Int. J. Human–Computer Studies 211 (2026) 103782



Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs





ISense: An assessment instrument that predicts self-regulated learning using Mobile sensing

Tongyu Zhao[a,b], Jiaying Gao[c], Yu Feng[b,a], Yatong Zu[d], Adriano Tavares[e], Tiago Gomes[b], Sandro Pinto[b], Hao Xu[a,f,*]

[a] College of Computer Science and Technology, Jilin University, 130012 Changchun, China
[b] ALGORITMI Center, School of Engineering, University of Minho, 4804-533 Guimarães, Portugal
[c] Changchun University of Science and Technology, 130022 Changchun, China
[d] College of Physical Education, Jilin University, 130012 Changchun, China
[e] Department of Industrial Electronics, University of Minho, 4704-553 Braga, Portugal
[f] Jilin University Library, Jilin University, 130012 Changchun, China

HIGHLIGHTS

- We develop and implement the iSense system for SRL assessment.
- Collecting mobile sensing and self-reported data from 211 students over a year.
- Establishing an SRL assessment model using mobile sensing data and deep learning.
- We observe that SRL highly correlates with app usage patterns and sensing data.
- Our study can achieve longitudinal monitoring and early intervention of SRL.

ARTICLE INFO

Keywords:
Lifelong learning
Mobile learning
Improving classroom teaching
Evaluation methodologies
Human-computer interface

ABSTRACT

Self-regulated learning significantly influences students' academic behavior and performance. Traditional SRL assessment heavily relies on subjective and self-evaluation methods, which are susceptible to personal biases and cognitive limitations. In this study, we leverage mobile devices with GPS and sensors to collect self-reported and passive mobile sensing data from 211 college students over a year. Our aim is to conduct a passive assessment of self-regulated learning. To achieve this, we apply four deep learning models to analyze behavioral features associated with self-regulated learning using students' life record data. Our findings demonstrate a precise assessment of student self-regulated learning, encompassing the following subscales. Environment structuring (MAE = 2.88, r = 0.54) represents participation in the organization and construction of the learning environment. Time management (MAE = 2.82, r = 0.57) represents planning study time and balancing study with other activities; Help seeking (MAE = 4.86, r = 0.53) represents the tendency and frequency of students to seek help during the learning process. Our study helps inspire new forms of education to assess self-regulated learning and paves the way for individualized interventions in future studies.

1. Introduction

Self-regulated learning (SRL) plays a pivotal role in education, enabling learners to monitor and optimize their learning processes and outcomes (Pintrich, 2004). The definition of SRL has evolved over time, alongside advances in assessment methodologies (Pintrich et al., 2000). According to social cognitive theory, SRL encompasses student behavioral paradigms that emphasize internal maintenance activities and consciously controlled behaviors (Bandura, 1991; Schunk, 2013). However, traditional assessment methods, such as self-report surveys, think-aloud protocols, and interviews, are prone to biases and fail to capture real-time behavior changes (Azevedo & Gašević, 2019; van Halema et al., 2020; Saint et al., 2020; Du et al., 2023).

* Corresponding author at: College of Computer Science and Technology, Jilin University, 130012 Changchun, China.
Email addresses: id9269@alunos.uminho.pt (T. Zhao), jiayingg20@mails.jlu.edu.cn (J. Gao), id8946@alunos.uminho.pt (Y. Feng), zuyatong@jlu.edu.cn (Y. Zu), atavares@dei.umihno.pt (A. Tavares), mr.gomes@dei.uminho.pt (T. Gomes), sandro.pinto@algoritmi.uminho.pt (S. Pinto), xuhao@jlu.edu.cn (H. Xu).

https://doi.org/10.1016/j.ijhcs.2026.103782
Received 12 July 2024; Received in revised form 26 August 2025; Accepted 28 February 2026
Available online 6 March 2026
1071-5819/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Recognizing SRL as a dynamic interplay of behaviors necessitates real-time, fine-grained analysis for accurate assessment (Saint et al., 2022; Li et al., 2020b). Consequently, new methods are emerging to assess SRL evolution, increasingly explored in diverse learning environments (Viberg et al., 2020; Araka et al., 2020). Recent research has explored data tracking methods to objectively assess SRL, offering insight into student learning behaviors (Araka et al., 2020; Li et al., 2020a). Although studies have predicted, described, and guided SRL behaviors in online environments, challenges in reliability and replicability persist (Hachey et al., 2015; Zimmerman & Schunk, 2011; Yang et al., 2015; Araka et al., 2019; Wong et al., 2019).

Embedded sensors in mobile devices provide insights into students' location and behavior (Harari et al., 2016). Mobile devices have become ubiquitous in students' lives, influencing both classroom and off-task behaviors (Chen & Yan, 2016; May & Elder, 2018; Hawi & Samaha, 2016; Samaha & Hawi, 2016; Nikou & Economides, 2017). However, they also present opportunities to mitigate learning-related risks, which have not been fully explored in educational research. Addressing the need for innovative psychological measurement methods, our work investigates a quantitative and continuous approach to assess SRL levels within students' daily lives. Our study integrates passive mobile sensing with self-report responses to explore correlations between daily behaviors and SRL. Specifically, we aim to answer the following research questions (RQ): **RQ1:** What are the behavioral patterns related to SRL in students' daily lives? **RQ2:** Is it feasible to predict students' SRL using passive behavioral sensing assessments obtained from smartphones?

To address the inquiries posed above, we collect a dataset of online self-regulated learning (OSLQ) measurements from college students at a university in China, along with multimodal mobile sensing data. Barnard et al. (2009) have been proven evidence toward the reliability and validity of the OSLQ to assess SRL skills in blended and online course formats. SRL is a process that fluctuates with changes in the learning environment. The advantage of the OSLQ is that it extends the study of SRL to various learning contexts.

To gain a deeper understanding of SRL skills and changes in learning behavior in different learning formats, we need to collect high-quality longitudinal data across time spans to capture the diversity and changes in learning behavior patterns. We develop a mobile application (iSense) to gather real-time data from students. Subsequently, we utilize one-year mobile sensing data from a cohort of 211 college students to identify behavioral patterns. The key contributions of this paper are summarized as follows:

- We present the first study to leverage mobile sensing data to explore the relationship between daily behavioral patterns and SRL, revealing a strong correlation between these patterns and SRL based on the analyzed sensor data.
- We discover a set of behavioral patterns associated with SRL and provide fundamental conditions to predict SRL using mobile sensing.
- The iSense system, developed and deployed as an Android application, is a practical tool that integrates passive multimodal mobile sensing with self-reported data. The data were then processed using our proposed models. The system's efficacy in automatically assessing SRL was trained and validated, demonstrating its effectiveness in real-world applications.

2. Related work

2.1. Conceptualization of self-regulated learning

Self-regulated learning (SRL) is recognized as a critical set of skills that students use to ensure effective learning in various situations by consistently adapting their learning locations, goals, environments, and learning strategies (Zimmerman & Schunk, 1994; Paris & Paris, 2003). The concept of SRL posits that students have clear learning goals and can autonomously decide how to achieve these goals given a set of external and internal conditions (Zimmerman, 1986; Pintrich, 2004; Schmitz & Wiese, 2006). In this case, the student might consult the course materials in the library and watch lecture videos in their dormitory at night. These learning episodes represent typical learning scenarios. Students learn in fragmented time periods, in different learning environments, and using various learning forms and technologies. Therefore, the final task may represent a single SRL cycle or a series of ongoing cycles and subcycles (Saint et al., 2022). In this sense, SRL is viewed as an ongoing process that unfolds and develops over time (Azevedo et al., 2010; Winne & Perry, 2000). SRL behaviors interact with each other over time to create internal dynamics that constitute the essence of the learning process (Bringmann et al., 2016; Engelmann & Bannert, 2021).

2.2. Assessment of self-regulated learning

There are various methods for assessing SRL. One common approach is through retrospective self-reports (Roth et al., 2016). Roth et al. (2016) conducted a comprehensive examination of the self-report tools frequently used in SRL studies, including questionnaires, interviews, verbal protocols, and learning diaries. Generally, students are required to record or assess their SRL skills. For example, they may keep structured learning diaries or record their time investment in each task. Retrospective self-reports may be suitable for capturing stable SRL across time and context. However, SRL is a learning event that fluctuates and changes as it evolves over time and context. Because participants tend to summarize their experiences when retrieving past events, data from retrospective assessments may be insufficient to study SRL as a series of learning events (McCardle & Hadwin, 2015; Torrano Montalvo et al., 2004; Schmitz & Wiese, 2006; Shiffman et al., 2008; Veenman, 2011).

Online tracking technology allows the learning process to be captured in real time by manipulating the learning conditions and using methods other than self-reports. In some studies, questionnaire-based assessments focus on specific learning tasks or learning activities (Ifenthaler, 2012). Other research designs employ techniques, include monitoring students' emotional states in classrooms through camera-based behavior engagement and computer activities to provide feedback for improving SRL (Delen & Liew, 2016; Davis et al., 2016; Aslan et al., 2019; Järvelä & Bannert, 2019); using wearable devices to analyze students' facial expressions and postures to enhance SRL (Shih et al., 2010; Nussbaumer et al., 2014); employing eye-tracking technologies to assess gaze patterns and behaviors in predefined Areas of Interest (AOIs) to assess metacognitive monitoring and SRL (Taub & Azevedo, 2019; Holmqvist et al., 2024); recording students' attention during classes with cameras and VR to improve their SRL (Ahuja et al., 2021); promoting SRL through dashboard visualizations of learner activities (Davis et al., 2016; Malmberg et al., 2021); and designing multifunctional mobile applications that enhance SRL via peer feedback (Cárdenas-Robledo & Peña-Ayala, 2019; Dai & Wu, 2023). However, a common limitation of online tracking technologies is that they suffer from artificial learning environments (Loeffler et al., 2013). Moreover, the latter methods have been concentrated on lab-based simulations of real-world behaviors, which can introduce artificial constraints that could disrupt the learning process or overlook the underlying motivations and cognitive processes (Loeffler et al., 2019).

2.3. Mobile technology supports self-regulated learning

In fact, mobile devices may be the only items with which students coexist throughout all their scattered learning moments and environments during the day. Therefore, we recommend using personal mobile devices to record learning events. This efficient method accurately measures how students learn independently of the materials they use (e.g., textbooks, paper notes, tablets, computers), independently of location (e.g., waiting times, commuting, classrooms, dorm and independently of the duration of the learning sessions (e.g., from 1 to n minutes).

Recent advances in mobile sensing technology have transformed behavioral assessments, enabling unobtrusive and continuous tracking of

Table 1
Research using mobile sensing to assess various aspects of students' psychological variables (2014–2024).

Author(s) and Publication year	Psychological variables	Duration (N = participants)	Features
Nepal et al. (2024)	Mental health	3 years (N = 186)	**Sensing data:** Physics activity, APP Usage, Location, Screen time, Audio play, step counts. **Self-reported data:** COVID, Patient Health, PHQ4, Self Esteem, Social Level, Stress.
Feng et al. (2023)	Perceived relatedness	3 months (N = 58)	**Sensing data:** Phone activity. **Self-reported data:** SRQ-A.
Kammoun et al. (2023)	Eating	13 weeks (N = 678)	**Sensing data:** App tools, Physics activity, Location, Screen time. **Self-reported data:** Eating, etc.
Wang et al. (2022)	Mental health	1 year (N = 180)	**Sensing data:** Location, physical activity, class attendance, sleep onset. **Self-reported data:** CIRP Freshmen Survey (TFS).
Xu et al. (2019)	Depression	106 days (N = 455)	**Sensing data:** Campus Map, Bluetooth, Sleep, Location. **Self-reported data:** BDI-II.
Wang et al. (2018b)	Personality	2 weeks (N = 646)	**Sensing data:** physical activity, ambient sound, phone usage location. **Self-reported data:** Big Five.
Wang et al. (2018a)	Depression	9 weeks (N = 83)	**Sensing data:** physical activity, location, sleep, conversations, heart rate, Screen time. **Self-reported data:** PHQ-8, PHQ-4.
Boukhechba et al. (2018)	Mental health	2 weeks (N = 72)	**Sensing data:** GPS, accelerometer, calls and SMS. **Self-reported data:** Social Interaction Anxiety Scale, Positive and Negative Affect Schedule.
Zhang et al. (2018)	Mood	1 month (N = 42)	**Sensing data:** APP Usage, ambient light, sounds, step count, accelerometer, indoor locations.
Farhan et al. (2016)	Depression	5 months (N = 79)	**Sensing data:** GPS. **Self-reported data:** PHQ-9.
Huang et al. (2016)	Social anxiety	10 days (N = 18)	**Sensing data:** GPS. **Self-reported data:** Social Interaction Anxiety.
Saeb et al. (2015)	Mental health	30 days (N = 66)	**Sensing data:** sleep, phone usage, number of email. **Self-reported data:** Pittsburg Sleep Quality Index, perceived stress scale, mental health composite.
Wang et al. (2014)	Mental health	10 weeks (N = 60)	**Sensing data:** Sleep, conversations, co-locations. **Self-reported data:** PHQ, PHQ-9.

behaviors through embedded mobile sensors such as accelerometers and microphones (Harari et al., 2017). Human-computer interaction researchers have explored how to use passively collected sensor data to infer both individual and communal behaviors and psychological states in various contexts, such as stress, anxiety, and depression, showcasing its potential for tracking and modeling student behavior (Saeb et al., 2015; Wang et al., 2018a). The feasibility of mobile sensing methods to monitor health-related behaviors in student daily lives has been demonstrated by study (He & Agu, 2017), which found significant correlations between mobile sensing data and mental health and academic performance. Furthermore, collaborations between engineers and psychologists have shown statistically significant relationships between mobile sensing data and levels of social anxiety among college students, with experiments using accelerometer and GPS data highlighting significant behavioral differences in various environments (Gong et al., 2019). Projects like StudentLife (Wang et al., 2014) and EDUM (Zhou et al., 2016) utilize mobile sensing for in-depth analysis of students' mental health based on indicators such as activity, location, and WiFi traffic patterns. Our previous research has also successfully identified various behavioral patterns using mobile sensing data, and investigated correlations between perceived relatedness and these behavioral patterns (Feng et al., 2023). Table 1 provides an overview of research using mobile sensing to assess various psychological variables in students' daily lives.

Although mobile sensing is increasingly used to monitor changes in student mental health, its application in the SRL domain remains underexplored. This study aims to bridge this gap by extending mobile sensing methods to assess SRL, thereby contributing a novel perspective to the literature. In this paper, personal mobile devices are instantiated as tools for recording and tracking learning events, serving as a new measure for assessing SRL.

3. ISense study: Design and apparatus

We collaborate with a representative team from the fields of psychology, computer science, user experience design, and ethics at a university in China to design an exploratory study and develop a mobile sensing application. This app is intended to collect both passive mobile sensing data and self-reported data from participants to gain insights into students' SRL skills and behavioral patterns. Below, we will discuss the study design process in detail.

3.1. ISense system framework design

The iSense predictive system consists of the iSense app, which runs on Android phones, and the iSense data analysis service, which operates in the cloud. The iSense app collects participants' passive sensing data and periodically administers self-reported SRL surveys, sending the survey responses to the iSense analysis system. The iSense data analysis service processes these data, building training and test datasets for further analysis and prediction. The passive sensing data and survey data will be discussed in the next section. Fig. 1 summarizes the system components and workflow.

3.1.1. ISense app and sensing system

The iSense app is designed to continuously collect and record participants' passive sensing data, which includes various types of data such as physical activity, app usage and location data. The app also periodically administers self-reported SRL surveys to gather additional insights.

When the iSense app is first launched, participants will see a consent screen that includes an agreement and privacy statement, with links to these documents within the app. After completing the consent process, the app generates a unique ID for each participant, which is displayed on the app's main interface. The app also includes a self-check feature, permission management, real-time behavior recognition, switching of current behavior states (e.g., sitting, studying, etc.), and notification listening status (participants can opt to authorize or deny), as shown in Fig. 2(a).

To ensure data security and privacy, the app stores the collected data locally on the phone until the device is charged and connected to Wi-Fi. Once these conditions are met, the app uploads the data securely to the cloud-based iSense data analysis service using SSL encryption. The uploaded data is stored in the cloud for further extraction and analysis.

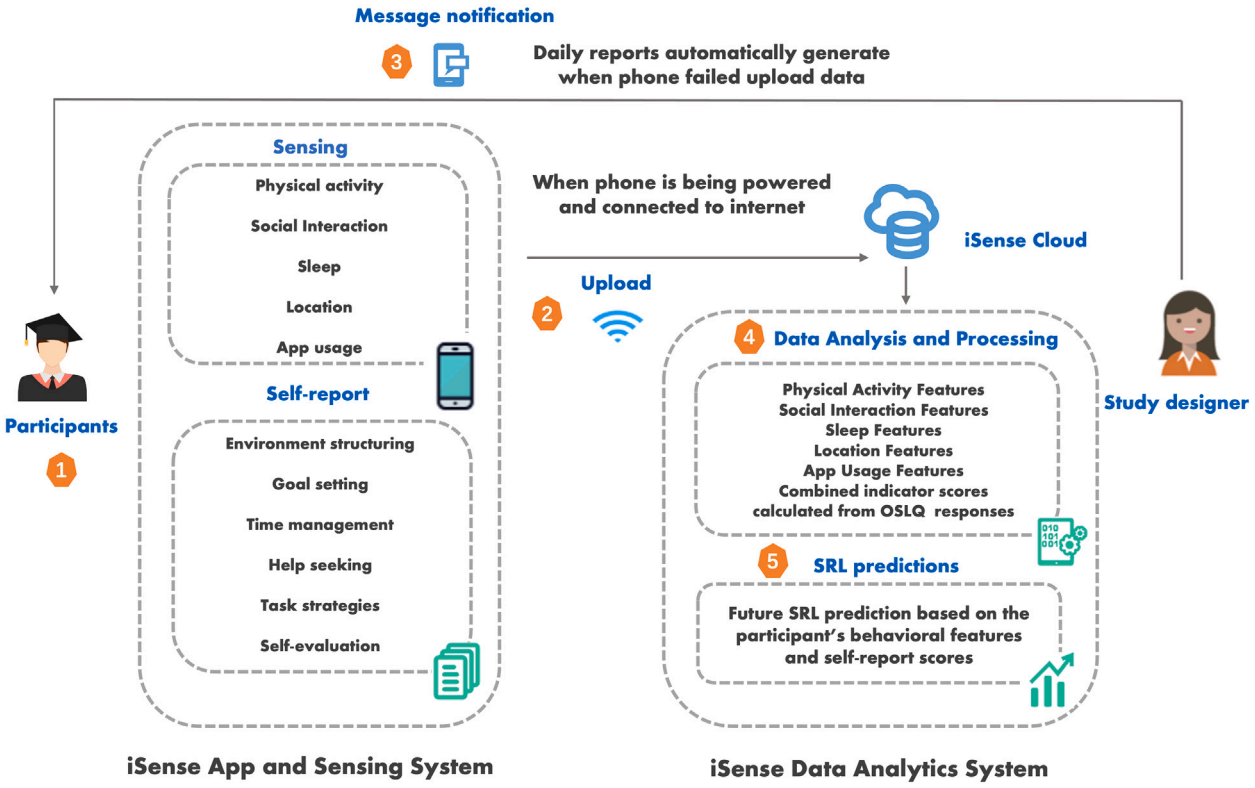
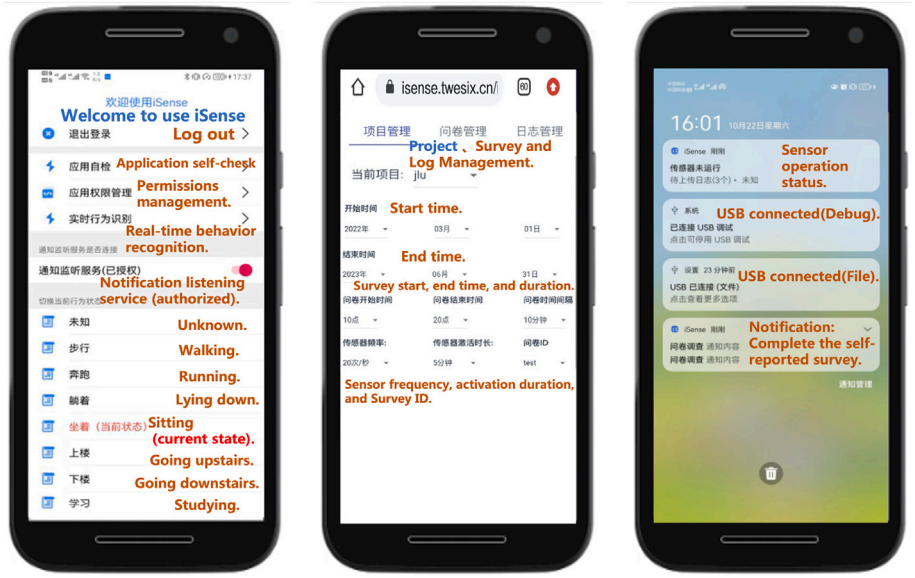


Fig. 1. iSense system and study design.



(a) (b) (c)

Fig. 2. Screenshots of iSense system: (a) shows the user interface of the iSense app; (b) shows the iSense system's control panel, primarily used by researchers for data detection and analysis; (c) illustrates the notification sent to participants when the app encounters technical issues preventing data upload, guiding them to get back on track.

3.1.2. iSense data analytics system

The iSense data analysis system receives and processes data from the app, helping researchers to monitor study compliance and track changes in participants' behavior.

Smartphone data processing. The analysis system receives passive sensing and SRL survey data from the iSense app and stores it in a secure cloud database, as shown in Fig. 2(c). The system automatically extracts features from the raw data for further analysis and model training. These features form the basis for monitoring study compliance and gaining insights into participants' behavior.

Daily study adherence reports. To monitor participant compliance and detect any technical issues preventing the application from

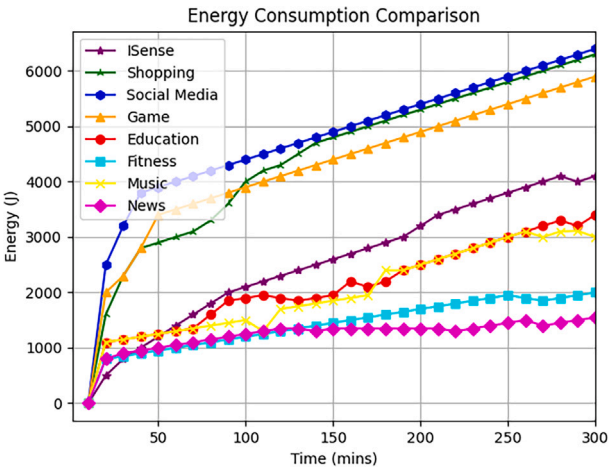


Fig. 3. Energy consumption of different APPs.

uploading passive sensing data, the system generates daily reports highlighting participants with unuploaded data. Once participants with technical issues are identified, researchers will send notifications to help them resolve the problem and get back on track, as shown in Fig. 2(b). We consider this feature to be a technical extension rather than an extension related to SRL prediction. Examples of compliance issues due to technical problems include device malfunctions, insufficient data coverage, inability to charge the device, or the phone being lost or stolen. Our primary focus is on the SRL behavior changes of students throughout the year, and this feature ensures that we can collect a large amount of data over the year.

Model training and validation. The system uses the extracted features to train detection models for data analysis and prediction generation. The performance of the models is validated to ensure their accuracy and reliability. The specific process of model training and validation will be explained in detail in a later section.

3.2. Energy consumption of iSense

To evaluate energy consumption, we conduct a comparative analysis between the iSense app and several widely used apps: news (Today's Hot), music (NetEase Music), social media (Weibo), education (Chinese MOOC), gaming (Honor of Kings), shopping (Taobao), and fitness (Keep). Each app was operated continuously for five hours, and their energy usage was monitored through battery logs. As depicted in Fig. 3, the results indicate that the energy consumption of iSense is modest, closely resembling that of educational (Chinese MOOC) and music (NetEase Music) apps, and only half as much as the gaming app (Honor of Kings). Social media (Weibo) and shopping apps (Taobao), which often require frequent sensor data collection and network communication, consumed more energy than iSense. These findings demonstrate that sampling sensor data every 5 minutes with iSense has a minimal impact on mobile usage, with energy consumption well within acceptable limits.

4. Method

4.1. Ethics clearance

This study received approval from the university's Institutional Review Board (IRB) on March 4, 2022. The application does not record privacy-sensitive information such as the content of text messages and voice calls. All student phone numbers and IDs are anonymized. For app usage, only categories (e.g., music, entertainment, social media) are recorded, without logging the actual names. For sensor data such as sound and light, aggregate metrics (mean, variance) are calculated and uploaded, without retaining the raw data. All participants have a clear method to withdraw their consent.

4.2. Participants

The data collected, analyzed, and discussed in this study comes from an ongoing long-term mobile sensing research project that primarily utilizes smartphone sensors and self-reported surveys. Jilin University, a prestigious academic institution situated in a small university town in northeast China, was chosen as the subject of this study. Unlike universities located in large cities, Jilin University is situated on a self-contained campus where all students are required to reside, which makes mobile sensing within the campus meaningful as it can capture the activities of all students. Almost all campus buildings have primary functions, such as dormitories, classrooms, libraries, gyms, cafeterias, and social gathering places. The number of multifunctional high-rise buildings, thus precise indoor positioning may not be necessary. Broader GPS signals can indicate the main area type where students are located. The university has two semesters: the spring semester typically runs from February to June, and the fall semester from September to December. There are two academic holidays (winter and summer breaks) during which undergraduates usually leave campus to go home, vacation, or intern in their fields. The semester progresses quickly, with students facing the pressure of midterm exams soon after the semester begins.

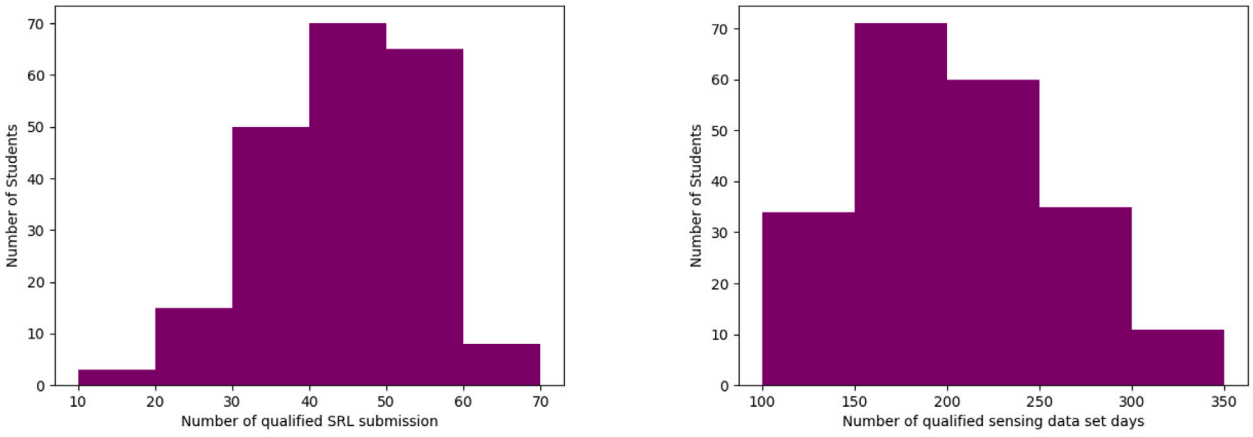
The participants were from this university, with a total of 289 students volunteered to participate in this one-year study (from March 2022 to March 2023). Participants were informed of the study's purpose in advance, and all were provided with a clear way to withdraw from the study. Specifically, students were recruited and provided consent one week prior to the start of the spring semester, using an initial sample size of 281 on March 6, 2022. Data collection and analysis began at the beginning of the study. All participants installed a continuous mobile sensing application (iSense APP) on their Android smartphones. The application was installed with the students' consent and collected data continuously for one year. To improve compliance and data quality, we administered weekly self-reported surveys to participants via a mobile sensing app and offered a weekly incentive of RMB 50 for completing the survey. Ultimately, only 211 students adhered to the study requirements and completed the weekly SRL assessments. Some participants were ineligible or withdrew from the study due to extensive gaps in their mobile sensing data. The demographic data of the 211 students used in our analysis include: most (55.45%, N = 117) were male. Regarding their departments, 23. 70% (N = 50) were in Social Sciences, 14. 22% (N = 30) in Engineering, 26. 54% (N = 56) in Physical Education, 21. 32% (N = 45) in Information Science, and 14. 22% (N = 30) in Medicine. Our participants covered a wide range of majors and grades throughout the university.

4.3. Data collection

Our analysis uses aggregated daily data, such as hours spent studying, with special attention to handling missing data within each day. Missing data can significantly skew derived values and inaccurately represent participant behavior. Therefore, we determine the number of sensed hours in a day using the ratio of three forms of continuously sampled data: activity, location, and screen lock time. Days with less than 19 hours of sensor data are excluded from the analysis to maintain data integrity.

Fig. 4(a) illustrates the distribution of participants and their corresponding number of responses in the self-reported dataset, which was collected weekly. In particular, approximately 31% (n = 65) of the students submitted more than 50 qualified reports, and 8 participants continued to submit reports for two months beyond the end of the experiment. The majority of participants submitted over 40 qualifying reports.

Fig. 4(b) illustrates the distribution of the participants in the passive sensing dataset and their corresponding number of days. Approximately 84% (n = 177) of the participants logged more than 150 days of mobile data. The minimum number of days recorded by any participant was 113, with a maximum of 287 days. On average, each participant



(a) (b)

Fig. 4. The distribution of data. (a) Distribution of the number of qualified submissions; (b) Distribution of participants' sensing data set days.

contributed 225.93 days (approximately 32 weeks) of sensing data and 39.4 self-reported responses.

In total, the analysis encompasses 287 days of sensing data and 3758 self-reported responses from 211 participants. This comprehensive dataset forms the basis for the training and testing sets used in our model, enabling robust and detailed analysis of SRL behaviors.

4.3.1. Survey questionnaires

To rigorously assess SRL among participants, we employ the Online Self-Regulated Learning Questionnaire (OSLQ), a tool renowned for its robust validity (Barnard et al., 2009). SRL is indeed a process that fluctuates and changes with the learning environment. The advantage of using the OSLQ is that it extends SRL research to other learning scenarios. Researchers can use the OLSQ to assess SRL skills and changes in the use of these skills across multiple forms of learning, thus providing critical information about the student and the learning environment (Barnard et al., 2009; Bruso & Stefaniak, 2016).

The OSLQ is comprised of six subscales that evaluate distinct aspects of SRL: a) Goal setting, b) Environment structuring, c) Task strategies, d) Time management, e) Help seeking, and f) Self evaluation. Each subscale is assessed with a 5-point defined on a Likert-type scale, where 1 represents "strongly disagree" and 5 "strongly agree". Higher scores indicate that students have a high level of SRL. Table 2 presents the mean scores and standard deviations for each OSLQ subscale, based on the scoring methodology outlined by Ulfatun et al. (2021). Throughout the study, participants demonstrated varying degrees of SRL, aligning with findings from existing research on SRL dynamics (Zimmerman, 2013; Mou, 2023). Notably, individual scores varied across different subscales, reflecting the diverse strengths and challenges experienced by students in managing their learning. We also observe that participants may score higher on some subscales and lower on others. For example, one participant scored high on time management and environment structuring but low on help seeking.

Fig. 5 displays annual fluctuations in the student SRL, as captured by the OLSQ scale. During the spring semester start in March, there was a noticeable increase in goal setting and help seeking scores, while time management scores declined sharply. This reflects students' challenges in adapting to the new semester environment, particularly during the first month. However, as the semester ends, time management scores improve significantly, which may be related to students' intensive revision behavior for final exams.

During the summer break, students' task strategies and goal setting scores returned to normal levels, demonstrating a regular pattern of learning behavior. When the fall semester resumed in September, and students returned to campus, their scores in environment structuring declined, indicating that they may have difficulty readjusting to the on-campus learning environment. At the same time, the improvement in self evaluation scores shows that students became more accurate and proactive in evaluating their learning progress and effectiveness. Task strategies scores continued to increase during the fall semester, reaching a year-round high in December when environment structuring scores were at their lowest. This suggests that students can independently adopt effective learning strategies even when support from the external environment diminishes.

Table 2
Statistics analysis of OSLQ Scale.

Subscale	Response range	Overall mean (std)	Within-person range*
Goal setting	1–5	3.49 (1.23)	3 (2–4)
Environment structuring	2–5	4.07 (0.87)	3 (3–4)
Task strategies	1–5	3.79 (1.29)	2 (1–3)
Time Environment	1–5	4.24 (1.79)	2 (2–4)
Help seeking	2–5	4.78 (1.46)	3 (1–4)
Self evaluation	1–5	3.54 (1.43)	2 (2–3)

* Values are presented as the median (LQ-UQ), where 'LQ' stands for the lower quartile (25th percentile) and 'UQ' for the upper quartile (75th percentile).

The data for the full year show that SRL gradually improves during the semester and decreases slightly during the holidays. However, the overall improvement during the semester exceeds the regression during the holidays, showing a cyclical growth pattern of SRL. This observation highlights the important role of ongoing academic activity and a structured learning environment in the promotion of SRL.

4.3.2. Mobile sensing

The iSense application is adept at collecting various behavioral sensing features from raw sensor data, which are instrumental in depicting the daily behavior patterns of participants. These features span across Physical activity, Calls and SMS, Sleep, App usage, Location, and the ambient environment, providing a comprehensive view essential for understanding SRL. Next, we describe into the specifics of these features and explain their relevance to our analysis:

1. **Physical activity.** Physical activity features are crucial for assessing how students allocate their time between active and stationary states, which can influence learning efficiency. We utilize the Android Activity Recognition API to identify different states such as stationary, in-vehicle, on-bicycle, on-foot, running, tilting, walking, and unknown, with updates every 30 minutes. The focus is particularly on the time students spend stationary

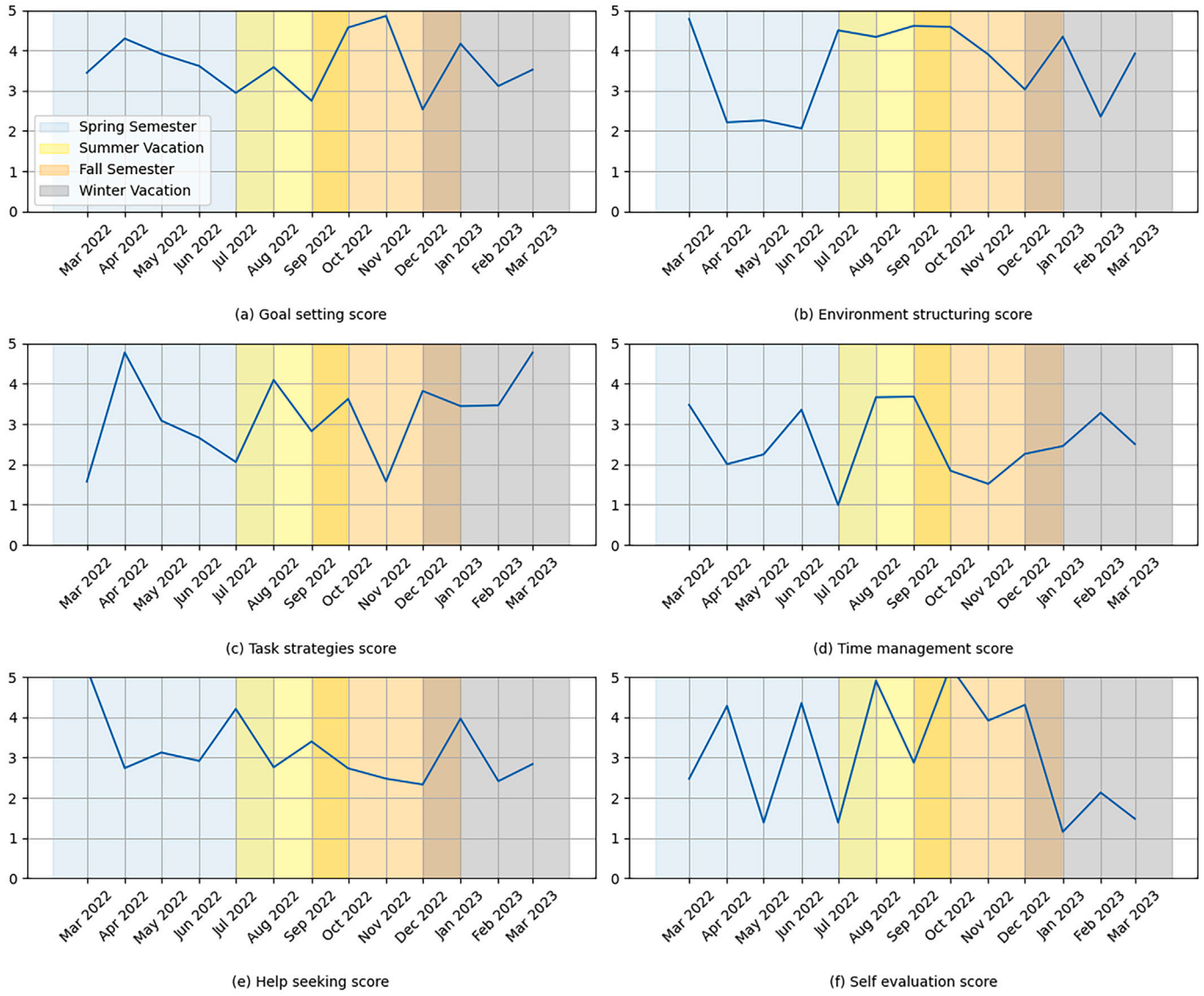


Fig. 5. Trends in OLSQ scale among participants. The plot is based on each assessment among participants' mean values and the 25th/75th percentiles.

and walking, which indicates potential study periods. The API, noted for its high accuracy (95%) in distinguishing between walking and stationary states (Wang et al., 2016), helps quantify daily physical activity effectively.

2. **Calls and SMS.** According to SRL theory, interpersonal interactions are significant for students seeking help during learning (Zimmerman, 2013). We track call duration and call/SMS frequency from logs to gauge social connections. These features reflect the frequency and duration of communications and provide insight into students' sociability, which can be critical for collaborative learning and support. Sociability-related features that are derived from calls and SMS inferences are incoming phone calls, outgoing phone calls, incoming SMS, outgoing SMS, and duration of communications regarding the use of applications for voice or video calls.
3. **Sleep.** Quality sleep is foundational for effective learning (Alshammari et al., 2023). We calculate three sleep features to evaluate sleep change: sleep duration, start time, and wake time. Sleep patterns are inferred using data from ambient light, audio amplitude, activity, and screen on/off, providing a comprehensive picture of students' sleep habits.
4. **Location.** Students' activity patterns are related to the learning environment created (Clough, 2009). By utilizing GPS and WiFi data, we analyze location and mobility patterns to infer whether students are engaging in academic or social activities. Using GPS for outdoor tracking and WiFi for indoor positioning, we analyze mobility and location data to proxy for student learning and social activities. We also incorporate a semantic understanding of university campus locations to identify and record the time spent in dorms, classrooms, canteens, study rooms, and other relevant campus sites. Related features derived from semantic location inferences are dorm, classroom, art & entertainment, canteen, study room, parks & outdoors, library, shop, gym, cafe, not found.
5. **App usage.** The types of apps students engage with can offer insights into their learning habits and distractions (Lin et al., 2021). Due to the large number of APPs, we do not use specific APPs as functions. Instead, we divide APPs into 26 categories. Such categories are available directly from the Android Market. The 26 categories are as follows: Education, Games, Art & Design, Beauty, Books, Comics, Communications(e.g., WeChat, QQ), Dating, Entertainment, Finance, Food & Beverage, Health & Fitness, Maps & Navigation, Browser, Medical, Music, News &

Magazines, Parenting, Personalization, Photography, Shopping, Social, Video Players & Editors, Weather, word, not found. Each app launched by a student is represented as a tuple that details the category of app used and the duration of usage. This categorization allows us to respect privacy while still gaining valuable insights into how students use their devices for educational and leisure activities.

4.4. Association analysis

In this study, we utilize Principal Component Analysis (PCA) (Schölkopf et al., 1997) and Generalized Estimating Equations (GEE) (Hardin & Hilbe, 2002) to explore the relationship between mobile sensing data and the scores from the OSLQ. These statistical methods are chosen for their effectiveness in reducing dimensionality (PCA) and handling longitudinal data (GEE), allowing us to draw meaningful inferences from the complex and multi-dimensional data collected.

4.4.1. Feature selection

The analysis of daily behavioral features using PCA is crucial for understanding the relationship between mobile usage patterns and SRL behavior. PCA was chosen because it effectively reduces dimensionality while preserving variance, thus making complex data more manageable and interpretable.

We began by calculating the average of each daily feature, such as app usage duration, across all participants throughout the year. This standardization ensures equitable representation of each participant's behavior. These averaged features were then compiled into matrix N, where each column represents an independent behavioral feature, and each row represents a dataset derived from 3578 OSLQ responses.

In subsequent analyses, we utilize scores from these PCs instead of direct behavioral features. This approach allows each PC to be interpreted as representing specific behavioral patterns based on its weighted features. For example, a PC with high positive weights on "the duration of using the entertainment app" can interpret high mobile phone usage.

Fig. 6 depicts the cumulative variance of features explained by the first n principal components. Ultimately, we select 17 PCs that together explain 91% of the variance for further analysis. These selected PCs not only simplify the data structure but also significantly enhance the model's capacity to elucidate SRL-related behavioral patterns.

4.4.2. Interpreting selected features

In this study, we apply GEE method to investigate the relationship between the 17 selected PCs and the OSLQ subscale scores. Given that the OSLQ data is collected repeatedly from participants over the course of the year-long iSense study, the responses from the same participant are likely to be correlated. GEE is an appropriate method for analyzing such longitudinal data, as it accounts for the within-participant correlation of repeated measures and provides robust estimates of population-averaged effects.

We conduct separate bivariate regressions between each of the 17 PCs and each of the six OSLQ subscale scores, resulting in 408 unique combinations (17 PCs × 6 OSLQ subscales). The GEE method is applied to these combinations to assess the relationship between behavioral patterns (represented by the PCs) and SRL behaviors (represented by the OSLQ scores). The GEE model is specified as follows:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \epsilon_{it}$$

where Y_{it} represents the OSLQ score for participant i at time t, X_{it} is the corresponding PC score, β_0 is the intercept, and β_1 is the regression coefficient. The term ϵ_{it} captures the residual error for participant i at time t. The GEE approach estimates the population-averaged coefficients, accounting for the correlation structure within the data (i.e., within participants).

To control for the false discovery rate (FDR) in this exploratory regression analysis, we apply the Benjamini-Hochberg procedure (BH) (Benjamini et al., 2006). This procedure adjusts the p-values obtained from the regressions to account for multiple comparisons and reduces the likelihood of Type I errors. By controlling the FDR, the BH procedure helps ensure that the significant relationships between behavioral features (represented by the PCs) and SRL subscales (represented by the OSLQ) are reliable and not due to random fluctuations in the data. This method enables us to focus on the most meaningful associations while minimizing the risk of false positives.

4.5. Predicting OLSQ subscale

We explore various methods to identify the optimal model for predicting SRL. Specifically, our experiments involve testing different time windows for data aggregation and employing various deep-learning models. In the following sections, we detail the methodologies implemented in our study.

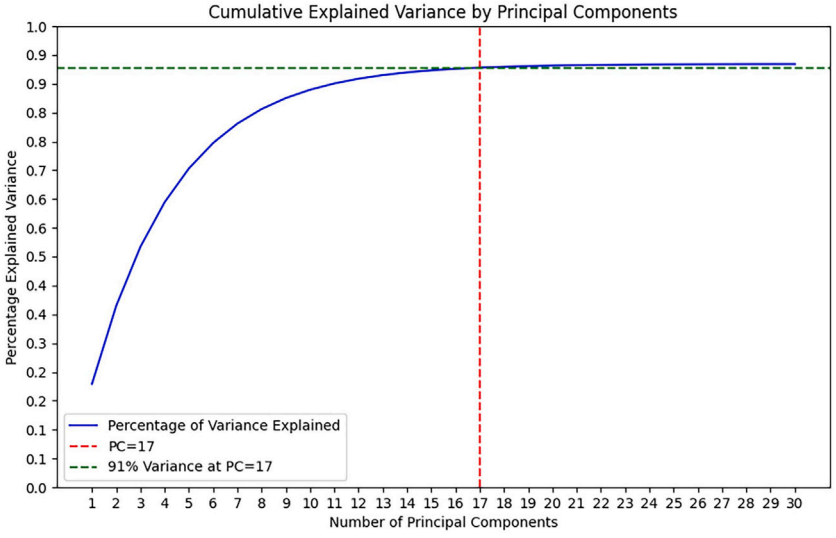


Fig. 6. The total variance in the sensed data accounted for by the top-n principal components.

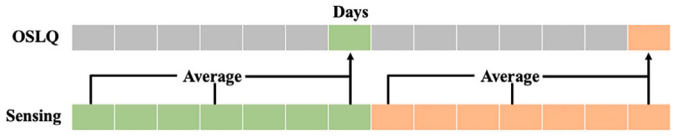


Fig. 7. Assess the prediction performance with 7-day time windows.

4.5.1. Time window analysis

Based on the observed annual fluctuations in student SRL, we recognize that these behaviors are not static but vary over different periods due to changes in the academic environment and students' adaptive processes. To capture these dynamics effectively, we experiment with multiple time windows: 7-day, 14-day, 21-day, and 30-day intervals—that align with the cyclical patterns seen throughout the year.

Shorter time windows (7-day) allow us to detect immediate, reactive changes in SRL behaviors, such as the initial adaptation struggles at the start of a semester or the sharp adjustments seen during specific academic events like the start of the spring or fall semesters. Meanwhile, longer windows (14-day, 21-day, and 30-day) help to capture more comprehensive patterns that evolve gradually, such as the steady improvement in time management skills as the semester progresses or the changes in environment structuring and self-evaluation scores over more extended periods.

Our approach aligns with the periodicity of the OSLQ self-reported survey, which is administered every seven days. By evaluating the predictive performance of our model using these different time intervals, we can correlate the aggregated daily features—computed as the average of each feature within each time window—with the corresponding OSLQ scores collected at the end of each period. This method enables us to understand both short-term fluctuations and longer-term trends in SRL behaviors. Fig. 7 illustrates this data aggregation strategy, using a 7-day window as an example: for each participant, we calculate the mean of each feature (e.g., average learning duration) from day 1 to day 7 to correlate with the OSLQ scores at the end of that period.

4.5.2. Evaluation of deep learning models

We use several deep learning models to model behavioral temporal patterns and assess SRL, including Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Long Short-Term Memory Network (LSTM), and Temporal Convolutional Network (TCN).

Multilayer Perceptron (MLP): The first model we consider is the MLP, known for its straightforward architecture. The MLP in this study features an input layer, two hidden layers, and an output layer. Each layer is fully connected to the next, allowing the model to learn deep representations of the input data. The input layer size matches the number of extracted features from the mobile sensing data and self-reports, while each hidden layer contains 128 neurons. ReLU activation functions introduce non-linearity in the hidden layers. The output layer consists of six neurons corresponding to the six OLSQ subscales, employing a sigmoid activation function.

Convolutional Neural Network (CNN): The second model we consider is CNN. In this study, we design the CNN architecture with multiple convolutional layers to extract higher-level features indicative of SRL behaviors. Each convolutional layer applies learnable filters that identify crucial temporal patterns, followed by a max pooling layer that reduces data dimensionality. This reduction not only enhances the model's generalization capabilities by focusing on dominant features but also decreases the computational demand. The output from the final pooling layer is flattened into a one-dimensional array, which feeds into a fully connected dense layer. This layer integrates the learned features to predict the scores for each OSLQ subscale. The output layer consists of six neurons with a sigmoid activation function, providing a probabilistic estimate of each subscale's score.

Long Short-Term Memory (LSTM): Next, we consider the LSTM, which excels at capturing temporal dependencies crucial for predicting SRL as indicated by the OLSQ subscales. The LSTM architecture in our study processes temporal sequences from mobile sensing data to predict continuous OLSQ subscale scores. The input layer receives the raw features structured as sequences over time. Multiple LSTM layers are stacked to enhance the model's ability to learn complex temporal patterns. Each LSTM unit includes a cell state and three gates (input, output, and forget gates) that manage information flow. The output from the LSTM layers is then passed to a dense layer, integrating the learned features into predictions for each OLSQ subscale. The final output layer comprises six units with sigmoid activation functions, producing probabilistic outputs between 0 and 1 for each subscale.

Temporal Convolutional Network (TCN): Finally, we consider is the TCN, a convolutional neural network capable of capturing long-term dependencies. The causal convolution layer captures long-term dependencies, while the output layer is used for final prediction. The TCN architecture we employ in our study is the causal convolution layer, which enables the model to efficiently capture long-term dependencies in sequence data without introducing future information. Multiple layers of 1D convolutions enable the TCN to learn spatial hierarchies of features. Each layer applies filters that capture patterns across different timescales. To increase the receptive field without losing resolution or coverage, we use dilated convolutions. This allows the network to learn from data points that are further apart, capturing long-term dependencies. These connections help train deeper networks by allowing gradients to flow through the network directly, preventing the vanishing gradient problem. Ensures that the convolution output for a specific time step is not influenced by the data from the future, adhering to the causality principle.

4.5.3. Evaluation metrics and model validation

In this study, we utilize Bayesian optimization to efficiently determine the best model configurations by maximizing function f at the sampling point:

$$x^{+} = \arg\max_{x \in M} f(x) \tag{1}$$

where M represent the search space,and x is the hyperparameter set.

To quantitatively assess the accuracy and effectiveness of the developed models in predicting SRL behaviors from mobile sensing data, we employ several evaluation metrics.

Mean Absolute Error (MAE): measures the average magnitude of errors in predictions:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{2}$$

where y_i represent the true value and $\hat{y}_i$ is the predicted value.

Root Mean Square Error (RMSE): emphasizes larger errors by squaring the differences before averaging:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{3}$$

Pearson Correlation Coefficient: quantifies the linear correlation between actual and predicted values:

$$\rho_{y,\hat{y}} = \mathrm{cov}(y, \hat{y})/\sigma_y \sigma_{\hat{y}}, \tag{4}$$

where cov represent the covariance, and σ_y and $\sigma_{\hat{y}}$ are the standard deviations of the actual and predicted values, respectively.

To better understand model performance under various educational scenarios, we assess the accuracy using Leave-One-Subject-Out (LOSO) and Specific-Leave-One-Subject-Out (SLOSO) cross-validation techniques. In LOSO, the model is trained on the data from all participants except for one, effectively simulating the introduction of a new student without any prior SRL data. This helps evaluate how well the model can adapt to new, unseen students.

Additionally, we use SLOSO, which involves training the model using one student's initial data point, along with the complete data sets from other participants, to predict that student's later responses. This approach is particularly relevant in practical educational settings where an application might continuously monitor and adapt to a student's social and SRL behaviors over time.

5. Results and discussion

5.1. RQ1: What are the behavioral patterns related to SRL in students' daily lives?

5.1.1. Behavioral patterns analysis

As previously detailed, we examine the relationship between the scores of 17 PCs and the OLSQ subscale scores using GEE and FDR. Each PC is labeled based on its loadings, which represent the weight associated with features. We concentrate on the highest-ranked behavioral features to gain insights into the typical behaviors represented by each PC. Our findings indicate that 11 out of the 17 behavioral patterns (PCs) are correlated with OLSQ subscales.

We detail the 11 behavioral patterns related to the OSLQ subscales in Table 3, which shows the features of each behavioral pattern indexed by PCA. Specifically, Pattern 1 describes students who frequently make phone calls or use social apps to chat in the library, classroom, and study room; Pattern 2 describes students who usually use social media to socialize with others in the library, classroom, and study room; Pattern 3 describes students who use educational software (such as office applications, collaboration applications, etc.) in the library, classroom, and study room; Pattern 4 describes students who spend more time listening to music in the library, classroom, and study room; Pattern 5 describes students who like to spend time in the library, classroom, and study room, sleep late, and do not play games; Pattern 6 describes students who are often in the library and travel longer distances in the evening and at night; Pattern 7 describes students who like to be active in different places and spend less time in the dorm; Pattern 8 describes students who like to watch live broadcasts on their mobile phones in the dorm and often order takeaways; Pattern 9 describes students who spend more time studying in their spare time; Pattern 10 describes students who engage in frequent browsing of entertainment, beauty, lifestyle content, and shopping activities on their smartphones, particularly before bedtime; Pattern 11 describes students who tend to get up early and enjoy jogging outdoors during the day.

5.1.2. Associations between behavioral patterns and OLSQ subscales

In this section, we delve into the associations between selected PCs and the scores on the OLSQ to identify behavioral patterns closely linked with SRL assessment. Table 4 presents the correlations between these behavioral patterns and the six OSLQ subscales. The positive correlation between behavioral patterns and subscales indicates that students with these behavioral patterns tend to achieve higher scores in the corresponding subscales, and thus have higher SRL skills, and vice versa. Next, we describe our findings.

The Goal setting subscale measures an individual's ability and habit of setting goals in school and daily activities. It asks how a person sets learning goals, schedules study time, prioritizes study tasks, and adjusts study plans to accommodate new situations. We found that the goal setting subscale was negatively correlated with Pattern 5 (preferring late-night activities) and positively correlated with Pattern 9 (increasing study time outside normal hours) and Pattern 11 (early risers, preferring morning jogging). Our results showed that participants who spent more time in the study settings, had higher sleep quality, and used their mobile phones less scored higher on the goal setting subscale.

The Environment structuring subscale measures how students organize and structure their learning environment to optimize learning. This involves choosing a learning location, arranging physical and digital resources, and setting up an environment conducive to learning. We observed that the environment structuring subscale had a positive

Table 3
Behavioral patterns and their corresponding features represented by the n-th principal component of PCA.

Pattern	Features
1. Use applications for calls/communication frequently in the library, classroom, and study room	high call-in-and-out numbers during the library, classroom, and study room, high unlock numbers, less sedentary duration,durationof using SMS and communication apps.
2. Use the phone for social media apps in the library, classroom, and study room	high SMS-in-and-out numbers during the library, classroom, and study room, high unlock numbers, spend increased time on social media apps (i.e., Blog, etc.).
3. Use study apps frequently in the library, classroom, and study room	spending more time the library, classroom, and study room, spending less time in the dorm, having shorter sleep periods, longer phone usage duration, playing fewer games but frequently using web browsers and music/video apps.
4. Spend additional time on music and be sedentary in the library, classroom, and study room	showing up more frequently in the library, classroom, and study room, duration of sedentary, less time at the dorm, early to bed, longer lock duration, using more music apps.
5. Spend time in the library, classroom, and study room and sleeping late and gaming less	less time playing games, not often playing game, showing up more frequently in the library, classroom, and study room, earlier sleep start time.
6. Frequent in the library and walk long distances to the library during the evening	showing up more frequently in the library, frequent fewer locations in the evening hours, rarely engage with games, video, and interactive apps, however, use social network apps.
7. Visit and spend time in different places at night	longer unlock duration, spending less time on games but frequently using browser and music/video applications, more active on walk.
8. Spend substantial time to watching videos and consuming food or beverages in the dorm	longer unlock duration, going to bed early and waking up early, less physically active, particularly during study times, frequent utilize food and drink apps, particularly those that offer food delivery services more usage of engagement apps.
9. Spend time in the study room at spare time	showing up more frequently in study room, frequent fewer locations at spare time, less time at dorm, shorter sleep duration, longer lock duration, using learning app or music apps.
10. Engage in shopping and browsing content related to lifestyle sharing, entertainment and beauty around bedtime	The duration of using shopping, lifestyle, entertainment, and beauty apps, as well as the time when sleep begins.
11. Prefer to wake up early and enjoy jogging outdoors during daylight hours	early sleep end time, sleep duration, jogging, spending time in brighter ambient light places in the morning and afternoon.

Table 4
Patterns associated with OLSQ scale.

Sub-scale	Associated behavioral pattern
Goal setting	pattern 3 (+), **pattern 9 (+)**, **pattern 5 (−)**, pattern 6 (+), pattern 10 (−), pattern 11 (+)
Environment structuring	pattern 4 (+), pattern 5 (−), pattern 7 (−), pattern 6 (+), pattern 8 (−), **pattern 9 (+)**
Task strategies	pattern 3 (+), pattern 5 (−), pattern 9 (+)
Time management	pattern 1 (−), pattern 2 (+), pattern 4 (−), pattern 5 (−), pattern 6 (+), **pattern 9 (+)**
Help seeking	pattern 1 (+), pattern 2 (+), pattern 3 (+), pattern 5 (−), **pattern 8 (−)**, pattern 9 (+)
Self evaluation	pattern 9 (+), pattern 11 (+)

Negative association (−), Positive association (+), Associations involving $p < 0.05$. FDR <0.1 in bold.

correlation with Pattern 6 (frequent library visits and extensive travel in the evening) and Pattern 9 (increased study time outside normal hours). In contrast, it showed a negative correlation with Pattern 5 (tendency for late-night activities) and Pattern 8 (preference for watching live broadcasts and ordering takeout in the dorm). Our results showed that participants who visited different places and spent longer walking times scored lower on the environment structuring subscale.

The Time management subscale measures how individuals arrange and manage time to complete study tasks. This includes planning study time, setting deadlines, and balancing study with other activities. We found that the time management subscale was positively correlated with pattern 9 (increasing study time outside normal hours) and pattern 1 (frequently making phone calls or using chat applications during study) and negatively correlated with Pattern 4 (mainly listening to music while studying). We found that a number of traits were related to the time management subscale. Generally, higher scores on the time management subscale were associated with spending more time in educational settings (especially in study spaces), being sedentary, frequent use of educational/browser applications and music/video applications, and spending time in dorm time is less relevant. Interestingly, we observed that students who prefer to use catering apps tend to have low time management skills.

The Help seeking subscale assesses students' tendency and frequency to seek help during learning, such as asking teachers, classmates, or other resources to understand learning materials or solve learning problems. We found that the Help seeking subscale was positively correlated with pattern 2 (extensive use of social media for socializing during study) and pattern 3 (use of study-oriented software during study), and negatively correlated with Pattern 8 (preference for watching live broadcasts and ordering takeout in the dormitory). Our results showed that participants who scored higher on the help seeking subscale were more likely to rely on internet searches, online videos, and instant messaging tools to seek help from others.

The Task strategies subscale measures how students select and apply different learning strategies to handle learning material. This includes using methods such as generalizing, reorganizing information, and applying memory techniques. We found that the task strategies subscale was positively correlated with pattern 3 (use of learning-oriented software during study) and pattern 9 (increased study time outside of normal hours). Students who actively participated in social learning and used learning tools scored higher on the task strategies subscale, demonstrating that using additional learning tools further stimulated their active learning.

The Self evaluation subscale assesses students' ability to evaluate their own learning progress, understanding, and learning effects during the learning process. We found that the self evaluation subscale was positively correlated with pattern 9 (increasing study time outside normal hours) and pattern 11 (early riser, preferring morning jogging). Similar to the Help seeking subscale, students who scored higher on the Self evaluation subscale tended to utilize other students' online discussions or instant messaging programs to deepen their understanding of the content and examine their ideas. Furthermore, we found fewer perceived behavioral features associated with the Self evaluation subscale, possibly because Self evaluation skills are often internal and difficult to capture through passive sensing.

In summary, our results demonstrate that mobile sensing can be used to assess SRL. Important indicators of SRL include sedentary time, type of application usage, number of places visited, and duration in dorm and study places.

5.2. RQ2: Is it feasible to predict students' SRL using passive behavioral sensing assessments obtained from smartphones?

5.2.1. Experimental setup

In this section, we discuss the hyperparameter tuning results and prediction performance for each deep learning model separately. By tuning different hyperparameters, we obtain optimal model configurations and evaluate their performances in the task of predicting SRL.

Table 5
Result of the tuning process for MLP, CNN, LSTM and TCN.

Model	Layers	Unit/filter size	Nodes/dense nodes	Kelnel size	Dropout	Epochs	Batch
MLP	2	100	32	–	0.764	64	128
CNN	1	256	64	16	0.486	64	256
LSTM	1	64	64	–	0.861	32	128
TCN	1	32	–	4	0.803	64	256

We employ Bayesian optimization to determine the optimal combination of hyperparameters for each model. This optimization method utilizes a Gaussian process to decide which combination of hyperparameters is more beneficial at each iteration. We achieve this using functionalities supported by the Scikit-Optimize 5 library. Each tuning process was evaluated up to 100 times on different hyperparameter configurations. The tuning process was terminated early if the surrogate model converges before 100 evaluations. Ultimately, the tuning process produced a hyperparameter configuration that achieved optimal performance. Each proposed model has unique hyperparameters that need to be tuned. However, all models share some hyperparameters, such as the number of epochs and batch size, which are related to the number of iterations and the iteration method of the training data set. Table 5 shows the results of the hyperparameter tuning related to the proposed model, as well as the prediction performance.

5.2.2. Prediction performance analysis

Usually, in small sample datasets, LOSO and SLOSO are the best choices for evaluating the performance of the model from a small human-centered dataset in existing studies (Wang et al., 2020, 2018a, 2016). However, the estimation is also very specific for this particular dataset, which may lead to larger variance compared to the model's performance on a new dataset. Therefore, we also used 5-fold cross-validation. For the OLSQ scale, we employed ten times of 5-fold cross-validation and reported the average values. We also compared our predictive model with three baseline models. Baseline 1 predicts by using the mean of the scores as the prediction. This is the simplest regression model with only an intercept (Martin et al., 2022). Baseline 2 randomly generates a sample from the already known distribution (i.e., the OLSQ score distribution) and uses it as the prediction value (Chen, 2022). Baseline 3 predicts by using the most frequent OLSQ subscale score among all samples as the prediction for each instance (Brokenshire & Kumar, 2009). These three baseline methods provide a performance threshold for comparison with more complex models (Poliak et al., 2018).

We present in Table 6 the best combination of time windows and prediction models for predicting OLSQ subscale scores. The performance reports include MAE, RMSE, Pearson correlation coefficient, and the correlation between predicted and actual values across cross-validation (grouped by 5-fold). The baseline models set the minimum performance threshold for the prediction models. For example, Baseline 1 achieved an MAE of 4.15 and RMSE of 4.75 for "Environment Structuring," Baseline 2 had an MAE of 3.85 and RMSE of 4.51, while Baseline 3 performed slightly better, achieving an MAE of 3.25 and RMSE of 4.35. These results remain consistent across other subscales, providing a reference point for evaluating the predictive model's performance. The predicted OLSQ scores show a positive correlation with their actual responses. All prediction models performed better than the baseline models.

Overall, the TCN model demonstrated superior performance compared to other models, indicating high accuracy in predicting SRL. This outcome suggests that the TCN model is particularly effective in capturing SRL-associated behavioral patterns, as evidenced by a significant positive correlation between the predicted OLSQ scores and the

Table 6
Patterns associated with OLSQ scale.

Sub-scale	Our model window/method	5-fold cross validation					LOSO		SLOSO	
		Our model MAE/RMSE	Baseline 1 MAE/RMSE	Baseline 2 MAE/RMSE	Baseline 3 MAE/RMSE	Pearson r- value	Pearson r- value	Per-user error Mean (std)	Pearson r-value	Per-user error Mean (std)
Environment structuring	21/LSTM	**2.88/1.09**	4.15/4.75	3.85/4.51	3.25/4.35	0.54	0.48	2.90 (2.85)	0.52	2.82 (1.78)
Goal setting	30/TCN	**4.85/2.06**	6.25/5.25	5.15/4.85	4.15/5.05	0.56	0.50	3.85 (3.80)	0.55	2.75 (1.75)
Time management	30/TCN	**2.82/1.03**	5.50/3.50	4.35/3.60	3.90/3.60	0.57	0.52	1.78 (1.77)	0.58	2.70 (1.72)
Help seeking	14/TCN	**4.86/3.07**	8.10/4.10	6.85/4.20	5.50/4.30	0.55	0.49	3.88 (2.82)	0.53	3.80 (2.76)
Task strategies	7/LSTM	**7.89/5.11**	13.52/5.20	12.25/5.50	11.10/6.40	0.53	0.47	9.93 (5.84)	0.51	12.85 (6.79)
Self-evaluation	7/TCN	**10.90/9.12**	13.52/8.30	12.50/7.95	12.23/7.54	0.51	0.45	9.98 (7.88)	0.50	6.90 (4.83)

actual observed values. In particular, all predictive models exceeded the performance benchmarks set by the baseline model, highlighting their enhanced predictive capabilities.

To better understand the performance of the model under various scenarios, we analyze the accuracy of the LOSO and SLOSO strategies. Generally, we found that the LOSO approach underperformed, indicating that our dataset may not sufficiently capture individual differences necessary for effective model training using only data from other participants. This issue is somewhat expected since the OLSQ subscales are composed of items that measure specific functionalities from multiple behaviors, which can be highly personalized.

To mitigate this, we incorporate each participant's first chronological response as an additional data point in the SLOSO strategy. This approach significantly improved performance, making it comparable to the results of 5-fold cross-validation. We also calculate the MAE for each participant during the LOSO and SLOSO processes, allowing us to analyze the distribution of per-user errors. The findings showed that the absolute errors per participant from SLOSO consistently remained below 15% of the total range for the majority of OLSQ subscales. This enhanced error distribution demonstrates the effectiveness of SLOSO in capturing and predicting based on broad behavioral trends and specific individual differences.

Our analysis reveals that utilizing behavioral data from specific time windows significantly enhances the predictive accuracy for certain OLSQ subscales. For subscales like goal setting and time management, which focus on students' short-term and long-term planning and adherence to schedules, a 30-day window was most effective. This longer window captures a broader range of behaviors related to consistent goal-oriented activities. Conversely, for subscales that assess immediate conditions and do not specify a time range—such as environment structuring and task strategies—shorter time windows yielded better predictions. These subscales benefit from capturing more dynamic, current state behaviors; thus, a more immediate snapshot of student activity is preferable for accurate assessments.

Additionally, while previous research suggested that monitoring changes or trajectories in perceived behavior could provide valuable insights for assessing SRL, we found that incorporating these dynamic features did not significantly enhance our model's evaluations. We hypothesize that these trajectory-based features are not critical for assessing SRL, as they predominantly relate to recent or past learning activities, which may not directly influence current SRL performance. Consequently, we adopt a more streamlined approach, utilizing only the average values of perceived behaviors from the days preceding the assessment date (specifically, $d = 7, 14, 21, 30$ days) as predictors in our models. This simplified method focuses on static snapshots of behavior, which our findings suggest are more predictive of SRL outcomes.

After considering the features of each subscale, we identify optimal prediction windows for each. This validation affirms that mobile sensing technology can successfully captures daily behaviors for automatic assessment. For subscales focused on the current state without a specified time range, such as environment structuring and task strategies, shorter time windows were more beneficial. However, predictions regarding self evaluation were the least accurate, possibly because mobile sensors struggle to capture intrinsic psychological changes. Future research might explore using a broader array of mobile sensors to better capture data related to self evaluation.

6. Case study

We focus on a case study to demonstrate the effectiveness of the model. The student in our case study is a male student from the School of Physical Education who reported at the beginning of the study that he usually woke up at 8 am, attended classes, and had lunch in the cafeteria; in the afternoon, he attended classes and exercised; in the evening, he played games and went to bed at 11 pm. During the initial two months, he was confident in his SRL abilities. However, significant changes occurred in his daily routine according to the assessment in May 2022. During that period, he typically slept until 2–3 in the afternoon, ordered food using delivery apps, attended classes, played games with friends in the evening, and then going to bed in the early morning. He still believed that he had the ability to pass the exam, but he rarely chose study locations and allocated additional study time. By the assessment in June 2022, his situation worsened. He spent almost all his mornings and afternoons playing games and watching videos in his dorm, ordering food using delivery apps. He anticipated difficulties in studying, resulting in frequent absences from classes and group discussions during that period. Throughout the autumn semester (September to December 2022), his behavior patterns did not significantly improve. Even under pressure from mid-term and final exams, he maintained a relaxed attitude towards studying. Daily activities continued to focus on gaming and videos, with occasional participation in classes and training. His sleep patterns also did not revert to the regular schedule observed at the beginning of the spring semester. The predicted values shown in Fig. 7 indicate that around September 2022, his scores on the OLSQ subscales began to decline. During the summer break (July to August 2022), he stayed on campus for physical training. Although there was some improvement in his sleep, he did not engage in systematic study preparation. More time was spent on leisure activities such as sports, gaming, and watching videos, with no significant increase in study time. Over the winter break (December 2022 to February 2023), he returned home but did not significantly increase his study time despite the absence of campus life distractions. His focus on studying remained low throughout the entire winter break. At the beginning of the spring semester in 2023 (February 2023), he attempted to readjust his daily routine and study habits, but with limited success. Despite attending more classes, his after-class time was still primarily spent on gaming and entertainment, resulting in low study efficiency. In his self-evaluation, he expressed some frustration and acknowledged the need for more self-discipline to improve his academic performance.

Fig. 8 illustrates how the SLOSO model predicted this student's SRL. The initial assessment was included in the training set alongside knowledge from the population. From the predicted values, we observe: (a) The participant maintained low study time throughout the study period. (b) Around September 2022, his OLSQ sub-scale showed deterioration.

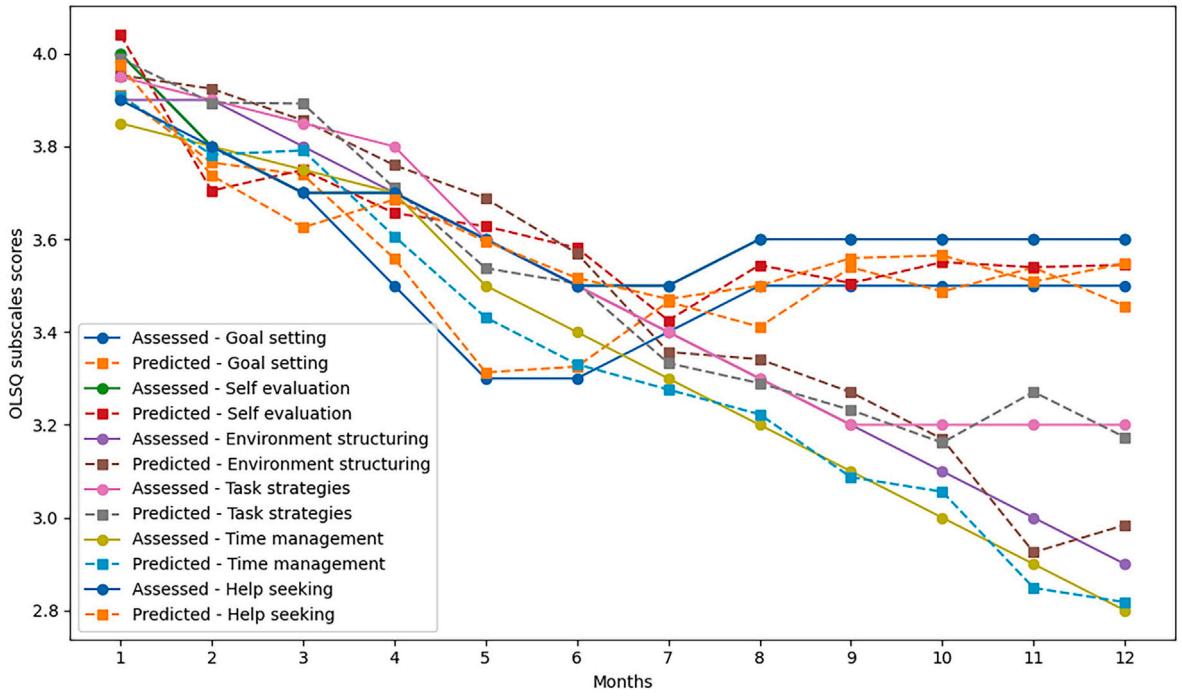


Fig. 8. Case study: The dynamics of a student's assessed and predicted SRL using SLOSO.

Although the SLOSO model still shows predictions biased towards initial inputs, researchers should focus closely on the changes highlighted by the model. This case study demonstrates the importance and effectiveness of the proposed system. Through this year-long observation, we identified several critical factors in the student's lifestyle influencing his SRL behaviors. Particularly, his sleep patterns, app usage characteristics, and daily activity habits showed significant changes closely related to fluctuations in his study efficiency. This case study demonstrates the potential of the proposed system in assessing and predicting SRL behaviors among university students, providing important insights for further research and practical applications. While acknowledging the limitations in dataset size, we recognize the potential for further enhancement of the model through inclusion of data from additional participants. Nonetheless, the system has already demonstrated its ability to detect behavioral patterns within the OLSQ subscales and track these dynamic changes.

7. Discussion

We employ a methodology that integrates mobile sensing and computational interaction to more effectively capture and comprehend students' learning behavior patterns associated with SRL. We identify a set of behavioral patterns that show positive or negative associations with the target psychological measures. To our knowledge, our work represents the pioneering use of mobile sensing to investigate SRL in a real-world setting. Appropriately, the methods we propose for assessing SRL in this paper could be applicable to other psychological measures in education or related fields, such as evaluating students' perceived autonomy in their learning process. Our research offers a novel approach to developing new assessment methods and opens avenues for timely personalized interventions in domains relevant to the human-computer interaction (HCI) community.

7.1. Implications for assessing SRL through mobile social behavioral sensing in education

We adopt computational methods to improve the assessment and understanding of SRL from mobile sensing data encompassing behavioral features. Our approach highlights the benefits of utilizing mobile sensing data to assess specific psychological variables by continuously capturing various components of behaviors associated with SRL through multiple modes. Furthermore, our approach demonstrates the utilization of diverse deep learning architectures and classical PCA techniques to classify and interpret behavioral patterns linked to psychological measures. This process serves as a crucial foundation for assessing SRL and other specific psychological variables within educational contexts. It encourages researchers to contemplate future possibilities for integrating mobile sensing into the design and development of predictive models and intervention systems.

7.2. Significance for human-computer interaction (HCI) and design

We believe our approach introduces new directions for assessing SRL to better understand students during the learning process and, consequently, improve their overall learning experience. Hopefully, for future research, smartphone-based learning assisting systems combined with passive assessing will be able to detect SRL levels in real-time to inform and help students with timely intervention.

Moreover, our approach offers another implication for designing real-time automated learning analytic tools that could be integrated into the educational or psychological systems for passive assessment of various psychological variables in education. For example, in the teaching scenario, by leveraging PCA of sensor data from smartphones, teachers could be informed promptly and gain a clearer understanding of students' behavioral patterns linked to their evolving levels and gain a clearer understanding of students' behavioral patterns linked to their evolving levels of SRL and psychological states. It also provides opportunities for educational institutions (e.g., universities) to integrate such SRL sensing techniques into their systems for timely prediction and personalized intervention; for example, once the system detects the students are in a low level of SRL, a notification would be sent to their teachers and tailored interventions that are designed to match their behavioral habits would be triggered. Furthermore, our study ensures that no privacy-sensitive information, such as text message content or voice call records, is collected. All student phone numbers and IDs are anonymized, and participants have the option to withdraw their consent

at any time. In this way, our approach not only pushes forward the design of novel learning assistance systems in educational research but also avoid privacy concerns.

7.3. Implications for sensing students in need

On a broader level, mobile sensing devices such as smartphones allow us to gain deeper insights into students' behavioral patterns associated with their psychological states. This continuous and real-time assessment capability enables us to identify students in need across different scenarios, for example, under the pandemic condition, learning in the authentic classrooms may not be possible, blended learning such as online learning or self-paved offline learning may shift to the new normal in education. Under such circumstances, teachers or instructors may not always be able to promptly notice and observe students, possibly causing the lack of support or intervention to the students in need.

8. Conclusions

The outcomes of this study represent a significant advancement in using passive smartphone sensing data to predict SRL. This research contributes in several ways, such as designing a streamlined mobile sensing platform that minimizes user burden, ensures high compliance, and can operate continuously for up to a year, providing new research insights into the learning behaviors of university students, and evaluating various design factors of predictive models. By comparing and analyzing different models, we aim to accurately predict SRL behaviors.

From a system development perspective, we improved the design of the iSense application by innovatively integrating self-reported survey push notifications within the Android sensing application. This integration ensures both high compliance and minimal battery consumption. This integration also allows researchers to contact students immediately when missing data is detected. The studies that included user input played a critical role in ensuring the reliability of the data collection phase. As a result, we noted considerably higher retention rates and improved data quality compared to previous long-term mobile sensing studies. We believe our experiences and insights can be valuable for other field researchers.

From a student well-being standpoint, we explored how different behaviors on campus before university influenced learning outcomes across various semesters and holiday periods. While previous studies have investigated numerous factors concerning student learning behaviors, they have primarily relied on surveys and occasional self-reports. By collecting data continuously and passively throughout the year, we provided unprecedented data and insights. Furthermore, our research delved into the predictable characteristics and behavioral patterns of SRL, offering important clues and insights for further research.

We anticipate that the findings of this study will inspire more research in related fields, promoting better dataset construction and model design. Through deeper research and practice, we believe that predictive models for SRL will become more accurate and reliable, providing more support for data-driven decision-making in education.

9. Limitations and future work

Despite the significant results achieved, our study has a few limitations that should be considered. First, our research focuses on a specific student population from a single university in China, which may limit the generalizability of our findings. To address this, we suggest that future studies extend the sample to include students from different universities and countries, as well as different educational settings (e.g., high school students or online learners), in order to verify the applicability of the findings in broader contexts.

Second, while the study leverages mobile devices for passive sensing data collection, we acknowledge that other multimodal sensing technologies could have contributed to a more comprehensive analysis. In particular, integrating data from a broader array of sensor devices—such as wearable devices or environmental sensors—could have provided additional insights into the multidimensional nature of SRL behaviors. Future research should explore the potential of combining mobile sensing with other sensing technologies to better capture the complexity of SRL.

These limitations also open up valuable avenues for future work. By expanding the participant pool and incorporating additional sensors, future studies can deepen our understanding of the factors that influence SRL across diverse populations and educational environments. Moreover, the integration of multimodal data would enable a more holistic approach to SRL prediction, leading to more accurate and robust models. We believe that these future directions will enhance the applicability and predictive power of SRL assessment methods, contributing to more personalized educational interventions.

CRediT authorship contribution statement

Tongyu Zhao: Writing – review & editing, Writing – original draft, Software, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Jiaying Gao:** Writing – review & editing, Software, Methodology, Formal analysis. **Yu Feng:** Writing – review & editing, Software, Methodology. **Yatong Zu:** Validation, Supervision, Methodology. **Adriano Tavares:** Validation, Supervision, Methodology. **Tiago Gomes:** Validation, Supervision, Methodology. **Sandro Pinto:** Validation, Supervision, Methodology. **Hao Xu:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (No. 62077027), the Department of Science and Technology of Jilin Province, China (No. 20230201086GX, No. 20200801002GH), the Education Department of Jilin Province, China (No. JJKH20200993K), the Ministry of Education of the People's Republic of China (No. 2021180010)., and the Humanity and Social Science Youth Foundation of Ministry of Education of China (No. 20YJC190034).

Data availability

The data that has been used is confidential.

References

Ahuja, K., Shah, D., Pareddy, S., Xhakaj, F., Ogan, A., Agarwal, Y., Harrison, C., 2021. Classroom digital twins with instrumentation-free gaze tracking. In: Proceedings of the 2021 Chi Conference on Human Factors in Computing Systems, pp. 1–9.

Alshammari, T.K., Rogowska, A.M., Basharahil, R.F., Alomar, S.F., Alseraye, S.S., Al Juffali, L.A., Alrasheed, N.M., Alshammari, M.A., 2023. Examining bedtime procrastination, study engagement, and studyholism in undergraduate students, and their association with insomnia. Front. Psychol. 13, 1111038.

Araka, E., Maina, E., Gitonga, R., Oboko, R., 2019. A conceptual model for measuring and supporting self-regulated learning using educational data mining on learning management systems. In: 2019 IST-Africa Week Conference (IST-Africa). IEEE, pp. 1–11.

Araka, E., Maina, E., Gitonga, R., Oboko, R., 2020. Research trends in measurement and intervention tools for self-regulated learning for e-learning environments—systematic review (2008–2018). Res. Pract. Technol. Enhanc. Learn. 15, 1–21.

Aslan, Sinem and Alyuz, Nese and Tanriover, Cagri and Mete, Sinem E and Okur, Eda and D'Mello, Sidney K and Arslan Esme, Asli, 2019. Investigating the impact of a real-time, multimodal student engagement analytics technology in authentic classrooms. In: Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems, pp. 1–12.

Azevedo, R., Gašević, D., 2019. Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: issues and challenges.

Azevedo, R., Moos, D.C., Johnson, A.M., Chauncey, A.D., 2010. Measuring cognitive and metacognitive regulatory processes during hypermedia learning: issues and challenges. Educ. Psychol. 45, 210–223.

Bandura, A., 1991. Social cognitive theory of self-regulation. Organ. Behav. Hum. Decis. Process. 50, 248–287.
Barnard, L., Lan y, W., To, Y.M., Paton, V.O., Lai, S.-L., 2009. Measuring self-regulation in online and blended learning environments. Internet High. Educ. 12, 1–6.
Benjamini, Y., Krieger, A.M., Yekutieli, D., 2006. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93, 491–507.
Boukhechba, M., Daros, A.R., Fua, K., Chow, P.I., Teachman, B.A., Barnes, L.E., 2018. Demonicsalmon: monitoring mental health and social interactions of college students using smartphones. Smart Health 9, 192–203.
Bringmann, L.F., Pe, M.L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., Tuerlinckx, F., Kuppens, P., 2016. Assessing temporal emotion dynamics using networks. Assessment 23, 425–435.
Brokenshire, D., Kumar, V., 2009. Discovering causal models of self-regulated learning. In: Artificial Intelligence in Education. IOS Press, pp. 257–264.
Bruso, J.L., Stefaniak, J.E., 2016. The use of self-regulated learning measure questionnaires as a predictor of academic success. TechTrends 60, 577–584.
Cárdenas-Robledo, L.A., Peña-Ayala, A., 2019. A holistic self-regulated learning model: a proposal and application in ubiquitous-learning. Expert Syst. Appl. 123, 299–314.
Chen, J., 2022. The effectiveness of self-regulated learning (srl) interventions on l2 learning achievement, strategy employment and self-efficacy: a meta-analytic study. Front. Psychol. 13, https://doi.org/10.3389/fpsyg.2022.1021101. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.1021101.
Chen, Q., Yan, Z., 2016. Does multitasking with mobile phones affect learning? A review. Comput. Hum. Behav. 54, 34–42.
Clough, G., 2009. Geolearners: location-based informal learning with mobile and social technologies. IEEE Trans. Learn. Technol. 3, 33–44.
Dai, Y., Wu, Z., 2023. Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: a mixed-methods study. Comput. Assist. Lang. Learn. 36, 861–884.
Davis, D., Chen, G., Jivet, I., Hauff, C., Houben, G.-J., 2016. Encouraging metacognition & self-regulation in moocs through increased learner feedback. In: LAL@ LAK, pp. 17–22.
Delen, E., Liew, J., 2016. The use of interactive environments to promote self-regulation in online learning: a literature review. Eur. J. Contemp. Educ. 15, 24–33.
Du, J., Hew, K.F., Liu, L., 2023. What can online traces tell us about students' self-regulated learning? A systematic review of online trace data analysis. Comput. Educ. 104828.
Engelmann, K., Bannert, M., 2021. Analyzing temporal data for understanding the learning process induced by metacognitive prompts. Learn. Instr. 72, 101205.
Farhan, A.A., Yue, C., Morillo, R., Ware, S., Lu, J., Bi, J., Kamath, J., Russell, A., Bamis, A., Wang, B., 2016. Behavior VS. Introspection: refining prediction of clinical depression via smartphone sensing data. In: 2016 IEEE Wireless Health (WH). IEEE, pp. 1–8.
Feng, Y., Zhao, T., Li, J., Zu, Y., Tavares, A., Gomes, T., Xu, H., 2023. Social behavioral sensing: an exploratory study to assess learning motivation and perceived relatedness of university students using mobile sensing. In: Proceedings of the Annual Meeting of the Cognitive Science Society, vol. 45.
Gong, J., Huang, Y., Chow, P.I., Fua, K., Gerber, M.S., Teachman, B.A., Barnes, L.E., 2019. Understanding behavioral dynamics of social anxiety among college students through smartphone sensors. Inf. Fusion 49, 57–68.
Hachey, A.C., Wladis, C., Conway, K., 2015. Prior online course experience and gpa as predictors of subsequent online STEM course outcomes. The Internet and Higher Education 25, 11–17.
van Halema, N., Van Klaveren, C., Drachsler, H., Schmitz, M., Cornelisz, I., 2020. Tracking patterns in self-regulated learning using students' self-reports and online trace data. Frontline Learn. Res. 8, 140–163.
Harari, G.M., Gosling, S.D., Wang, R., Chen, F., Chen, Z., Campbell, A.T., 2017. Patterns of behavior change in students over an academic term: a preliminary study of activity and sociability behaviors using smartphone sensing methods. Comput. Hum. Behav. 67, 129–138.
Harari, G.M., Lane, N.D., Wang, R., Crosier, B.S., Campbell, A.T., Gosling, S.D., 2016. Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. Perspect. Psychol. Sci. 11, 838–854.
Hardin, J.W., Hilbe, J.M., 2002. Generalized Estimating Equations. Chapman and Hall/CRC.
Hawi, N.S., Samaha, M., 2016. To excel or not to excel: strong evidence on the adverse effect of smartphone addiction on academic performance. Comput. Educ. 98, 81–89.
He, Q., Agu, E.O., 2017. A rhythm analysis-based model to predict sedentary behaviors. In: 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). IEEE, pp. 383–391.
Holmqvist, K., Örbom, S.L., Hooge, I.T.C., Niehorster, D.C., Alexander, R.G., Andersson, R., Benjamins, J.S., Blignaut, P., Brouwer, A.-M., Chuang, L.L., et al., 2024. Retraction note: eye tracking: empirical foundations for a minimal reporting guideline.
Huang, Y., Xiong, H., Leach, K., Zhang, Y., Chow, P., Fua, K., Teachman, B.A., Barnes, L.E., 2016. Assessing social anxiety using GPS trajectories and point-of-interest data. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 898–903.
Ifenthaler, D., 2012. Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. J. Educ. Technol. Soc. 15, 38–52.
Järvelä, S., Bannert, M., 2019. Temporal and adaptive processes of regulated learning: what can multimodal data tell?
Kammoun, N., Meegahapola, L., Gatica-Perez, D., 2023. Understanding the social context of eating with multimodal smartphone sensing: the role of country diversity. In: Proceedings of the 25th International Conference on Multimodal Interaction, pp. 604–612.
Li, Q., Baker, R., Warschauer, M., 2020a. Using clickstream data to measure, understand, and support self-regulated learning in online courses. Int. High. Educ. 45, 100727.
Li, S., Du, H., Xing, W., Zheng, J., Chen, G., Xie, C., 2020b. Examining temporal dynamics of self-regulated learning behaviors in STEM learning: a network approach. Comput. Educ. 158, 103987.
Lin, Y., Liu, Y., Fan, W., Tuunainen, V.K., Deng, S., 2021. Revisiting the relationship between smartphone use and academic performance: a large-scale study. Comput. Hum. Behav. 122, 106835.
Loeffler, S.N., Bohner, A., Stumpp, J., Limberger, M.F., Gidion, G., 2019. Investigating and fostering self-regulated learning in higher education using interactive ambulatory assessment. Learn. Individ. Differ. 71, 43–57.
Loeffler, S.N., Myrtek, M., Peper, M., 2013. Mood-congruent memory in daily life: evidence from interactive ambulatory monitoring. Biol. Psychol. 93, 308–315.
Malmberg, J., Fincham, O., Pijeira-Díaz, H.J., Järvelä, S., Gašević, D., 2021. Revealing the hidden structure of physiological states during metacognitive monitoring in collaborative learning. J. Comput. Assist. Learn. 37, 861–874.
Martin, H., Craigwell, R., Ramjarrie, K., 2022. Grit, motivational belief, self-regulated learning (srl), and academic achievement of civil engineering students. Eur. J. Eng. Educ. 47, 535–557.
May, K.E., Elder, A.D., 2018. Efficient, helpful, or distracting? A literature review of media multitasking in relation to academic performance. Int. J. Educ. Technol. High. Educ. 15, 1–17.
McCardle, L., Hadwin, A.F., 2015. Using multiple, contextualized data sources to measure learners' perceptions of their self-regulated learning. Metacognition and Learning 10, 43–75.
Mou, T.-Y., 2023. Online learning in the time of the Covid-19 crisis: implications for the self-regulated learning of university design students. Act. Learn. High. Educ. 24, 185–205.
Nepal, S., Liu, W., Pillai, A., Wang, W., Vojdanovski, V., Huckins, J.F., Rogers, C., Meyer, M.L., Campbell, A.T., 2024. Capturing the college experience: a four-year mobile sensing study of mental health, resilience and behavior of college students during the pandemic. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 8, 1–37.
Nikou, S.A., Economides, A.A., 2017. Mobile-based assessment: investigating the factors that influence behavioral intention to use. Comput. Educ. 109, 56–73.
Nussbaumer, A., Kravcik, M., Renzel, D., Klamma, R., Berthold, M., Albert, D., A framework for facilitating self-regulation in responsive open learning environments, arXiv preprint arXiv:1407.5891, 2014.
Paris, S.G., Paris, A.H., 2003. Classroom applications of research on self-regulated learning. In: Educational Psychology. Routledge, pp. 89–101.
Pintrich, P.R., 2004. A conceptual framework for assessing motivation and self-regulated learning in college students. Educ. Psychol. Rev. 16, 385–407.
Pintrich, P.R., Wolters, C.A., Baxter, G.P., 2000. 2. Assessing metacognition and self-regulated learning.
Poliak, A., Naradowsky, J., Haldar, A., Rudinger, R., Van Durme, B., Hypothesis only baselines in natural language inference, arXiv preprint arXiv:1805.01042, 2018.
Roth, A., Ogrin, S., Schmitz, B., 2016. Assessing self-regulated learning in higher education: a systematic literature review of self-report instruments. Educ. Assess. Eval. Account. 28, 225–250.
Saeb, S., Zhang, M., Karr, C.J., Schueller, S.M., Corden, M.E., Kording, K.P., Mohr, D.C., et al., 2015. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J. Med. Internet Res. 17, e4273.
Saint, J., Fan, Y., Gašević, D., Pardo, A., 2022. Temporally-focused analytics of self-regulated learning: a systematic review of literature. Comput. Educ.: Artif. Intell. 3, 100060.
Saint, J., Whitelock-Wainwright, A., Gašević, D., Pardo, A., 2020. Trace-srl: a framework for analysis of microlevel processes of self-regulated learning from trace data. IEEE Trans. Learn. Technol. 13, 861–877.
Samaha, M., Hawi, N.S., 2016. Relationships among smartphone addiction, stress, academic performance, and satisfaction with life. Comput. Hum. Behav. 57, 321–325.
Schmitz, B., Wiese, B.S., 2006. New perspectives for the evaluation of training sessions in self-regulated learning: time-series analyses of diary data. Contemp. Educ. Psychol. 31, 64–96.
Schölkopf, B., Smola, A., Müller, K.-R., 1997. Kernel principal component analysis. In: International Conference on Artificial Neural Networks. Springer, pp. 583–588.
Schunk, D.H., 2013. Social cognitive theory and self-regulated learning. In: Self-Regulated Learning and Academic Achievement. Routledge, pp. 119–144.
Shiffman, S., Stone, A.A., Hufford, M.R., 2008. Ecological momentary assessment. Annu. Rev. Clin. Psychol. 4, 1–32.
Shih, K.-P., Chen, H.-C., Chang, C.-Y., Kao, T.-C., 2010. The development and implementation of scaffolding-based self-regulated learning system for e/m-learning. J. Educ. Technol. Soc. 13, 80–93.
Taub, M., Azevedo, R., 2019. How does prior knowledge influence eye fixations and sequences of cognitive and metacognitive srl processes during learning with an intelligent tutoring system? Int. J. Artif. Intell. Educ. 29, 1–28.
Torrano Montalvo, F., González Torres, M., et al., 2004. Self-regulated learning: current and future directions.
Ulfatun, T., Septiyanti, F., Lesmana, A.G., 2021. University students' online learning self-efficacy and self-regulated learning during the Covid-19 pandemic. Int. J. Inf. Educ. Technol. 11, 597–602.
Veenman, M.V.J., 2011. Alternative assessment of strategy use with self-report instruments: a discussion. Metacognition and learning 6, 205–211.
Viberg, O., Khalil, M., Baars, M., 2020. Self-regulated learning and learning analytics in online learning environments: a review of empirical research. In: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, pp. 524–533.

Wang, R., Aung, M.S.H., Abdullah, S., Brian, R., Campbell, A.T., Choudhury, T., Hauser, M., Kane, J., Merrill, M., Scherer, E.A., et al., 2016. Crosscheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 886–897.

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., Campbell, A.T., 2014. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 3–14.

Wang, R., Wang, W., DaSilva, A., Huckins, J.F., Kelley, W.M., Heatherton, T.F., Campbell, A.T., 2018a. Tracking depression dynamics in college students using mobile phone and wearable sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2, 1–26.

Wang, W., Harari, G.M., Wang, R., Müller, S.R., Mirjafari, S., Masaba, K., Campbell, A.T., 2018b. Sensing behavioral change over time: using within-person variability features from mobile sensing to predict personality traits. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2, 1–21.

Wang, W., Mirjafari, S., Harari, G., Ben-Zeev, D., Brian, R., Choudhury, T., Hauser, M., Kane, J., Masaba, K., Nepal, S., et al., 2020. Social sensing: assessing social functioning of patients living with schizophrenia using mobile phone sensing. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–15.

Wang, W., Nepal, S., Huckins, J.F., Hernandez, L., Vojdanovski, V., Mack, D., Plomp, J., Pillai, A., Obuchi, M., Dasilva, A., et al., 2022. First-gen lens: assessing mental health of first-generation students across their first year at college using mobile sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 6, 1–32.

Winne, P.H., Perry, N.E., 2000. Measuring self-regulated learning. In: Handbook of Self-Regulation. Elsevier, pp. 531–566.

Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G.-J., Paas, F., 2019. Supporting self-regulated learning in online learning environments and moocs: a systematic review. Int. J. Hum.-Comput. Interact. 35, 356–373.

Xu, X., Chikersal, P., Doryab, A., Villalba, D.K., Dutcher, J.M., Tumminia, M.J., Althoff, T., Cohen, S., Creswell, K.G., Creswell, J.D., et al., 2019. Leveraging routine behavior and contextually-filtered features for depression detection among college students. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 3, 1–33.

Yang, X., Li, J., Guo, X., Li, X., 2015. Group interactive network and behavioral patterns in online english-to-chinese cooperative translation activity. The Internet and Higher Education 25, 28–36.

Zhang, X., Li, W., Chen, X., Lu, S., 2018. Moodexplorer: towards compound emotion detection via smartphone sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1, 1–30.

Zhou, M., Ma, M., Zhang, Y., SuiA, K., Pei, D., Moscibroda, T., 2016. Edum: classroom education measurements via large-scale wifi networks. In: Proceedings of the 2016 Acm International Joint Conference on Pervasive and Ubiquitous Computing, pp. 316–327.

Zimmerman, B.J., 1986. Becoming a self-regulated learner: which are the key subprocesses? Contemp. Educ. Psychol. 11, 307–313.

Zimmerman, B.J., 2013. Theories of self-regulated learning and academic achievement: an overview and analysis. Self-regulated learning and academic achievement 1–36.

Zimmerman, B.J., Schunk, D.H., 1994. Self-Regulation of Learning and Performance: Issues and Educational Applications. L. Erlbaum Associates.

Zimmerman, B.J., Schunk, D.H., 2011. Self-regulated learning and performance: an introduction and an overview. Handbook of Self-Regulation of Learning and Performance 15–26.