To Opt In or To Opt Out? Predicting Student Preference for Learning Analytics-Based Formative Feedback

Merikko, Joonas

2022-09-16

Merikko , J , Ng , K , Saqr , M & Ihantola , P 2022 , ' To Opt In or To Opt Out? Predicting Student Preference for Learning Analytics-Based Formative Feedback ' , IEEE Access , vol. 2022 , no. 10 , pp. 99195-99204 . https://doi.org/10.1109/ACCESS.2022.3207274

http://hdl.handle.net/10138/349609
https://doi.org/10.1109/ACCESS.2022.3207274

Received 12 August 2022, accepted 8 September 2022, date of publication 16 September 2022,
date of current version 26 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3207274

RESEARCH ARTICLE

To Opt in or to Opt Out? Predicting Student Preference for Learning Analytics-Based Formative Feedback

JOONAS MERIKKO[1], KWOK NG[2,3], MOHAMMED SAQR[4], AND PETRI IHANTOLA[1]
[1]Department of Education, Faculty of Educational Sciences, University of Helsinki, 00014 Helsinki, Finland
[2]Philosophical Faculty, School of Educational Sciences and Psychology, University of Eastern Finland, 80101 Joensuu, Finland
[3]Physical Activity for Health Research Cluster, Department of Physical Education and Sport Sciences, University of Limerick, Limerick, V94 T9PX, Ireland
[4]Faculty of Science and Forestry, School of Computing, University of Eastern Finland, 80101 Joensuu, Finland

Corresponding author: Joonas Merikko (joonas.merikko@gmail.com)

The work of Mohammed Saqr was supported in part by the Academy of Finland Project Topeila under Grant 350560. The work of Petri Ihantola was supported by the Ministry of Education and Culture in Finland under Grant OKM/307/522/2020.

ABSTRACT Teachers' work is increasingly augmented with intelligent tools that extend their pedagogical abilities. While these tools may have positive effects, they require use of students' personal data, and more research into student preferences regarding these tools is needed. In this study, we investigated how learning strategies and study engagement are related to students' willingness to share data with learning analytics (LA) applications and whether these factors predict students' opt-in for LA-based formative feedback. Students (N = 158) on a self-paced online course set their personal completion goals for the course and chose to opt in for or opt out of personalized feedback based on their progress toward their goal. We collected self-reported measures regarding learning strategies, study engagement, and willingness to share data for learning analytics through a survey (N = 73). Using a regularized partial correlation network, we found that although willingness to share data was weakly connected to different aspects of learning strategies and study engagement, students with lower self-efficacy were more hesitant to share data about their performance. Furthermore, we could not sufficiently predict students' opt-in decisions based on their learning strategies, study engagement, or willingness to share data using logistic regression. Our findings underline the privacy paradox in online privacy behavior: theoretical unwillingness to share personal data does not necessarily lead to opting out of interventions that require the disclosure of personal data. Future research should look into why students opt in for or opt out of learning analytics interventions.

INDEX TERMS Feedback, learning strategies, opt-in, privacy, self-regulation, study engagement, teaching augmentation.

I. INTRODUCTION

One of the most important tasks of a teacher is to support students' self-regulated learning (SRL) skills; in other words, how learners systematically activate and sustain their cognition, motivation, productive behavior, and positive affects toward the attainment of their learning goals [1]. In a traditional classroom setting, teachers support and promote learners' self-regulation by constantly monitoring how the learning process is going and providing personalized guidance and feedback for the learners. Especially, formative feedback, i.e., feedback on how to perform a task more effectively, is essential in developing self-regulatory skills [2], [3].

Student-teacher interaction has become increasingly mediated through digital platforms and data artifacts. The methods for supporting and promoting self-regulation need to be adjusted when there are no face-to-face interactions [4]. Instead, teachers monitor students' learning processes by investigating online data traces and provide feedback to students with messages in online learning environments. The

The associate editor coordinating the review of this manuscript and approving it for publication was Laxmisha Rai.

nature of the interaction is often asynchronous, as learners and teachers may perform their tasks at different times. These self-paced settings afford learners more autonomy [5], [6] but increase the need for self-regulation and complicate teachers' chances to support and guide students.

On the other hand, teachers' work is increasingly augmented with intelligent tools that extend their pedagogical abilities. When a real-time interaction with students is scarce or nonexistent, feedback becomes informed by data artifacts such as assignments, forum posts, and log data on students' online learning behavior. The teacher can analyze these data manually and with the help of technological tools (i.e., teaching augmentation [7]). These tools allow a scale of education otherwise unattainable (e.g., MOOCs). However, while using such technologies, besides their effectiveness, it is essential to consider how students welcome them.

A. AUGMENTING TEACHING WITH LEARNING ANALYTICS-BASED FORMATIVE FEEDBACK

In an increasingly digital world, teachers' job is less about helping students acquire knowledge and more about designing digital environments and making meaning of the data students produce [8]. Teachers use their pedagogical knowledge to define which data to collect and how to analyze them to support learning [9]. Technological tools help teachers with this process, and learning analytics (LA)-based personalized feedback is an excellent example of new technologies augmenting teachers' capabilities [10].

In LA-based personalized feedback, the teacher uses technology to provide customized feedback to students when it would be ordinarily impossible (e.g., in courses with hundreds or thousands of students). Rather than writing personal feedback messages to each student, the teacher writes feedback templates, i.e., IF-THIS-THEN-THAT scripts, which the system converts to personalized feedback using data about students and their learning processes [11].

Receiving personalized feedback with this approach has been associated with a positive impact on student perception of feedback quality [10] and higher academic achievement [12]. Furthermore, Lim and colleagues [12] found patterns of self-regulated learning (SRL) to differ between students who received personalized feedback and those who did not. Iraj and colleagues [13] found that early engagement with feedback was positively associated with passing the course and that most students found feedback messages helpful in their learning. Also, the large number of feedback messages combined with technological tools such as tracking links may be used to study students' engagement with the feedback, providing teachers with insights on feedback quality [13]. Such findings encourage pursuing an in-depth understanding of automatic feedback as an effective tool for teachers.

In a systematic literature review of automatic feedback, Cavalcanti and colleagues found that most studies showed no evidence of manual feedback being more efficient than automatic feedback [14]. We argue that challenges in LA-based personalized feedback are mainly the challenges of feedback in general. Both the feedback sender and the receiver contribute to the success of feedback. Jonsson [15] has presented five reasons why students may not use feedback: (1) it may not be useful; (2) it may not be sufficiently individualized; (3) it may be too authoritative; (4) students may lack strategies for using feedback; and (5) students may not understand the terminology used. Furthermore, Price and colleagues [16] suggest that students' 'readiness to engage' (i.e., motivation to receive feedback, emotional response, and assessment literacy skills) may contribute to the engagement with feedback. For example, students may misinterpret the feedback if they do not understand the difference between formative and summative assessment [16]. However, some challenges are unique to LA-based personalized feedback. For example, while automation decreases the human effort needed for providing feedback, there might be a temptation to send feedback more often than would be optimal from a student's perspective. Furthermore, Lim and colleagues [17] found that some students with a study strategy based on offline activities felt that the feedback overemphasized engagement with online learning tasks.

Finally, it is essential to note that feedback may also have a negative impact. Especially, feedback administered in a controlling manner may harm intrinsic motivation [18]. While feedback is one of the most potential areas where intelligent technologies may augment teachers' abilities, caution is needed not to scale up any adverse effects.

B. STUDENT PERSPECTIVES ON USING THEIR DATA

While students are increasingly aware of data mining used to monitor and influence buying behavior [19], they are not necessarily expecting the same in an educational context [20]. Students trust not-for-profit higher education institutions more than for-profit corporations and are comfortable with practices in a university setting that they were skeptical of in a corporate environment [21]. Interesting questions are how aware students are about education institutions using their data, which data and for which purpose they are willing to share with the institution, and whether students have agency in deciding how their information is used.

Jones and colleagues [21] found that undergraduate students at U.S. higher education institutions lack awareness of analytic practices and the data they rely on. Several students encountered the idea of the university collecting and analyzing information about them for the first time during the study [21]. In Australia, Roberts and colleagues [20] found that most students were unaware or unsure of what big data and learning analytics were. Furthermore, in Finland, Teräs and colleagues [22] found that most students did not know what data their institution collected from them and what purposes the data were used for.

Students' approval for using their data varies depending on which data are used and for what purpose. Ifenthaler and Schumacher [23] found that students, in general, were open to sharing data related to their university studies but were more skeptical about sharing information about their online behavior. The willingness to share data positively predicted the usage of learning analytics systems, and participants were open to sharing more data if the analytics system provided rich and meaningful information [23]. Arnold and Sclater [24] found in a survey carried out in the U.K. and American institutions that 71%–94% of students would be happy to provide their data if they were used for improving grades, 53%–76% if the data were used for preventing dropping out, and 25%–61% if the data were used for social comparison. American students showed higher acceptance than U.K. students throughout different use cases, underlining the cultural differences. Furthermore, Bennett and Folley [25] found that students wanted to know the specific source of data being ingested and analyzed by learning analytics dashboards and required that data be used for educational purposes. Tsai, Whitelock-Wainwright, and Gasevic [26] found that students were protective toward personal data and had high expectations of how the university should process their data. However, their actions to protect personal data did not reflect such awareness. This privacy paradox [27] is a common phenomenon in online privacy behavior. In theory, users are concerned about privacy but still choose to disclose personal data because finding the risk acceptable or the benefits outweigh the risks.

Much of the learning analytics literature reflects an academic, teacher-centric, or institutional view [28]. The field has been criticized for treating students as subjects instead of autonomous individuals with their values and interests [29]. For example, students expect to be able to choose whether to opt in for or opt out of learning analytics [20], [26], [30]. Still, some institutions deny students the option to opt out based on an argument that because the data may benefit students, there is an obligation to use it [21]. Moreover, Prinsloo and Slade [31] argue that the binary opt-in/opt-out discussion is too narrow – students should be seen as active collaborators in the harvesting, analysis, and use of their data [30], [32]. For example, students should be able to modify the features shown on a dashboard [33] and review opt-in decisions as their experience at the university increases [26].

Several things have been found when involving students in the design of learning analytics. In Germany, Schumacher and Ifenthaler [34] found that the most wanted learning analytics features students want to use for their studies were reminders for deadlines, a feature helping revise the learning content of former semesters, and prompts for self-assessment. Students often expect simple and practical rather than analytically sophisticated features. In Finland, Silvola and colleagues [35] investigated students' needs concerning LA-based support for student engagement. They found that students typically suggested the role of LA as mediating information between the student and institution (behavioral engagement), increasing students' awareness of themselves as learners (cognitive engagement), providing support in challenging situations (emotional engagement), and helping students adapt their learning conditions according to individual needs (agentic engagement) [35].

Furthermore, there are examples where student agency is an inherent feature of a learning analytics tool. Jivet and colleagues [36] allowed learners to choose which indicators to monitor on a learning analytics dashboard. This tool design revealed unexpected results: while authors expected most learners to choose six indicators (maximum number of indicators allowed), more than half of the learners chose less. Moreover, neither the number of indicators selected nor the percentage of learning behavior indicators was associated with self-reported learner goals or self-regulated learning skills [36].

Overall, there is a tendency toward greater student participation in the deployment of learning technologies that utilize student data. Opt-in procedures are only one yet essential part of student agency in LA.

C. AIMS OF THE CURRENT STUDY

Technological tools and learning analytics (LA) augment teachers' capabilities for providing formative feedback to support students' self-regulated learning. According to Hattie and Timperley [3], feedback regarding the processing of the task and self-regulation is more powerful than feedback regarding the task or the student as a person. Furthermore, Lim and colleagues [17] suggest providing students feedback on their time management and learning strategies to enhance personalized feedback further.

There is an increasing focus on student agency and ethical aspects of LA interventions in LA literature. Opt-in procedures with an opportunity to review one's decisions later are recommended [26]. However, little is known about how students use the possibilities for opting in or out of LA interventions and how different student characteristics are associated with the opt-in behavior. In the present study, we addressed this gap by investigating the interactions between student characteristics and opt-in behavior in a setting where students can manage their opt-in as self-service.

The first aim of our study was to examine the complex interactions between learning strategies, study engagement, and students' willingness to share data with LA applications. We were especially interested in whether it matters what data the students are requested to share: are different components of learning strategies and study engagement associated with willingness to share specific data types? We also investigated, to what extent students are willing to share their data. We hypothesized that students are less willing to share fine-grained data revealing their behavior than other data types.

The second aim of our study was to understand how the willingness to share data, learning strategies, and study engagement affect students' opt-in for LA-based formative feedback. Acknowledging the privacy paradox, students' low self-reported willingness to share data does not necessarily lead to opting out of LA interventions. We investigated whether students' self-reported willingness to share data, learning strategies, or study engagement predicted their initial opt-in for LA-based formative feedback or the choice to change their opt-in status later during the course.

II. METHODS

A. CONTEXT

The context of the present study is an undergraduate-level self-paced online course in business studies offered by a Finnish research university. The course was provided entirely online using the Moodle learning management system (LMS) and included three modules, each with an automatically assessed multiple-choice online exam (30 or 35 points each, a total of 100 points). Students could enroll and complete the throughout the academic year 2020–2021, but the scope of this research is the Fall semester of 2020.

In previous years, students often registered for the course early but started studying only when the end of the semester was approaching. The aim of the intervention described next (the academic year 2020–2021) was to reduce the average course completion time and increase the completion rate. Besides the intervention, the course design and materials were not changed compared to previous academic years.

B. INTERVENTION: LA-BASED FORMATIVE FEEDBACK

In this study, we used data on students' course progress and time management goals to provide the students with personalized formative feedback. Before accessing the course materials, we required students to fill in a course enrollment questionnaire in the Moodle LMS, where they set their personal goals by stating which month they intend to have the course completed and selected whether they wanted to receive personalized feedback. If opted-in, they would get individual emails regularly, encouraging them to keep on track with their completion goal (e.g., 'looks like you are behind the schedule you planned, here are some suggestions what to do next'). Students were also encouraged to resubmit the course enrollment questionnaire if they wanted to change their completion goal or feedback preference.

The course teacher used OnTask [10] to provide students personalized feedback on five feedback rounds during the semester. At each feedback round, the teacher created messages for meaningful combinations of completion goals (5 alternatives) and course progress (6 alternatives), (See Table 1 for examples). The number of feedback messages received by each student depended on the particular timing of course enrollment and completion, as the feedback rounds were fixed in time.

C. PARTICIPANTS

During the fall semester of 2020, 158 students started the self-paced online course. Of these students, 74 (47%) responded to the research survey. The age of the respondents varied between 20 and 54 ($M = 29.3$, $SD = 9.2$). Regarding gender, 47 (64%) identified themselves as female, 26 (35%) male, and 1 (1%) other. Students were in different phases within their studies: 32 (43 %) had started their studies 2019–2020, 27 (36 %) 2017–2018 and 15 (20 %) 2016 or earlier. Informed consent was obtained from the participants.

D. MEASURES

1) SURVEY

We sent a survey via email to the participants at the beginning of the Fall 2020 semester, before the students had access to the course materials. The survey consisted of items from the Motivated Strategies for Learning Questionnaire (MSLQ) [37], the Schoolwork Engagement Inventory, [38] and the Sharing of Data Questionnaire (SOD) [23].

From the Motivated Strategies for Learning Questionnaire (MSLQ) we used three sections: Self-efficacy for Learning and Performance (8 items, e.g. 'I expect to do well in this class'), Metacognitive Self-Regulation (12 items, e.g.,, 'If course materials are difficult to understand, I change the way I read the material.'), Time and Study Environment (8 items, e.g. 'I make good use of my study time for this course.'). Items were rated on a scale ranging from 1 ('not at all true of me') to 7 ('very true of me').

The Schoolwork Engagement Inventory (EDA) consists of nine items that load onto three factors: energy (3 items; e.g., 'When I study, I feel I'm bursting with energy'), dedication (3 items; e.g., 'I find my studies full of meaning and purpose') and absorption (3 items; e.g., 'Time flies when I'm studying'). The items were rated on a scale ranging from 1 ('Totally disagree') to 6 ('Totally agree').

The Sharing of Data (SOD) questionnaire focuses on whether students are willing to disclose specific data types (28 items; e.g., 'medical information,' 'records of my downloads in the learning environment') to a learning analytics system ('Please indicate whether you would agree to disclose the following data for a Learning Analytics system'). Items were rated on a Thurstone scale from 0 ('I do not agree') to 1 ('I agree').

2) COURSE ENROLLMENT QUESTIONNAIRE

Before starting the first module of the course, the students needed to fill out a course enrollment questionnaire where they selected their *Course Completion Goal* ('The course can be completed on your own schedule. In which month do you plan to complete the whole course?') and *Opt-in for Personalized Feedback* ('I WOULD LIKE to receive automated personalized feedback messages' or 'I DO NOT want to receive automated personalized feedback messages'). As students could change their answers throughout the course, we created two dummy variables: *Initial opt-in* and *Opt-in changed*.

E. ANALYSES

We applied exploratory factor analysis with generalized least squares estimation methods and oblique promax rotation to Sharing of Data (SOD) items using JASP software (Version 0.16) [39]. We found a factor structure with four factors. The first factor ('Performance data') was comprised of ten

TABLE 1. Examples of OnTask feedback templates for the feedback round at the end of September.

Completion Goal	Course Progress	Feedback message[1]
October	Enrollment questionnaire completed	Hello {studentname}! You announced that you plan to complete the course in October at the latest. I noticed that you have not yet completed the first module of the course. Have there been any changes to your plans? If you start now, you can still complete the course as planned. The Powerpoint slides found in Moodle help you check the things in the paragraphs of the book. Kind regards, {teachername}
October	Second module completed	Hello {studentname}! You announced your intention to complete the course by October. You've already completed two modules, so you're already ahead of schedule. Well done! There is only one final module and the related exam left. Kind regards, {teachername}*

[1] {studentname} and {teachername} were replaced by actual names in each individual feedback message.

items (e.g. 'school history records', 'motivation questionnaire results') that explained 19.1% of the variance with factor loadings from.52 to 1.07. The second factor ('Process data') was comprised of seven items (e.g. 'records of my online times', 'records of my online user paths') that explained 13.5% of the variance with factor loadings from.44 to.80. The third factor ('Demographic data') was comprised of four items (e.g. 'name', 'information about employment during studies') that explained 10.4% of the variance with factor loadings from.52 to.90. Fourth and final factor ('Sensitive data') was comprised of four items (e.g. 'medical information, 'information about the family: marital status, names of children, etc.') that explained 8.0% of the variance with factor loadings from.49 to.83.

We calculated measures for internal consistency (Cronbach's alpha) for all sum variables using the psych package in R [40] and found them acceptable ($\alpha > 0.7$). Moreover, we ran Pairwise Pearson correlations between the sharing of data factors and other survey variables (i.e., self-efficacy, metacognitive self-regulation, time and study environment, energy, dedication, and absorption) using JASP (0.16).

Regarding our first research aim, we calculated a partial correlation network using LASSO regularization and EBIC model selection [41] using the Network feature with EBICglasso estimator in JASP. A partial correlation network shows correlated variables after controlling for all other variables in the network, i.e., dependent on each other. Such networks are beneficial when studying complex relationships between constructs and have recently been used in studying self-regulated learning [42], [43]. While using traditional metrics such as betweenness and closeness is often problematic in such network [44], we use degree and expected influence [45] to describe the centrality of different nodes.

Regarding our second research aim, we ran Mann-Whitney U-tests in JASP to determine any statistically significant differences between the opt-in and opt-out groups and opt-in changed and no changes groups. Then, we built binomial logistic regression models with *Initial opt-in* as a dependent variable using the glm function in the R stats package. As there were considerably more students opting in than opting out, we built models with original data and weighted data, where we weighed students opting out using a 4:1 ratio. We first built a naive model using all the survey variables as predictors and then chose a model by AIC in a stepwise algorithm using the stepAIC function in R. This resulted in a total of four models (original naive, original stepwise, weighted naive, and weighted stepwise). Sensitivity, specificity, and balanced accuracy metrics were calculated for each model. A small number of students changing their opt-in status during the course prevented building a model with *Opt-in changed* as a dependent variable.

III. RESULTS

A. PRELIMINARY RESULTS

Most students (N = 121, 77% initially; N = 114, 72% after changes during the semester) opted in for LA-based feedback regarding their progress toward their goals. Eight students who initially opted in decided to opt out later, whereas one student who initially opted out chose to opt in later.

In Table 2 we present the descriptive statistics, internal consistencies (i.e., Cronbach's alphas), and pairwise correlations of the measures used in the study. Correlation analyses revealed multiple significant correlations between the self-report measures. All six variables regarding learning strategies and study engagement were positively correlated ($r \geq .40$). Moreover, nearly all sharing of data variables were positively correlated ($r \geq .35$), a nonsignificant connection between Demographic data (DEM) and Sensitive data (SEN) being the only exception.

Regarding the sharing of data variables, students are, in general, more willing to share demographic data ($M = 0.62, SD = 0.37$) and performance data ($M = 0.60, SD = 0.36$) compared to process data ($M = 0.44, SD = 0.35$) and especially sensitive data ($M = 0.21, SD = 0.29$), which students are very hesitant to share. Students' willingness to disclose data about their performance was positively correlated with self-efficacy ($r = .34$) and all of the study engagement variables ($r \geq .24$). Moreover, willingness to disclose process data was positively correlated with absorption ($r = .25$) and willingness to share sensitive data with self-efficacy ($r = .25$).

Students who opted in for LA-based formative feedback had, on average, slightly higher scores in most of the learning strategies, study engagement, and sharing of data variables compared to students who opted out (See Table 3). However, none of these differences were statistically significant

TABLE 2. Correlations, descriptive statistics, and measures of internal consistency.

	SEL	MET	TIM	ENE	DED	ABS	PER	PRO	DEM	SEN
Motivated Strategies for Learning										
Self-efficacy (SEL)	1									
Metacognitive Self-Regulation (MET)	.55***	1								
Time and Study Environment (TIM)	.43***	.68***	1							
Schoolwork Engagement										
Energy (ENE)	.45***	.45***	.50***	1						
Dedicaton (DED)	.48***	.51***	.48***	.77***	1					
Absorption (ABS)	.40***	.48***	.54***	.77***	.79***	1				
Sharing of Data										
Performance data (PER)	.34**	.19	.23	.26*	.26*	.24*	1			
Process data (PRO)	.13	.19	.20	.21	.15	.25*	.63***	1		
Demographic data (DEM)	.14	0.01	.08	.03	-.00	-.04	.51***	.35**	1	
Sensitive data (SEN)	.25*	-.02	.07	.10	-.02	-.07	.49***	.41***	.23	1
N	*73*	*72*	*72*	*72*	*72*	*72*	*71*	*71*	*71*	*71*
M	*4.84*	*4.27*	*4.82*	*3.94*	*4.40*	*3.77*	*.60*	*.44*	*.62*	*.21*
SD	*1.16*	*1.02*	*0.99*	*1.04*	*1.03*	*1.22*	*.36*	*.35*	*.37*	*.29*
Cronbach's alpha	*.91*	*.88*	*.76*	*.86*	*.89*	*.86*	*.92*	*.85*	*.79*	*.71*

$^{*}\ p < .05$, $^{**}\ p < .01$, $^{***}\ p < .001$

as measured by Mann–Whitney U-test. Comparing students who changed their opt-in status during the course to students who did not, there was a significant difference in self-efficacy ($U = 71.5, p = .03, d = .58$) and dedication ($U = 69.5, p = .03, d = .59$). Students who changed their preference reported on average higher self-efficacy ($M = 5.88$, $SD = 0.67$) and dedication ($M = 5.33$, $SD = 0.85$) compared to students with no changes ($M = 4.76$, $SD = 1.16$ and $M = 4.33$, $SD = 1.01$, respectively). Regarding other measures, there were no significant differences between the groups.

B. COMPLEX INTERACTIONS BETWEEN LEARNING STRATEGIES, STUDY ENGAGEMENT, AND WILLINGNESS TO SHARE DATA

We calculated a partial correlation network (See figure 1) to answer our first research question. Willingness to share performance data, absorption and dedication are the most central nodes of the network based on centrality metrics (degree and expected influence). In contrast, the least central nodes are willing to share demographic and sensitive data. Two rather distinct groups of nodes can be distinguished: motivated strategies for learning and schoolwork engagement nodes form one group, and sharing of data nodes form another group. When looking at these two groups distinctly, each is highly interconnected. Among the sharing of data group, willingness to share performance data is the most central node, connecting with all the other sharing of data nodes ($.12 \leq r_p \leq .29$). Dedication is the most central node among the schoolwork engagement and motivated strategies for the learning group, connecting with four other nodes in the group ($.07 \leq r_p \leq .39$).

The willingess to share performance data node acts as a bridge between the two groups, being positively connected with self-efficacy ($r_p = .12$), dedication ($r_p = .04$) and energy ($r_p = .03$) nodes. Moreover, there is a weak connection between the self-efficacy node and sharing of sensitive data node ($r_p = .03$). To conclude, willingness to share data is rather weakly connected to different aspects of learning strategies and study engagement based on a partial correlation network, the strongest connection being between self-efficacy and sharing of performance data.

C. PREDICTING STUDENTS' OPT-IN FOR LA-BASED FORMATIVE FEEDBACK

The binomial regression models to predict students' initial opt-in for LA-based formative feedback are presented in Table 4. The first model (no weighting, naive) attempted to predict students' initial opt-in based on all survey metrics. Looking at the model coefficients, none of the predictors is statistically significant, and the model accuracy is very weak (balanced accuracy 0.49). Using a model selection algorithm to select variables for the second model (no weighting, stepwise AIC), a null model with only the intercept was created (balanced accuracy 0.50).

As the opt-in and opt-out groups were imbalanced (see Table 3), we also created weighted models where students opting out were weighted using a 4:1 ratio. The third model (weighted, naive) had a slightly higher balanced accuracy (0.65) than the first model, but none of the predictors were statistically significant. Using the model selection algorithm, we selected five predictors for the fourth model: self-efficacy, metacognitive self-regulation, energy, dedication, and willingness to share demographic data, of which metacognitive self-regulation was the only significant predictor ($p = 0.04$). While the fourth model had the highest balanced accuracy (0.67) of our models, it can still be considered weak.

TABLE 3. Group comparisons.

var	Opt in N	M	SD	Opt out N	M	SD	Mann–Whit. U	p	Opt in not changed N	M	SD	Opt in changed N	M	SD	Mann–Whit. U	p	d
SEL	59	4.84	1.13	14	4.83	1.34	420.0	.93	68	4.76	1.16	5	5.88	0.67	71.5	**.03**	**.58**
MET	58	4.34	1.02	14	4.01	1.00	322.0	.24	67	4.29	1.04	5	4.02	0.46	202.5	.44	.21
TIM	58	4.84	1.01	14	4.75	0.94	390.0	.83	67	4.82	0.99	5	4.80	1.06	163.5	.94	.02
ENE	58	3.96	1.04	14	3.88	1.08	397.0	.90	67	3.88	1.02	5	4.87	0.96	85.0	.07	.49
DED	58	4.46	0.98	14	4.14	1.21	352.5	.45	67	4.33	1.01	5	5.33	0.85	69.5	**.03**	**.59**
ABS	58	3.81	1.19	14	3.62	1.36	366.5	.58	67	3.77	1.20	5	3.80	1.57	167.5	1	.00
PER	57	0.61	0.37	14	0.54	0.35	349.5	.47	66	0.58	0.37	5	0.84	0.15	104.5	.17	.37
PRO	57	0.46	0.35	14	0.36	0.36	323.5	.27	66	0.44	0.35	5	0.53	0.34	133.5	.48	.19
DEM	57	0.64	0.37	14	0.54	0.35	329.5	.30	66	0.60	0.37	5	0.80	0.21	120.5	.31	.27
SEN	57	0.20	0.28	14	0.23	0.33	405.5	.92	66	0.19	0.29	5	0.40	0.29	91.5	.07	.45

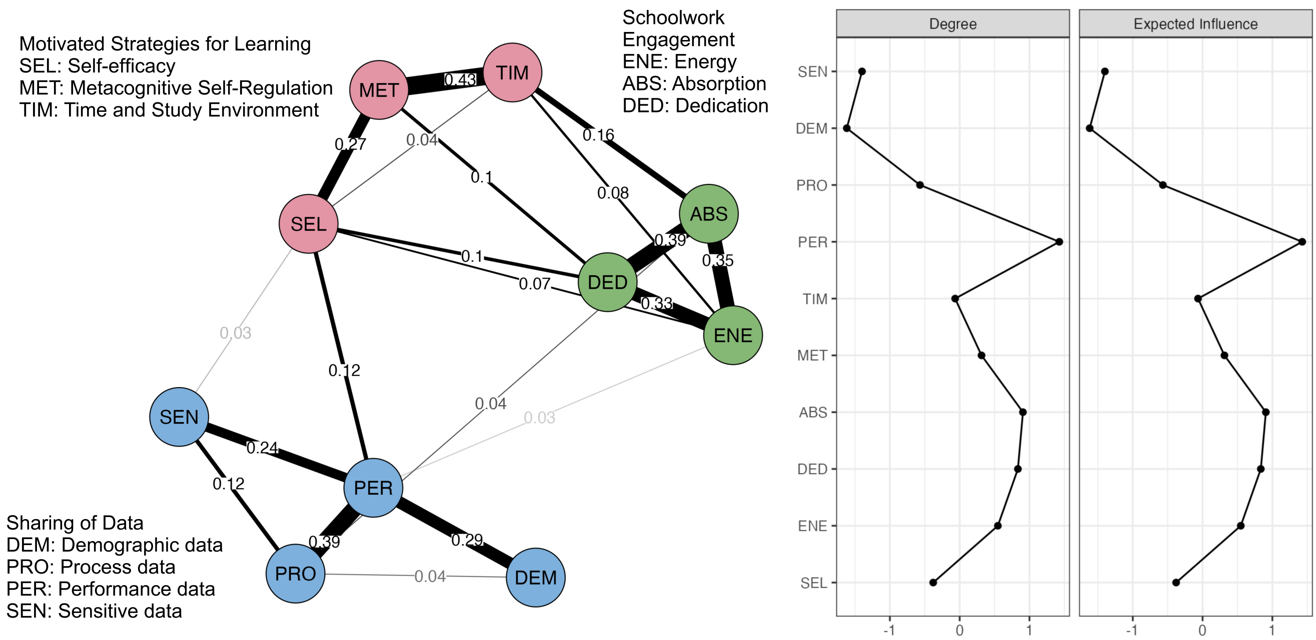


FIGURE 1. Partial correlation network of measures with centrality metrics.

In summary, none of the models were sufficient for predicting students' initial opt-in based on self-reported learning strategies, study engagement, and willingness to share data. Among the different predictors, only metacognitive self-regulation in the fourth model was statistically significant.

IV. DISCUSSION

A. DATA CATEGORIES AND WILLINGNESS TO SHARE

In general, the more fine-grained and personal the data are, the less likely students are willing to share them with learning analytics applications. Students were most willing to share demographic data ($M = .62$), followed by performance data ($M = .60$). Students were hesitant to share process data ($M = .44$) and sensitive data ($M = .21$). Especially, the results regarding sharing process data are noteworthy since different forms of process data are the most used data categories in learning analytics applications [46].

Our results agree with previous findings: applying the factor structure found here to results by Ifenthaler and Schumacher [23], we found little difference in results regarding performance data, process data, or sensitive data. The willingness to share demographic data here was moderately higher than results by [23] ($M = .50$), which contextual factors or cultural differences could explain.

B. SELF-EFFICACY AND PERFORMANCE DATA

The partial correlation network shows that students with higher self-efficacy were more willing to disclose performance data for learning analytics applications. In other words, students not expecting to perform well were less inclined to share their performance data.

While such data (e.g., grades, test results) are usually available to teachers and institutions, we specifically asked if the student would be willing to disclose these data for learning analytics systems. Perhaps the most prevalent use

TABLE 4. Logistic regression models to predict initial opt-in for LA-based feedback.

	No weighting						*Weighted Models (4:1)*					
	Model 1 (naive)			Model 2 (stepwise AIC)			Model 3 (naive)			Model 4 (stepwise AIC)		
	Est.	Std. Err.	p	Est.	Std. Err.	p	Est.	Std. Err.	p	Est.	Std. Err.	p
SEL	−.40	.51	.44				−.35	.34	.30	.14	.21	.13
MET	.71	.54	.19				.63	.34	.06	**.59**	**.29**	**.04***
TIM	−.11	.47	.82				−.10	.35	.77			
ENE	−.53	.69	.44				−.54	.47	.25	−.66	.41	.10
DED	.91	.73	.21				.86	.48	.07	.82	.43	.06
ABS	−.36	.73	.62				−.26	.47	.57			
PER	−.01	.49	.99				−.00	.35	.99			
PRO	.32	.46	.49				.43	.31	.16			
DEM	.36	.38	.34				.26	.26	.30	.39	.21	.07
SEN	−.19	.40	.64				−.25	.27	.35			
Constant	1.57	.35	<.01	1.39	.30	<.01	.18	.21	.41	.14	.21	.50
N	70			70			70			70		
Sensitivity	0.00			0.00			0.71			0.71		
Specificity	0.98			1.00			0.59			0.63		
Bal. Acc.	0.49			0.50			0.65			0.67		

$^{*}\, p < .05$

case for these systems to utilize performance data would be to differentiate content based on performance. Thus, it is logical that a high-performing student is willing to disclose data to get more personal experience. While the same could apply to students not expecting to perform well (i.e., students with low self-efficacy), literature on students' help-seeking strategies show that low self-efficacy correlates with help-seeking avoidance – a view that needing help is a sign of weakness [47]. In this sense, disclosing performance data might be seen as a threat to one's self-esteem.

C. WE COULD NOT PREDICT OPT-IN

Regarding our second research aim, none of the four logistic regression models using self-reported data on learning strategies, engagement, and willingness to share data as predictors sufficiently predicted students' opt-in decisions. Especially interesting is that self-reported unwillingness to disclose data for learning analytics systems did not translate into opt-out of a learning analytics intervention. While only 44% of students reported being willing to share data about their learning process, still 77% of the students agreed to participate in a feedback intervention that utilized data on their online learning behavior.

While earlier research has found that willingness to share data is related to the anticipated use of learning analytics systems [23], our results show that the same does not necessarily apply to the decision to opt in for a learning analytics intervention. In literature, this phenomenon is called the privacy paradox: users are, in theory, concerned about their privacy but still choose to disclose data because they assess the risk as being minor or the benefits outweighing the risks [27]. Students may be very strategic about when to disclose data and if it is beneficial for their learning. However, a more detailed analysis of the rationales behind students' opt-in decisions is beyond the scope of the current study.

While all the models failed to predict students' opt-in decisions, metacognitive self-regulation was a statistically significant predictor of one of the four models. This raises the question regarding the role of LA-based formative feedback as SRL support. For a student with excellent SRL and time management skills, feedback telling if you are on track with your goal may feel rather superficial. However, this kind of prompt might be optimal for a student with intermediate self-regulatory skills. Recognizing these nuances might be easy for a teacher when thinking about an individual student, but it is challenging to scale up in teaching augmentation tools.

D. REVISITING OPT-IN DECISIONS

Previous research has shown that students expect to be able to choose whether to opt in for or opt out of learning analytics [20], [26], [30]. Tsai and colleagues [26] have suggested that students should also be able to revisit their opt-in decisions during the semester. Being able to opt out later may even increase the probability of initial opt in, as students can actually reflect on whether the intervention is helpful for them, and then decide if they like to continue with it.

While we gave this opportunity to students in the current study, nine students (6%) ended up using the option – mainly to opt out later after initially opting in. The students using this opportunity had, on average, higher dedication (i.e., general meaningfulness of and enthusiasm for studies) and self-efficacy (i.e., expectation to perform well) compared to students who did not change their opt-in decision. The more dedicated students may see the choice of opting in or out as more important than other students and thus are more likely to use the opportunity to change their status.

E. LIMITATIONS

This study includes some limitations. First, the sample for the current study was limited in size and consisted of students of only one course, entailing limited generalizability of the results. Furthermore, the small sample size may also cause the logistic regression models to be unstable. Our findings should be validated with a larger sample in the future.

Another limitation is that we do not know to what extent students' opt-in choice was driven by the disclosure of data and to what extent by their general willingness to receive feedback (or some other factor). Since the logic of LA-based formative feedback is rather complex, some students might not have understood how their data is used in the intervention. One possibility in the future would be to allow students to choose between no feedback, general feedback, and personalized LA-based feedback, clearly indicating that the last one requires them to disclose data on their learning process. Furthermore, as students made their opt-in decision when starting the course and they could start the course any time during the semester, the time between the survey (collected at the beginning of the semester) and the opt-in decision was not constant between the participants.

Finally, it should be acknowledged that the current study used self-reports to measure learning strategies and study engagement. Self-report instruments have been found to largely measure students' intentions, which may differ from their actual behavior [48]. This should be considered when interpreting our results.

V. CONCLUSION

We investigated how learning strategies and study engagement relate to students' willingness to share data with learning analytics applications and whether these factors predict students' actual opt-in for LA-based formative feedback. We found that students with lower self-efficacy were more hesitant to share data about their performance. However, we could not sufficiently predict students' opt-in decisions based on their self-reported learning strategies, study engagement, or willingness to share data.

Our inability to predict students' opt-in decisions emphasizes the contextuality of opt-in behavior. Previous research has shown that students' willingness to share their data depends on the data to be shared [23] and the purpose for which the data are used [24]. Based on our findings, asking for opt-in for a specific intervention should be preferred over requesting consent for using particular data or data categories. Still, it is vital to acknowledge the differences between data categories. We found that students were more hesitant to share sensitive or process data than performance or demographic data. A balance between the data required and the usefulness of the intervention is needed: one might not want to share sensitive data for a marginal benefit [27].

Overall, our findings underline the importance of student agency in learning analytics, called for in previous research [20], [26], [32]. An opt-in/opt-out design with no predefined default and the ability to revisit choices as self-service puts students in control and generates valuable information. A low or decreasing opt-in rate is a sign that educators and learning analytics developers should look into: Why did students choose to opt out? Did they understand the intervention? Was there something that raised suspicion? For example, we found that students not expecting to perform well were more hesitant to share their performance data. Should this be due to the student being ashamed of their low performance, one could emphasize that the intervention aims to help and support the student, not to monitor performance or facilitate student competition. In this sense, opt-in rates and particularly changes in them are valuable tools that help learning analytics developers design and evaluate learning analytics interventions.

REFERENCES

[1] D. H. Schunk and J. A. Greene, *Historical, Contemporary, and Future Perspectives on Self-Regulated Learning and Performance*. Evanston, IL, USA: Routledge, 2017.

[2] I. Clark, "Formative assessment: Assessment is for self-regulated learning," *Educ. Psychol. Rev.*, vol. 24, no. 2, pp. 205–249, 2012.

[3] J. Hattie and H. Timperley, "The power of feedback," *Rev. Educ. Res.*, vol. 77, no. 1, pp. 81–112, 2007.

[4] L. Bol and J. K. Garner, "Challenges in supporting self-regulation in distance education environments," *J. Comput. Higher Educ.*, vol. 23, nos. 2–3, pp. 104–123, Dec. 2011.

[5] J. Rhode, "Interaction equivalency in self-paced online learning environments: An exploration of learner preferences," *Int. Rev. Res. Open Distrib. Learn.*, vol. 10, no. 1, pp. 1–23, Feb. 2009.

[6] P. Ihantola, I. Fronza, T. Mikkonen, M. Noponen, and A. Hellas, "Deadlines and MOOCs: How do students behave in MOOCs with and without deadlines," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Oct. 2020, pp. 1–9.

[7] P. An, K. Holstein, B. d'Anjou, B. Eggen, and S. Bakker, "The TA framework: Designing real-time teaching augmentation for K-12 classrooms," in *Proc. Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–17.

[8] J. M. Lodge, E. Panadero, J. Broadbent, and P. G. D. Barba, "Supporting self-regulated learning with learning analytics," in *Learning Analytics in the Classroom*. Evanston, IL, USA: Routledge, 2018, pp. 45–55.

[9] O. Viberg, M. Hatakka, O. Bälter, and A. Mavroudi, "The current landscape of learning analytics in higher education," *Comput. Hum. Behav.*, vol. 89, pp. 98–110, Dec. 2018.

[10] A. Pardo, J. Jovanovic, S. Dawson, D. Gašević, and N. Mirriahi, "Using learning analytics to scale the provision of personalised feedback," *Brit. J. Educ. Technol.*, vol. 50, no. 1, pp. 128–138, Jan. 2019.

[11] A. Pardo, K. Bartimote, S. Buckingham Shum, S. Dawson, J. Gao, D. Gašević, S. Leichtweis, D. Liu, R. Martínez-Maldonado, N. Mirriahi, A. C. M. Moskal, J. Schulte, G. Siemens, and L. Vigentini, "OnTask: Delivering data-informed, personalized learning support actions," *J. Learn. Anal.*, vol. 5, no. 3, pp. 235–249, Dec. 2018.

[12] L.-A. Lim, S. Gentili, A. Pardo, V. Kovanović, A. Whitelock-Wainwright, D. Gašević, and S. Dawson, "What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course," *Learn. Instruct.*, vol. 72, Apr. 2019, Art. no. 101202.

[13] H. Iraj, A. Fudge, H. Khan, M. Faulkner, A. Pardo, and V. Kovanović, "Narrowing the feedback gap: Examining Student engagement with personalized and actionable feedback messages," *J. Learn. Analytics*, vol. 8, no. 3, pp. 101–116, Nov. 2021.

[14] A. P. Cavalcanti, A. Barbosa, R. Carvalho, F. Freitas, Y.-S. Tsai, D. Gašević, and R. F. Mello, "Automatic feedback in online learning environments: A systematic literature review," *Comput. Educ., Artif. Intell.*, vol. 2, Jan. 2021, Art. no. 100027.

[15] A. Jonsson, "Facilitating productive use of feedback in higher education," *Act. Learn. Higher Educ.*, vol. 14, no. 1, pp. 63–76, Mar. 2013.

[16] M. Price, K. Handley, J. Millar, and B. O'Donovan, "Feedback: All that effort, but what is the effect?" *Assessment Eval. Higher Educ.*, vol. 35, no. 3, pp. 277–289, May 2010.

[17] L.-A. Lim, D. Gasevic, W. Matcha, N. A. Uzir, and S. Dawson, "Impact of learning analytics feedback on self-regulated learning: Triangulating behavioural logs with students' recall," in *Proc. 11th Int. Learn. Anal. Knowl. Conf.*, Apr. 2021, pp. 364–374.

[18] E. L. Deci, R. Koestner, and R. M. Ryan, "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation," *Psychol. Bull.*, vol. 125, no. 6, p. 627, 1999.

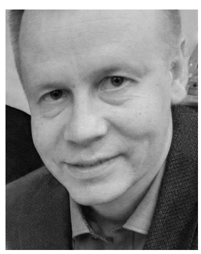
[19] S. Slade and P. Prinsloo, "Learning analytics: Ethical issues and dilemmas," *Amer. Behav. Scientist*, vol. 57, no. 10, pp. 1510–1529, 2013.

[20] L. D. Roberts, J. A. Howell, K. Seaman, and D. C. Gibson, "Student attitudes toward learning analytics in higher education: '*The fitbit version of the learning world*'," *Frontiers Psychol.*, vol. 7, p. 1959, Dec. 2016.
[21] K. M. Jones, A. Asher, A. Goben, M. R. Perry, D. Salo, K. A. Briney, and M. B. Robertshaw, "'We're being tracked at all times': Student perspectives of their privacy in relation to learning analytics in higher education," *J. Assoc. Inf. Sci. Technol.*, vol. 71, no. 9, pp. 1044–1059, 2020.
[22] M. Teräs, A. Silvola, H. Teräs, S. Hartikainen, S. Aksovaara, R. Hietaniemi, and H. Muukkonen, "Learning analytics for students: Synthesis of two user needs studies in finnish higher education," in *EdMedia+ Innovate Learning*. Chesapeake, VA, USA: Association for the Advancement of Computing in Education (AACE), 2020, pp. 455–463.
[23] D. Ifenthaler and C. Schumacher, "Student perceptions of privacy principles for learning analytics," *Educ. Technol. Res. Develop.*, vol. 64, no. 5, pp. 923–938, Oct. 2016.
[24] K. E. Arnold and N. Sclater, "Student perceptions of their privacy in leaning analytics applications," in *Proc. 7th Int. Learn. Anal. Knowl. Conf.*, Mar. 2017, pp. 66–69.
[25] L. Bennett and S. Folley, "Four design principles for learner dashboards that support student agency and empowerment," *J. Appl. Res. Higher Educ.*, vol. 12, no. 1, pp. 15–26, May 2019.
[26] Y.-S. Tsai, A. Whitelock-Wainwright, and D. Gašević, "The privacy paradox and its implications for learning analytics," in *Proc. 10th Int. Conf. Learn. Anal. Knowl.*, Mar. 2020, pp. 230–239.
[27] S. Barth and M. D. T. de Jong, "The privacy paradox—Investigating discrepancies between expressed privacy concerns and actual online behavior—A systematic literature review," *Telematics Informat.*, vol. 34, no. 7, pp. 1038–1058, Nov. 2017.
[28] D. West, A. Luzeckyj, D. Toohey, J. Vanderlelie, and B. Searle, "Do academics and university administrators really know better? The ethics of positioning student perspectives in learning analytics," *Australas. J. Educ. Technol.*, vol. 36, no. 2, pp. 60–70, 2020.
[29] A. Rubel and K. M. L. Jones, "Student privacy in learning analytics: An information ethics perspective," *Inf. Soc.*, vol. 32, no. 2, pp. 143–159, Mar. 2016.
[30] S. Slade and P. Prinsloo, "Student perspectives on the use of their data: Between intrusion, surveillance and care," in *Proc. EDEN Conf.*, no. 2, 2014, pp. 291–300.
[31] P. Prinsloo and S. Slade, "Student privacy self-management: Implications for learning analytics," in *Proc. 5th Int. Conf. Learn. Anal. Knowl.*, Mar. 2015, pp. 83–92.
[32] S. Slade, P. Prinsloo, and M. Khalil, "Learning analytics at the intersections of student trust, disclosure and benefit," in *Proc. 9th Int. Conf. Learn. Anal. Knowl.*, 2019, pp. 235–244.
[33] L. D. Roberts, J. A. Howell, and K. Seaman, "Give me a customizable dashboard: Personalized learning analytics dashboards in higher education," *Technol., Knowl. Learn.*, vol. 22, no. 3, pp. 317–333, Oct. 2017.
[34] C. Schumacher and D. Ifenthaler, "Features students really expect from learning analytics," *Comput. Hum. Behav.*, vol. 78, pp. 397–407, Jan. 2018.
[35] A. Silvola, P. Näykki, A. Kaveri, and H. Muukkonen, "Expectations for supporting student engagement with learning analytics: An academic path perspective," *Comput. Educ.*, vol. 168, Jul. 2021, Art. no. 104192.
[36] I. Jivet, J. Wong, M. Scheffel, M. V. Torre, M. Specht, and H. Drachsler, "Quantum of choice: How learners' feedback monitoring decisions, goals and self-regulated learning skills are related," in *Proc. 11th Int. Learn. Anal. Knowl. Conf.*, Apr. 2021, pp. 416–427.
[37] P. R. Pintrich, D. A. F. Smith, T. García, and W. J. McKeachie, "A manual for the use of the motivated strategies questionnaire (MSLQ)," Nat. Center Res. Improve Postsecondary Teach. Learn., Univ. Michigan, Ann Arbor, MI, USA, 1991.
[38] K. Salmela-Aro and K. Upadaya, "The schoolwork engagement inventory," *Eur. J. Psychol. Assessment*, vol. 28, no. 1, pp. 60–67, Sep. 2012.
[39] J. Love, R. Selker, M. Marsman, T. Jamil, D. Dropmann, J. Verhagen, A. Ly, Q. F. Gronau, M. Smíra, S. Epskamp, D. Matzke, A. Wild, P. Knight, J. N. Rouder, R. D. Morey, and E.-J. Wagenmakers, "JASP: Graphical statistical software for common statistical designs," *J. Stat. Softw.*, vol. 88, no. 2, pp. 1–17, 2019.
[40] W. Revelle and M. W. Revelle, "Package 'psych,'" *Comprehensive R Arch. Netw.*, vol. 337, p. 338, Jan. 2015.
[41] S. Epskamp and E. I. Fried, "A tutorial on regularized partial correlation networks," *Psychol. Methods*, vol. 23, no. 4, p. 617, 2018.
[42] J. Malmberg, M. Saqr, H. Järvenoja, and S. Järvelä, "How the monitoring events of individual students are associated with phases of regulation: A network analysis approach," *J. Learn. Anal.*, vol. 9, no. 1, pp. 77–92, Mar. 2022.
[43] M. Saqr, O. Viberg, and W. Peteers, "Using psychological networks to reveal the interplay between foreign language students' self-regulated learning tactics," in *Proc. STELLA2020*, vol. 2828, 2021, pp. 12–23.
[44] L. F. Bringmann, T. Elmer, S. Epskamp, R. W. Krause, D. Schoch, M. Wichers, J. T. Wigman, and E. Snippe, "What do centrality measures measure in psychological networks? *J. Abnormal Psychol.*, vol. 128, no. 8, p. 892, 2019.
[45] D. J. Robinaugh, A. J. Millner, and R. J. McNally, "Identifying highly influential nodes in the complicated grief network," *J. Abnormal Psychol.*, vol. 125, no. 6, p. 747, 2016.
[46] N. Nistor and Á. Hernández-Garcíac, "What types of data are used in learning analytics? An overview of six cases," *Comput. Hum. Behav.*, vol. 89, pp. 335–338, Dec. 2018.
[47] M. C. White and H. Bembenutty, "Not all avoidance help seekers are created equal: Individual differences in adaptive and executive help seeking," *Sage Open*, vol. 3, no. 2, 2013, Art. no. 2158244013484916.
[48] D. Gasevic, J. Jovanovic, A. Pardo, and S. Dawson, "Detecting learning strategies with analytics: Links with self-reported measures and academic performance," *J. Learn. Anal.*, vol. 4, no. 2, pp. 113–128, 2017.

JOONAS MERIKKO received the M.Sc. degree in mathematics education from the University of Helsinki, Finland, in 2013, where he is currently pursuing the Ph.D. degree with the Faculty of Educational Sciences, specializing in self-regulated learning in the age of AI. His research interests include student's and teacher's views on learning technologies and these technologies potential effects on how teaching and learning are organized.

KWOK NG received the Ph.D. degree. He is currently a Senior Researcher and a Postdoctoral Researcher with the Faculty of Education, University of Turku, Finland, the Department of Physical Education and Sport Sciences, Faculty of Education and Health Sciences, University of Limerick, Ireland, and the School of Educational Sciences and Psychology, University of Eastern Finland, Finland. He was the Principal Investigator of the UEFOT Project with the University of Eastern Finland.

MOHAMMED SAQR received the Ph.D. degree in learning analytics from Stockholm University. He currently works as a Senior Researcher in artificial intelligence, big data in education, network science, and scientometrics at the University of Eastern Finland. He is particularly interested in research methods, including network analysis, temporal networks, machine learning, process, and sequence mining, and temporal processes in general. He is also an active member of several scientific organizations and acts as an academic editor in leading academic publications.



PETRI IHANTOLA received the Ph.D. degree from Aalto University, in 2011. He currently works as an Associate Professor in big data learning analytics and the Director of the MOOC-Center, University of Helsinki, Finland. His research interests include educational data mining and building educational software with a particular focus on smart content, automated assessment, and learning analytics in computing education.