course interactions in a Massive Open Online Course platform. *PloS one*, *16*(1), e0245718. https://doi.org/10.1371/journal.pone.0245718

Zimmermann, C., Kopp, M., & Ebner, M. (2016). How MOOCs can be used as an instrument of scientific research. In *Proceedings of the European Stakeholder Summit on experiences and best practices in and around MOOCs (EMOOCS 2016)* (pp. 393–400). BoD. https://graz.elsevierpure.com/en/publications/how-moocs-can-be-used-as-an-instrument-of-scientific-research

Athabasca University