

# An Explorative Investigation into Leveraging LLMs to Predict University Students' Learning Motivation

Yu Feng<sup>1,2</sup> (yufeng3904@gmail.com), Tongyu Zhao<sup>1,2</sup> (id9269@alunos.uminho.pt),  
Yatong Zu<sup>3</sup> (zuyatong@jlu.edu.cn), Adriano Tavares<sup>4</sup> (atavares@dei.uminho.pt),  
Tiago Gomes<sup>5</sup> (mr.gomes@dei.uminho.pt), Hao Xu<sup>2,\*</sup> (xuhao@jlu.edu.cn)

<sup>1</sup>School of Engineering, University of Minho, Guimarães, Portugal

<sup>2</sup>College of Computer Science and Technology, Jilin University, Changchun, China

<sup>3</sup>College of Physical Education, Jilin University, Changchun, China

<sup>4</sup>Department of Industrial Electronics, University of Minho, Braga, Portugal

<sup>5</sup>ALGORITMI, University of Minho, Guimarães, Portugal

\*Corresponding Author

## Abstract

Learning motivation is a key variable in learning. Therefore, its assessment has consistently been a popular research topic. While traditional methods like self-report still dominate, methods integrating passive mobile sensing have emerged, using smartphones to collect behavioral data and assess learning motivation via statistical and machine learning techniques. Recent advances in large language models (LLMs) offer new perspectives for psychological measurements, yet their application in learning motivation assessment remains underexplored. To bridge this gap, we propose a novel approach that integrates LLMs with passive mobile sensing to assess and predict students' learning motivation. We constructed our dataset using mobile sensing data and self-report measures, then designed zero-shot and few-shot tasks embedded in LLMs to evaluate the performance. Ultimately, our findings indicate the feasibility and highlight the potential of leveraging LLMs to predict learning motivation levels based on mobile sensing data.

**Keywords:** learning motivation; large language models; passive sensing; education; psychology

## Introduction

Learning motivation, as a critical variable that profoundly influences students' learning experiences, academic engagement, and performance, has long been a focal point of research in education and psychology. According to Ryan and Deci's Self-Determination Theory (SDT) learning motivation encompasses both internal and external drivers that prompt students to pursue learning goals or fulfill personal expectations (Ryan & Deci, 2017). SDT differentiates motivation into intrinsic motivation, extrinsic motivation, and amotivation (Mitchell et al., 2012; Ryan & Deci, 2000). While intrinsic motivation reflects a natural propensity for learning and is the core element (Howard et al., 2017; Tohidi & Jabbari, 2012), our study concentrates on it as a key indicator of learning motivation within the SDT framework. As a fundamental trigger in the learning process, learning motivation has been extensively studied across multiple disciplines, including education, psychology, behavioral science, cognitive science, and neuroscience. A wide range of assessment approaches has been employed, spanning from traditional psychological methods such as self-report surveys and expert observations (Pekrun, 2020;

Fulmer & Frijters, 2009) to hybrid models integrating mobile sensing with machine learning (Feng et al., 2024), as well as neurophysiological approaches involving EEG-based assessment (Ng, 2018). However, conventional methods, such as self-reports and expert observations, which rely heavily on human interpretation and judgment, still dominate the field. In recent years, the emergence of LLMs has introduced new perspectives across various research domains, particularly in psychometrics, such as affective computing, affective state sensing, and human activity recognition. Given that LLMs are human language models, they offer a unique advantage over traditional machine learning and deep learning models by incorporating extensive pre-trained human linguistic knowledge, potentially enabling them to perceive and predict human behavior and psychological states. While some studies have explored the application of LLMs in the field of psychometrics (Peters & Matz, 2024), their use in learning motivation research remains largely unexplored.

Therefore, our study aims to propose an innovative hybrid approach that integrates traditional and emerging methodologies by leveraging LLMs in combination with passive mobile sensing to assess and predict learning motivation levels. In this study, we employ passive smartphone sensing to capture students' daily behavioral data and simultaneously collect their learning motivation level measures through self-report survey. LLMs are then used to explore the latent relationships between behavioral patterns and learning motivation levels. To achieve this research objective, we propose the following research question:

- Can we predict students' learning motivation levels by leveraging LLMs integrated with passive mobile sensing data collected from their smartphones?

Driven by the research question, we conducted a study to investigate whether LLMs could be utilized to predict students' learning motivation levels using the data collected via passive mobile sensing. In this study, we first constructed a dataset that includes tracked learning motivation data obtained through self-report surveys and concurrently collected daily behavioral data via passive smartphone sensing from undergraduate students at a university in China.

Next, we processed the collected dataset and incorporated prompt engineering to design zero-shot and few-shot tasks, embedding these tasks into LLMs to explore whether learning motivation levels could be predicted based on behavioral data collected from smartphones. Our findings suggest that a latent relationship exists between behavioral data collected via mobile sensing and students' learning motivation levels, demonstrating the feasibility and highlighting the potential of using LLMs combined with passive mobile sensing for learning motivation prediction. The key contributions of our study are as follows:

- We propose an innovative integrated approach by leveraging LLMs in conjunction with passive mobile sensing to predict students' learning motivation levels. Our study supports the feasibility of employing this novel method, thereby expanding the methodological framework for learning motivation assessment. Moreover, our approach offers a new methodological paradigm for future research in this domain.

This study was approved by the Institutional Review Board of the conducting institution. Participants were fully informed of all potential risks related to passive smartphone sensing and provided explicit consent. The consent form detailed these risks and was thoroughly explained during recruitment. To ensure privacy, all collected data was anonymized, securely stored on our private server, and used exclusively for this study.

The paper is organized as follows. Section 2 provides an overview of existing methodologies for assessing learning motivation levels. In Section 3, we elaborate on the study methodology, including study design, participant recruitment, data collection, data processing, tasks, and performance evaluation metrics. Section 4 presents the results, followed by a discussion of key findings and their implications in Section 5. Section 6 highlights the study's limitations and outlines potential future research directions. Finally, Section 7 concludes the paper.

## Related Work

Assessing learning motivation has long been a prominent research topic in pedagogy, educational psychology, and related psychological fields (Howard et al., 2020; Fredricks & McColskey, 2012). A substantial amount of research has focused on self-report survey type of methods (Pekrun, 2020; Fulmer & Frijters, 2009), including the development of various scales (Chiang, 2024; Sousa et al., 2024;), experience sampling methods or time diary methods (Martin et al., 2020; Ketonen et al., 2018). Besides these questionnaire-based methods, some studies have adopted a phenomenological perspective by employing in-depth interviews (Akteruzzaman et al., 2023) or expert/teacher scorings (Nair et al., 2024) to assess learning motivation. However, these approaches remain grounded within a relatively traditional educational and psychological framework. Though they are still dominant in measurement practices, their heavy reliance on human subjectivity and the inherent time-consuming nature of these methods presents unavoidable limitations, and

these methods remain insufficient for achieving unobtrusive and continuous measurement. With the advancement of HCI technology, the limitations of subjective methods have led to the exploration of multi-sensor techniques that capture more objective and granular data. Neuropsychological methods, including electroencephalographs (EEG) (Ng, 2018) and functional magnetic resonance imaging (fMRI) (Reeve & Lee, 2019), have also been explored (Di Domenic & Ryan, 2017). However, ensuring consistent reliability and reproducibility across different learning settings remains difficult, as these methods often depend on specialized tools and controlled conditions. Similarly, few studies have explored learning motivation at a more macro-level by integrating sensor-based approaches focusing on the behavioral level, such as digital phenotyping or mobile/passive sensing methods (Feng et al., 2024), since such blended methods are feasibly deployed to capture behavioral and motivation relevant physiological variables in similar educational settings (Ciordas-Hertel et al., 2021; Holstein et al., 2019; Wang et al., 2014). These studies integrated mobile sensing with traditional assessment methods and employed machine learning techniques to analyze the correlations between students' daily behaviors and learning motivation level, offering an illuminating methodological perspective. However, these methods have yet to achieve continuous assessment of learning motivation. Recently, the emergence of large language models has captured researchers' attention, leading them to conduct research by leveraging LLMs in psychological measurements (Zhang et al., 2024; Zhou et al., 2024). However, research on integrating LLMs with sensor data to assess learning motivation levels remains in its early stages. This gap underscores the need for further research that combines state-of-the-art sensing techniques with insights derived from LLMs to provide novel methods of assessing learning motivation in real-world educational environments.

## Methodology

In this section, we elaborate on the overall design of our study, then discuss participant recruitment and data collection, followed by data processing, including preprocessing and the behavioral features. We then describe the tasks assigned to LLMs, detailing the prompt design for zero-shot and few-shot tasks, as well as chain-of-thought reasoning. Finally, we explain the metrics employed to analyze the results.

## Study Design

To investigate whether large language models (LLMs) can leverage smartphone-based sensing data to measure and predict undergraduates' learning motivation, we conducted a continuous passive mobile sensing study tracking a group of 28 undergraduates at a college in China over 14 consecutive weeks, combining passive sensing with self-report surveys. We developed a native mobile application to collect daily behavioral data from participants and to assess their learning

motivation via weekly questionnaires. Beginning in the third week of February 2024, we spent two weeks recruiting participants on campus and guiding them through the installation of the mobile application and pretesting for data collection. When data collection officially commenced, we concurrently gathered both passive sensing data and participants' self-report responses, all participants were required to keep the sensing application installed and active on their smartphones throughout the consecutive 14-week period. After data collection concluded, we screened the datasets from the original 28 participants for quality, ultimately including 17 students in the final analysis, and constructed a new dataset from these 17 participants. Building upon this refined dataset, we adopted a prompt engineering approach by designing various tasks and presenting them as prompts to a large language model, thereby investigating how it assesses and predicts participants' learning motivation levels. Each of the 28 participants received a weekly compensation of 40 RMB during the study. The study was approved by the university's institutional review board.

### **Participants and Data Collection**

**Participants** We recruited 31 participants for our study at the very beginning, and 28 of them met all the requirements by consistently completing weekly self-report surveys on learning motivation and providing smartphone sensing data. However, upon reviewing the dataset, we excluded 11 participants due to extensive missing data in the smartphone sensing records. Consequently, we retained 17 participants to form the dataset used for the present research. All participants were Chinese undergraduates, including 11 males and 6 females. Concerning their majors, 29.4% (N=5) studied Computer Science, 23.5% (N=4) studied Clinical Medicine, 17.6% (N=3) were in Medical Imaging, 11.8% (N=2) majored in Economics, another 11.8% (N=2) in Advertising, and 5.9% (N=1) in Philosophy.

**Passive Sensing Data: Mobile Application** We developed a native Android application and utilized it to passively collect sensor data from participants while minimizing intrusiveness to capture authentic information from their daily lives. The collected sensor data is initially stored locally on each participant's smartphone; once the device connects to a network, the data is uploaded to our private server and subsequently deleted from the phone. Specifically, the application collects information about physical activities and states (e.g., walking, running, or remaining still), usage times of different apps (with only the names of currently running apps being uploaded), call and SMS counts, environmental information (ambient sound and light levels), battery status and screen events (unlock/lock actions), and location data. After collection, we perform preprocessing and data filtering on the collected raw data before extracting behavioral features from the refined dataset.

**Self-report Data: Learning Motivation Scales** We employed the subscales of the Academic Self-Regulation Questionnaire (SRQ-A) to assess participants' learning motivation, with evidence for its validity provided by Ryan and Connell (1989). The SRQ-A includes four distinct subscales: external regulation, introjected regulation, identified regulation, and intrinsic motivation; each is capable of generating its own score for research analysis. In this study, we adopted the intrinsic motivation subscale as our measure of learning motivation. Every week, at a randomly selected time, we distributed the SRQ-A via the mobile application, requiring participants to complete the questionnaire thereby providing self-reported measures of their learning motivation levels.

### **Data Processing**

**Pre-process** We began by cleaning and filtering the raw data collected from 31 participants, which included both self-reported responses and passive mobile sensing data, prior to engaging in data engineering processes. Specifically, we excluded data from participants with substantial missing information. Additionally, for participants with low data continuity, particularly in the passive mobile sensing data, we removed participants whose data quality did not meet the stringent requirements of this study. Consequently, we retained data from 17 participants, comprising continuously collected passive mobile sensing data and psychological measurement data over a 14-week period, to form the dataset utilized in this research.

**Behavioral Features** We identify and incorporate features that capture the following dimensions of daily behaviors derived from passive smartphone sensing data. We collected behavioral features across the following aspects and computed them on a daily basis. In terms of physical activity and status, we utilized an activity recognition API to identify participants' current activities and determine their physical states, as well as the duration of various activities, such as walking. Regarding semantic location, since all participants lived and studied on campus, we collected raw GPS coordinates through the mobile application and assigned semantic labels to these locations by integrating a third-party API with the university's campus map, which provided detailed labels for different areas and buildings (e.g., dormitory, library). For sleep patterns, sleep data were extracted from participants' smartphones to infer sleep duration, as well as sleep start and end times, with an approximate error margin of 40 minutes (Wang et al., 2014; Chen et al., 2013). In terms of the ambient environment, we collected environmental data from smartphones, including ambient sound levels and light levels, to assess participants' surroundings. Lastly, with regard to smartphone usage, we captured various aspects of smartphone activity, including screen events (e.g., unlock duration and frequency), battery status (e.g., charging frequency), communication patterns (e.g., the number and duration of incoming and outgoing calls), and app usage details. App usage data were collected

every five minutes, recording only the names of currently active applications without storing content, and 541 mobile applications were categorized into 43 types (e.g., education, entertainment, music), with usage durations calculated for each category. Acknowledging the limitations of LLMs in handling raw numerical data (Spathis & Kawsar, 2024) and their primary role as language models, as well as drawing on methods that utilize smartphones to measure other psychological variables, we opted to represent participants' daily behavioral features in clear and succinct English. Each feature was sequentially detailed, accompanied by the corresponding date and time range in the format YYYY-MM-DD HH:MM:SS to YYYY-MM-DD HH:MM:SS (e.g., 2024-03-01 00:00:00 to 2024-03-01 23:59:59). To maintain the temporal context, we preserved the time series structure within the data to determine whether the LLMs are capable of performing the tasks rather than developing a robust predictive model.

### Tasks

To investigate whether LLMs can understand the relationship between behavioral data passively collected via smartphones and students' learning motivation levels, tasks were designed and constructed to prompt LLMs to predict each participant's learning motivation level (as captured by intrinsic learning motivation scores) in a single inference. Specifically, we designed zero-shot and few-shot tasks using chain-of-thought (CoT) prompting. The study was conducted on Gemini 1.5 Pro, and we used approximately 5,000 tokens to describe participants' weekly behavioral data collected from smartphones. The LLM was required to output both a predicted score on the learning motivation level using the Likert scale and reasoning for the score, with reasoning based on provided student's weekly behaviors. To ensure the consistency and stability of the output as much as possible, we set the temperature to 0. In this investigative study, we designed and assigned individual tasks for each of the 17 participants. The continuous 14-week dataset for each participant was split into 10 weeks for training and 4 weeks for testing. For each participant, we performed three rounds of randomized training-testing splits. In the few-shot prompting tasks, we incrementally added examples from the training dataset, starting from one shot and increasing up to ten, to evaluate the model's ability to generalize across individual instances within the dataset.

**Zero-shot with Chain-of-Thought Reasoning** We used zero-shot as the baseline for our study to facilitate comparison, and we constructed the prompt as illustrated in Figure 1.

Below is a description of a university student's activities over a week, passively collected from their smartphone sensors. Based on the descriptions provided, select the option that best represents the student's learning motivation level for the provided week's description below.

For the learning motivation level, choose a Likert score ranging from 1 to 5 that best represents how the student generally felt during the week, where:

- 1 represents very low motivation,
- 5 represents very high motivation.

Description of the student's activities for the future week: *{feature description}*

Provide your choices in the following form and include reasoning:

- Learning Motivation Level: *[predicted score and reasoning]*

Figure 1: Prompt for Zero-Shot Tasks with Chain-of-Thought (CoT) Reasoning

**Few-shot with Chain-of-Thought Reasoning** We constructed the prompts for the few-shot tasks by randomly selecting one or more weeks of data from the training set as labeled examples, as illustrated in Figure 2.

Given a series of descriptions detailing a university student's weekly activities passively collected from their smartphones, along with their corresponding learning motivation levels, your task is to identify relationships between the student's daily activities and learning motivation levels. Based on these relationships, predict the student's learning motivation level during future activities.

According to the following behaviors of the student during a week, their learning motivation level is *score : weekly description*

\*<number-of-shot> weeks

Please select the option that best represents the student's learning motivation level for the future week's description based on the relationships you learned from the provided data.

For the student's learning motivation level, choose a Likert score ranging from 1 to 5 that best represents how the student generally felt during the week, where 1 represents very low motivation and 5 represents very high motivation

Description of the student's activities for the future week: *{feature description}*

Please provide your choices in the following format and include reasoning:

Learning Motivation Level: *[predicted score and reasoning]*

Figure 2: Prompt for Few-Shot Tasks with Chain-of-Thought (CoT) Reasoning

### Metrics for Performance Analysis

We adopted a macro-level approach using the Mean Absolute Error (MAE) as the metrics for evaluating the accuracy of our predictions. The overall Mean Absolute Error (MAE) is obtained using the formula:

$$MAE = \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{n_i} \sum_{j=1}^{n_i} |P_{ij} - T_{ij}| \right)$$

Here,  $P_{ij}$  denotes the predicted value,  $T_{ij}$  represents the actual value for the  $j$ -th observation of the  $i$ -th participant,  $n_i$  indicates the number of observations for participant  $i$ , and  $N$  is the total number of participants.

## Results

Under the zero-shot task condition, the MAE value indicates that the model's predictive capability is relatively poor, with an MAE of 1.45. Possibly due to the inherent limitations of the LLM's internal knowledge, these results suggest that its ability to predict individual learning motivation levels under zero-shot conditions remains limited. Compared to the zero-shot task condition, LLMs demonstrated improved predictive performance under the few-shot task condition. As shown in the table 1, a clear trend of decreasing MAE can be observed as the number of shots provided to the LLMs increases. Among these, the one-shot approach showed the most significant improvement, with a 26.2% reduction in MAE. However, as the number of shots continued to increase, the rate of performance improvement began to slow down. Nevertheless, compared to the zero-shot task, the few-shot task condition exhibited enhanced predictive capability, and the predictive performance of LLMs improved significantly as more examples were provided. Therefore, using behavioral data collected through smartphone sensing to predict students' learning motivation levels via LLMs is feasible, offering a promising approach to exploring the relationship between the two.

Table 1: The averaged Mean Absolute Error across the 17 participants for learning motivation level. The bold results indicate the best-performing shot.

Shot	MAEs
0	1.45
1	1.07
2	1.41
3	1.38
4	1.13
5	1.11
6	0.93
7	0.92
8	0.92
9	0.85
10	<b>0.83</b>

To explore the relationship between students' daily behavioral activities and their learning motivation levels using LLMs, we provided the model with an overall description of the task and employed chain-of-thought prompting to elicit its reasoning. Specifically, as shown in Table 2, certain behavioral patterns exhibit associations with different levels of learning motivation. For example, a longer duration of using productivity-related apps in a relatively quiet environment, as well as frequent use of educational apps, correlates with relatively higher learning motivation levels. Conversely, frequent phone unlocking and excessive use of short-video apps are associated with lower learning motivation levels. Meanwhile, we selected two participants,

one with a relatively high learning motivation level and one with a relatively low level, and presented a comparison of the reasoning generated by the LLMs under zero-shot and ten-shot conditions for each participant.

Table 2: Zero-Shot Chain-of-Thought reasoning and Ten-Shot Chain-of-Thought reasoning for participants.

Participants	Predicted Score	Reasoning
Participant No.5 (Zero-shot)	3	Long duration of using productivity-related applications (average 7.3 hours weekly), in a relatively quiet environment (low ambient sound level), frequent use of educational applications
Participant No.5 (Ten-shot)	4	This student appears to be highly focused and academically oriented. They dedicate a significant amount of time to productivity apps in a quiet environment, which likely supports sustained concentration. Additionally, their frequent use of educational apps suggests a strong commitment to learning and self-improvement.
Participant No.13 (Zero-shot)	2	A relatively high number of unlocking episodes, prolonged use of short-video applications (average 16.5 hours weekly), and a moderate number of text messages sent and received.
Participant No.13 (Ten-shot)	1	This student appears to be highly engaged with their smartphone, frequently unlocking it, which may indicate a tendency toward constant checking or digital distraction. The prolonged use of short-video applications suggests a strong inclination toward consuming bite-sized, entertaining content. Meanwhile, the moderate volume of text messages sent and received implies that, although they maintain some level of communication, their primary focus seems to be on media consumption rather than social interaction.

The results indicate that as more example data is provided, the LLM's reasoning ability is enhanced, demonstrating its capacity to learn and adjust optimally. While the accuracy and validity of the analytical results remain subjects for further investigation, this approach still holds research value. Features collected via smartphone sensors serve as potential indicators associated with learning motivation levels. Compared to traditional correlation analysis and manual interpretation, LLMs can mitigate the limitations and biases of human-driven analysis while enhancing sensitivity to

overall information. Leveraging their prior knowledge, LLMs can conduct a more comprehensive analysis of these relationships and generate logical reasoning aligned with expectations, offering an analytical perspective at the detailed level and revealing the latent relationships between smartphone sensor features and learning motivation levels, thereby providing new insights and inspiration for further research.

## Discussion

Our study explored the use of LLMs to predict students' learning motivation levels based on passive mobile sensing data collected from their daily behavioral patterns. Compared to traditional sensor-based assessment approaches, conventional statistical methods and expert-driven correlation analyses used in previous research, our findings provide preliminary evidence supporting the feasibility of employing LLMs for assessing learning motivation. This contributes to the broader field of psychological measurement by expanding existing methodologies for evaluating learning motivation. Furthermore, as an exploratory study, our approach introduces a methodological paradigm that offers insights and potential directions for future research.

## Implications

In the domain of learning motivation assessment, our approach enhances traditional psychological measurement approaches while simultaneously optimizing sensor-based hybrid methods. Beyond integrating passive mobile sensing, we leveraged LLMs to analyze the relationships between collected sensing data and psychological measures. Compared to conventional statistical methodologies and the machine learning or deep learning techniques commonly employed in prior research, our study offers a novel perspective on assessing learning motivation. Moreover, regarding data analysis, our method mitigates certain biases inherent in traditional techniques, such as manual interpretation or expert-driven analyses. More importantly, our findings demonstrate the feasibility of using LLMs for predicting and assessing learning motivation levels. By introducing a new integrative approach, our study proposes a transferable methodological paradigm that can be applied to the measurement of learning motivation and its associated psychological variables in future research.

In the field of HCI research, our study offers valuable insights. Our findings reveal that LLMs analyze students' daily behavioral data by first extracting features from smartphone sensor data and then performing reasoning and inference based on these patterns. This characteristic presents a novel research direction for future HCI studies, suggesting the potential of using LLMs to identify sensor-derived features that exhibit deep latent associations with learning motivation levels. These identified features could then serve as the foundation for sensor-driven measurement and intervention studies. Furthermore, given that LLMs are inherently language models, they offer unique advantages in

human interaction. For example, in designing intelligent agents for learning motivation interventions, our study provides a preliminary foundation and insights for future research.

## Limitations and Future Work

Though our study provides preliminary evidence supporting the feasibility of using LLMs to predict learning motivation levels, several limitations remain. First, the sample size and demographic representation in our study are constrained, as our dataset is primarily derived from students at a single university. This lack of population diversity limits the generalizability of our findings to broader populations and more generalized or common contexts. Second, the ground truth data for learning motivation levels was collected using the traditional self-report method mostly employed in psychology, which introduces the possibility of subjective bias from participants and may affect data reliability. Additionally, when processing the passive sensing data, due to the nature of LLMs as language models, we transformed log data into verbal descriptions as input for the model, rather than providing a continuous stream of raw sensor data. This limitation restricts the extent to which our study fully realizes the potential of sensor-based measurement for assessing learning motivation.

However, our study offers several constructive insights for future research. Primarily, we employed prompt engineering to design tasks that utilize LLMs for predicting learning motivation levels. Future studies could explore fine-tuning LLMs to integrate a broader range of data, potentially enhancing predictive performance. Additionally, increasing the scale and diversity of data collection could help expand the population representation in the dataset, thereby improving the model's generalizability. Furthermore, our study focused on evaluating the feasibility of using LLMs integrated with passive sensing for learning motivation prediction without conducting a comparative analysis across different motivation-level subgroups. This presents another promising direction for future research.

## Conclusion

Our study proposes a novel integrated approach that leverages LLMs combined with passive mobile sensing. We employed zero-shot and few-shot learning tasks to predict learning motivation using LLMs and utilized chain-of-thought reasoning to explore the potential relationships between daily behaviors and learning motivation levels. Our results indicate that it is feasible to leverage LLMs integrated with passive mobile sensing to predict students' learning motivation based on daily behavioral data from smartphones, and they further underscore the considerable potential of this approach. Moreover, compared to zero-shot tasks, providing LLMs with labeled information enhances the accuracy of their predictions. Overall, our research enriches the assessment methods for learning motivation and provides an inspiring new paradigm for future studies.

## Acknowledgments

This research was supported by the Humanity and Social Science Youth Foundation of Ministry of Education of China (No.20YJC190034).

## References

- Akteruzzaman, M., Sattar, S. B., & Anwar, S. J. (2023). Motivation in L2 learning: An analysis with self-determination theory. *Multidisciplinary Reviews*, 6.
- Ciordas-Hertel, G. P., Rödling, S., Schneider, J., Di Mitri, D., Weidlich, J., & Drachler, H. (2021). Mobile sensing with smart wearables of the physical context of distance learning students to consider its effects on learning. *Sensors*, 21(19), 6649.
- Di Domenico, S. I., & Ryan, R. M. (2017). The emerging neuroscience of intrinsic motivation: A new frontier in self-determination research. *Frontiers in human neuroscience*, 11, 145.
- Feng, Y., Zhao, T., Zhang, Y., Zu, Y., Tavares, A., Gomes, T., & Xu, H. (2024). Innovative Attempt at Enhancing Psychological Assessment: A Preliminary Investigative Study of Measuring College Students' Learning Motivation Levels through the Lens of Passive Sensing via Smartphones. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 46).
- Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In *Handbook of research on student engagement* (pp. 763-782). Boston, MA: Springer US.
- Fulmer, S. M., & Frijters, J. C. (2009). A review of self-report and alternative approaches in the measurement of student motivation. *Educational Psychology Review*, 21(3), 219-246.
- Holstein, K., McLaren, B. M., & Alevin, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher-AI complementarity. Grantee Submission
- Howard, J. L., Gagné, M., & Bureau, J. S. (2017). Testing a continuum structure of self-determined motivation: A meta-analysis. *Psychological bulletin*, 143(12), 1346.
- Howard, J. L., Gagné, M., Van den Broeck, A., Guay, F., Chatzisarantis, N., Ntoumanis, N., & Pelletier, L. G. (2020). A review and empirical comparison of motivation scoring methods: An application to self-determination theory. *Motivation and Emotion*, 44, 534-548.
- Ketonen, E. E., Dietrich, J., Moeller, J., Salmela-Aro, K., & Lonka, K. (2018). The role of daily autonomous and controlled educational goals in students' academic emotion states: An experience sampling method approach. *Learning and Instruction*, 53, 10-20.
- Martin, A. J., Mansour, M., & Malmberg, L. E. (2020). What factors influence students' real-time motivation and engagement? An experience sampling study of high school students using mobile technology. *Educational Psychology*, 40(9), 1113-1135.
- Mitchell, J. I., Gagné, M., Beaudry, A., & Dyer, L. (2012). The role of perceived organizational support, distributive justice and motivation in reactions to new information technology. *Computers in Human Behavior*, 28(2), 729-738.
- Nair, R., Shmon, T., Neufeld, A., Malin, G., & McKague, M. (2024). When medical students are autonomously motivated to mentor: a pilot study on confidence in clinical teaching and psychological well-being. *Canadian Medical Education Journal*, 15(4), 56-62.
- Ng, B. (2018). The neuroscience of growth mindset and intrinsic motivation. *Brain sciences*, 8(2), 20.
- Pekrun, R. (2020). *Self-report is indispensable to assess students' learning*. *Frontline Learning Research*, 8 (3), 185-193.
- Peters, H., & Matz, S. C. (2024). Large language models can infer psychological dispositions of social media users. *PNAS nexus*, 3(6), pga231.
- Reeve, J., & Lee, W. (2019). A neuroscientific perspective on basic psychological needs. *Journal of personality*, 87(1), 102-114.
- Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: examining reasons for acting in two domains. *Journal of personality and social psychology*, 57(5), 749.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1), 54-67.
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.
- Sousa, R. M., Gonçalves, N. G. S., Pereira, V. H., & Norcini, J. (2024). Motivating Medical Students: Adaptation of the Academic Motivation Scale within the Framework of the Self-Determination Theory. *Acta Médica Portuguesa*, 37(11), 757-766.
- Spathis, D., & Kawsar, F. (2024). The first step is the hardest: Pitfalls of representing and tokenizing temporal data for large language models. *Journal of the American Medical Informatics Association*, 31(9), 2151-2158.
- Tohidi, H., & Jabbari, M. M. (2012). The effects of motivation in education. *Procedia-social and behavioral Sciences*, 31, 820-824.
- Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., ... & Campbell, A. T. (2014, September). 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).
- Zhang, T., Teng, S., Jia, H., & D'Alfonso, S. (2024, October). Leveraging llms to predict affective states via smartphone sensor features. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 709-716).
- Zhou, H., Muppalla, S., Yildiz, Z., Kim, M., Sahai, A., & Peng, N. (2024, November). Measuring psychological

depth in language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 17162-17196).