Universidade do Minho
Escola de Engenharia

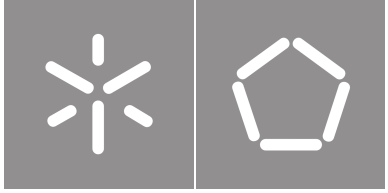
Yu Feng

A Hybrid Psychometric Approach Integrating Self-Determination Theory and Passive Smartphone Sensing for Continuous Assessment of Learning Motivation

Maio de 2026

A Hybrid Psychometric Approach Integrating Self-Determination Theory and Passive Smartphone Sensing for Continuous Assessment of Learning Motivation

UMinho | 2026



Universidade do Minho
Escola de Engenharia

Yu Feng

A Hybrid Psychometric Approach Integrating Self-Determination Theory and Passive Smartphone Sensing for Continuous Assessment of Learning Motivation

Tese de Doutoramento
Programa Doutoral em Engenharia Eletrónica e Computadores

Trabalho efetuado sob a orientação de:
Professor Doutor Tiago Manuel Ribeiro Gomes
Professor Doutor Adriano José da Conceição Tavares
Professor Doutor Hao Xu

Maio de 2026

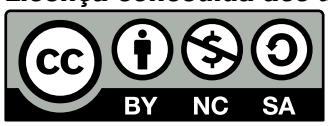
COPYRIGHT AND TERMS OF USE OF THIS WORK BY A THIRD PARTY

This is academic work that can be used by third parties as long as internationally accepted rules and good practices regarding copyright and related rights are respected.

Accordingly, this work may be used under the license provided below.

If the user needs permission to make use of the work under conditions not provided for in the indicated licensing, they should contact the author through the RepositoriUM of Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

CC BY-NC-SA 4.0

`https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en`

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v7.0.2) .

Acknowledgements

My PhD has been a wonderful, adventurous journey. It stretched me, surprised me, and changed how I see the world. I could not have walked it alone. Many people offered help at key moments; the few words below are only a small way to say thank you.

I am deeply grateful to my supervisor team, Professor Doctor Tiago Gomes, Professor Doctor Adriano José da Conceição Tavares, and Professor Doctor Hao Xu. Thank you for your guidance, patience, and trust. Your careful feedback, clear directions, and high standards shaped this work and shaped me as a researcher. You opened doors, challenged my ideas with kindness, and stood by me through setbacks and revisions. I have learned not only methods and results, but also how to think, write, and collaborate.

My heartfelt thanks go to my parents, Li Chang and Guojun Feng. Your love, encouragement, and quiet sacrifices have carried me from the first step to the last page. Thank you for believing in me when I doubted myself, for every message, call, and prayer, and for giving me the courage to finish. I dedicate this dissertation to you.

I also thank my friends, labmates, and colleagues for the conversations that sparked new ideas, the late-night debugging and coffee, the help with data collection, and the honest criticism that made the work stronger. I am grateful to the staff who supported the everyday logistics that made research possible.

Finally, I want to thank the person who kept showing up—myself. Thank you for staying curious, doing the hard work, and moving forward one small step at a time. To everyone mentioned and unmentioned who helped along the way: thank you.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

_____, _____

(Place) (Date)

_____

(Yu Feng)

Resumo

Uma abordagem psicométrica híbrida que integra a Teoria da Autodeterminação e a monitorização passiva via telemóvel para a avaliação contínua da motivação para a aprendizagem

A motivação para a aprendizagem molda profundamente o comportamento e o envolvimento dos estudantes, sendo central na psicologia da educação e na pedagogia. À luz da Teoria da Autodeterminação (SDT), a motivação intrínseca é um conceito estável de impulso genuíno. As avaliações convencionais baseadas em autorrelatos ou observações de peritos são trabalhosas, propensas a distorção e incapazes de captar a dinâmica temporal. Embora a monitorização móvel esteja difundida em psicometria, continuam escassos os trabalhos que visam diretamente a motivação para a aprendizagem e uma abordagem integrada, alinhada com a SDT e com avaliação contínua. Esta dissertação enfrenta essa lacuna.

Propomos uma abordagem psicométrica híbrida que integra autorrelatos fundamentados na SDT com sensorização passiva por telemóveis para uma avaliação contínua, de validade ecológica, e sustentada por estudos longitudinais que emparelham dados comportamentais recolhidos via telemóveis com escalas validadas de motivação. Um fluxo de trabalho operacional de cinco camadas abrange perceção, pré-processamento, representação, análise e relato, e o trabalho organiza-se em três questões de investigação faseadas: (i) quais são as associações comportamento–motivação? (ii) quais são os indicadores comportamentais refinados? e (iii) qual é a estimação ao nível individual por integração sensível ao contexto com inferência aumentada por LLM? Mantém-se a governação ética rigorosa em todo o processo.

Os resultados mostram que a combinação de sensorização passiva contínua com autorrelatos SDT identifica padrões comportamentais associados à motivação intrínseca e produz indicadores com maior ranularidade, sensibilidade temporal superior e fiabilidade psicométrica mais robusta. Evidenciam também uma estimativa simultânea interpretável e individualizada, viabilizada pela inferência contextualizada e aumentada pelo LLM. Globalmente, o trabalho estabelece um paradigma sistemático que liga a teoria psicológica ao comportamento no mundo real, progredindo da associação para a otimização e para a estimação. Oferece também orientações práticas para uma avaliação pouco intrusiva da motivação para a aprendizagem e para a análise da aprendizagem.

Palavras-chave: Avaliação, Motivação académica, Monitorização passiva por sensores, Psicométrico

Abstract

A Hybrid Psychometric Approach Integrating Self-Determination Theory and Passive Smartphone Sensing for Continuous Assessment of Learning Motivation

Learning motivation profoundly shapes students' behavior and engagement, making it central to educational psychology and pedagogy. In Self-Determination Theory (SDT), intrinsic motivation is a stable construct of genuine drive. Conventional assessments based on self-reports or expert observations are labor-intensive, bias-prone, and miss temporal dynamics. Although mobile sensing is widely used in psychometrics, work directly targeting learning motivation and a SDT-aligned, continuously assessed, integrated approach remains limited. This dissertation addresses that gap.

We propose a hybrid psychometric approach integrating SDT-grounded self-reports with passive smartphone sensing for continuous, ecologically valid assessment, supported by longitudinal studies that pair behavioral data gathered using smartphone with validated motivation scales. An operational five-layer workflow spans perception, preprocessing, representation, analysis, and reporting, and the work is structured by three staged research questions: (i) what are the behavior–motivation associations? (ii) what are refined the behavioral indicators? and (iii) what is the individual-level estimation via context-aware integration with LLM-augmented inference? Strict ethical governance applies throughout.

Results show that combining continuous passive sensing with SDT self-reports identifies behavioral patterns linked to intrinsic motivation and yields refined indicators with finer granularity, greater temporal sensitivity, and stronger psychometric reliability. They also demonstrate interpretable, individualized concurrent estimation enabled by context-aware, LLM-augmented inference. Overall, the research establishes a systematic paradigm connecting psychological theory to real-world behavior, progressing from association to optimization to estimation, and offering practical guidance for unobtrusive learning motivation assessment and learning analytics.

Keywords: Assessment, Learning Motivation, Passive Smartphone Sensing, Psychometric

Contents

List of Figures **xi**

List of Tables **xii**

Acronyms **xiv**

1 Introduction **1**

1.1 Research Background 1

1.2 Problem Statement 4

1.2.1 Limitations of Traditional Learning Motivation Assessment 4

1.2.2 Existing Mobile Sensing Studies for Learning motivation and Related Psychological Constructs 5

1.2.3 From Traditional Measures to Mobile Sensing: Research Gaps and Challenges in Motivation Assessment 6

1.3 Research Problem and Objectives 8

1.3.1 Research Problem 9

1.3.2 Research Questions 9

1.4 Research Scope, Boundaries and Limitations 12

1.4.1 Research Boundaries 14

1.4.2 Research Limitations 15

1.5 Methodological Overview 17

1.5.1 Comprehensive Research Design 18

1.5.2 Data Collection and Preprocessing 18

1.5.3 Progressively Staged Study Design 20

1.5.4 Overall Architectural Framework and Data Flow 23

1.6 Research Contributions 30

1.7 Human Subjects' Protection 32

1.8 Structure of the Dissertation . . . 33
1.9 Summary . . . 35
1.10 List of Publication . . . 35

2 Background and Related Work 36
2.1 Theoretical Background . . . 36
2.1.1 Motivational Continuum, Internalization, and Assessment Implications . . . 37
2.1.2 Temporal Dynamics of Intrinsic Learning Motivation . . . 40
2.1.3 Conceptual Bridge to Sensing . . . 43
2.2 Technical Foundations . . . 44
2.2.1 Passive Smartphone Sensing and Digital Phenotyping . . . 44
2.2.2 Behavioral Representation . . . 48
2.2.3 Learning Motivation Modeling (Large Language Model (LLM) Modeling) . . . 51
2.3 Review of Related Work: Assessing Learning Motivation in Higher Education . . . 54
2.3.1 Traditional Assessment Methods for Learning Motivation in Higher Education 54
2.3.2 Using Passive Smartphone Sensing to Assess Learning Motivation and Related Constructs . . . 55
2.3.3 Method Comparison and Research Gaps . . . 57
2.4 Chapter Summary . . . 62

3 Passive Smartphone Sensing Design, Data Collection, and Preprocessing 63
3.1 Passive Smartphone Sensing Design . . . 63
3.1.1 Modalities and Sampling Strategy . . . 64
3.1.2 *iSense* (Android-native) App: Functional Flow and Screens . . . 64
3.1.3 Backend and Version Compatibility (from *isense-console* source) . . . 75
3.1.4 Power, Privacy, and Security by Design . . . 75
3.2 Data Collection . . . 75
3.2.1 Scales and Cadence (Co-period Collection) . . . 75
3.2.2 Sensor Lists . . . 76
3.2.3 Standard Operating Procedure . . . 84
3.3 Preprocessing / Reprocessing . . . 85

4 Association Analysis of Daily Behaviors and Intrinsic Learning Motivation 88
4.1 Alignment with Research Questions (RQ1) . . . 88
4.2 Method Design Overview (Pointers to Shared Procedures) . . . 89
4.3 Contributions of This Chapter . . . 89
4.4 Methods . . . 89
4.4.1 Study Design . . . 89

4.4.2 Demographics . . . 90
4.4.3 Mobile Sensing Data . . . 91
4.4.4 Intrinsic Motivation Assessment . . . 92
4.4.5 Behavioral Features . . . 92
4.4.6 Association Analysis Pipeline . . . 93
4.4.7 Ethics and Privacy . . . 93
4.5 Results . . . 94
4.5.1 Behavioral Patterns (from Principal Component Analysis (PCA)) . . . 94
4.5.2 Associations with Intrinsic Motivation (intrinsic learning motivation) . . . 95
4.6 Discussion . . . 96
4.7 Limitations . . . 97
4.8 Summary . . . 97

5 Optimizing Behavior Identification by Refining Indicators from Behavioral Changes 98
5.1 Alignment with Research Questions (RQ2) . . . 98
5.2 Study-Specific Method Design Beyond Chapter 3 . . . 98
5.3 Methods . . . 99
5.3.1 Demographics . . . 99
5.3.2 Data Collection . . . 100
5.3.3 Behavioral Feature Set . . . 101
5.3.4 Behavior Identification via Clustering . . . 102
5.3.5 Association Analysis and Between-Group Comparison . . . 102
5.4 Results . . . 102
5.4.1 Nine Behavior Patterns (Clusters) and Interpretations . . . 103
5.4.2 Associations Between Behavior Patterns and intrinsic learning motivation . . . 104
5.4.3 Refined Behavioral Indicators . . . 106
5.5 Discussion . . . 107
5.6 Limitations . . . 108
5.7 Summary . . . 109

6 Concurrent Estimation of Learning Motivation with LLM-Augmented Modeling 110
6.1 Alignment with Research Questions (RQ3) . . . 111
6.2 Methods . . . 111
6.2.1 Scope and Study-Specific Additions Relative to Chapter 3 . . . 112
6.2.2 Study Design and Participants . . . 112
6.2.3 Passive Sensing and Self-Reports . . . 112
6.2.4 Representation: Textualization of Behavioral Features . . . 113
6.2.5 Tasks and Prompting . . . 113

6.2.6 Evaluation Metric . . . 115
6.3 Results . . . 115
6.3.1 Overall Estimation Accuracy (Shot Scaling) . . . 116
6.3.2 LLM Reasoning Patterns (Chain-of-Thought (CoT) Illustrations) . . . 117
6.4 Discussion . . . 117
6.5 Limitations . . . 118
6.6 Implications . . . 119
6.7 Ethics and Privacy . . . 120
6.8 Summary . . . 120

7 Conclusion and Future Work 122
7.1 Summary of Key Findings and Contributions . . . 122
7.1.1 Integrated Mobile Sensing for Learning Motivation: An Initial Longitudinal Study Series . . . 123
7.1.2 Methodology: An Innovative Hybrid Psychometric Approach and An Operationalizable Framework . . . 123
7.1.3 Self-Determination Theory: Theoretical Framing and Ecologically Valid Feedback 124
7.1.4 Practice: Guidelines for Unobtrusive, Privacy-Preserving Use . . . 124
7.2 Limitations . . . 124
7.2.1 Psychological Theory Scope . . . 125
7.2.2 Methodological and Data-Related . . . 125
7.2.3 Computational and Modeling . . . 126
7.3 Future Work . . . 127
7.3.1 Diverse and Cross-Cultural Validation . . . 127
7.3.2 Develop Automated Tools or Applications Utilizing Motivation-Aligned Behavioral Indicators to Assess Learning Motivation . . . 128
7.3.3 Advanced Modeling Towards Real-Time Estimation and Prediction . . . 128
7.3.4 Personalization with Learning Analytics, Adaptive Feedback and Intervention 129
7.3.5 Ethical and Responsible Deployment . . . 129
7.4 Closing Conclusion . . . 132

Bibliography 134

List of Figures

1 Conceptual framework of learning motivation grounded in Self-Determination Theory (Self-Determination Theory (SDT)). 2
2 The motivational continuum in Self-Determination Theory (SDT). 3
3 A five-layer architectural framework of the hybrid approach 24

4 Icon of *iSense*. 65
5 User interface for joining the project. 65
6 User interface for logging in. 66
7 Granting required permissions (overview). 67
8 Granting notification access. 69
9 Permission and service status check. 71
10 Add *iSense* to battery optimization whitelist. 72
11 Locate *iSense* in the “Not allowed” list and set to “Not optimized”. 73

12 Cumulative variability in the sensed data explained by the top-n principal components. . 93

13 Average silhouette vs k fork-means (Euclidean) on z-scored features. The maximum occurs at k=9 (0.28), which we adopt for subsequent clustering and per-cluster interpretation . 102
14 Features with significant differences between high- and low-intrinsic learning motivation groups (positive bars indicate higher values in the high-intrinsic learning motivation group). . . . 105

15 Zero-shot CoT prompting schematic. 114
16 Few-shot CoT prompting schematic. 115
17 Subject-wise absolute error distributions ($|P-T|$; 1–5 scale). Ten-shot CoT; $N = 17$; subjects ordered by median error. 116

List of Tables

1 Method comparison across approaches used to assess learning motivation and related constructs. . . . 61

2 Intrinsic Learning Motivation Items . . . 76
3 Event-driven modalities (API/fields/privacy) . . . 77
4 Active-triggered sensors (API/fields/privacy) . . . 81

5 Demographics of participants used in the analysis ($N = 58$). . . . 91
6 Statistics of Intrinsic Motivation (Self-Regulation Questionnaire—Academic (SRQ-A) intrinsic subscale). . . . 92
7 Features extracted and grouped by social interactivity. . . . 92
8 Behavioral patterns and their matching features (indexed by the n-th PCA component). . 94
9 Behavioral patterns associated with intrinsic learning motivation (intrinsic learning motivation). . . . 96

10 Participant demographics and sample flow . . . 100
11 Behavior patterns (clusters) identified from 59 sensing features and their top feature signals. 103
12 Behavior patterns associated with intrinsic learning motivation (intrinsic learning motivation). . . . 104
13 Refined behavioral features differentiating high vs. low intrinsic motivation. Values reflect daily averages; a positive sign indicates the feature is higher in the high-intrinsic learning motivation group. All displayed differences remain significant after Benjamini–Hochberg False Discovery Rate (FDR) correction at $p < 0.05$ (Figure 15). . . . 106
14 High-level mapping from refined indicators to expected direction w.r.t. intrinsic learning motivation 107

15 Average Mean Absolute Error (MAE) across shots (row = metric, columns = shots). Lower is better; best in bold. . . . 116
16 Zero-shot vs. Ten-shot CoT predictions and rationales (per participant). . . . 117

17 Future Work: Directions, Overviews, and Potential Applications 132

Acronyms

AMS Academic Motivation Scale

BLE Bluetooth Low Energy

COPUS Classroom Observation Protocol for Undergraduate STEM
CoT Chain-of-Thought
CPU Central Processing Unit

DB Database
DST Daylight Saving Time

EMA Ecological Momentary Assessment
ESM Experience Sampling Method

FDR False Discovery Rate

GEE Generalized Estimating Equations
GPA Grade Point Average
GPS Global Positioning System

HCI Human Computer Interface
HDBSCAN Hierarchical Density-Based Spatial Clustering of Applications with Noise
HE Higher Education
HTTPS Hypertext Transfer Protocol Secure

IDE Integrated Development Environment

IM	Instant Message
IMI	Intrinsic Motivation Inventory
IRB	Institutional Review Board
JSON	JavaScript Object Notation
LLM	Large Language Model
LMS	Learning Management System
MAE	Mean Absolute Error
MDUI	Material Design User Interface
MongoDB	A Document-oriented NoSQL Database
MSLQ	Motivated Strategies for Learning Questionnaire
OS	Operating System
OSS	Object Storage Service
PCA	Principal Component Analysis
SDK	Software Development Kit
SDT	Self-Determination Theory
SHAP	SHapley Additive exPlanations
SMS	Short Message Service
SRQ	Self-Regulation Questionnaire
SRQ-A	Self-Regulation Questionnaire—Academic
TLS	Transport Layer Security
TSBH	Two-Stage Benjamini–Hochberg
Vite	A Modern Frontend Build Tool and Development Server
Vue 2	Vue.js (The Progressive JavaScript Framework), version 2 (major release)

1 Introduction

This chapter provides a concise overview of the problem space, the overall methodological approach, and the comprehensive research design for this dissertation on assessing students' learning motivation with a novel hybrid approach. It first motivates why learning motivation, especially intrinsic motivation within Self-Determination Theory (SDT), matters for engagement and achievement, and why conventional psychological assessments are limited by episodic administration, reporting bias, and coarse temporal granularity. It then discusses the advantages of mobile sensing as a new measurement approach in this problem space and critically reviews the limitations of existing studies that combine mobile sensing to measure learning motivation and closely related constructs. Building on these limitations, the chapter introduces passive smartphone sensing as an unobtrusive and scalable complement for continuously capturing everyday behavioral data, and it outlines the practical challenges of this route, including inferring latent constructs from raw data, accommodating individual differences, modeling temporal dynamics, and maintaining unobtrusiveness. In response, the chapter proposes a hybrid approach integrating passive smartphone sensing with self-reports, which aligns SDT-grounded measures with behavioral features and, in later stages, incorporates large language model (LLM). Approaching augmented inference to connect sensor traces to high-level motivational constructs, enables continuous, concurrent estimation of learning motivation. Moreover, human subjects protection is treated as a primary concern throughout the research design and conduct. This chapter provides a detailed discussion of the associated ethical approvals, safeguards, and compliance procedures; it then previews the research contributions and presents the overall structure of the dissertation. Collectively, this chapter establishes the theoretical and methodological foundations and positions the dissertation to advance an innovative hybrid approach to continuously assess students' learning motivation in authentic settings.

1.1 Research Background

Learning motivation plays a pivotal role in the learning process, as it profoundly impacts students' learning behaviors and engagement [1–3]. Consequently, it has been extensively studied across disciplines such as educational psychology and pedagogy [4, 5]. This dissertation investigates learning motivation from

a psychological perspective, focusing specifically on students' dynamic motivation levels during learning, as conceptualized under the framework of Self-Determination Theory (SDT) proposed by Deci and Ryan [1, 2, 6–9] and shown in Figure 1.

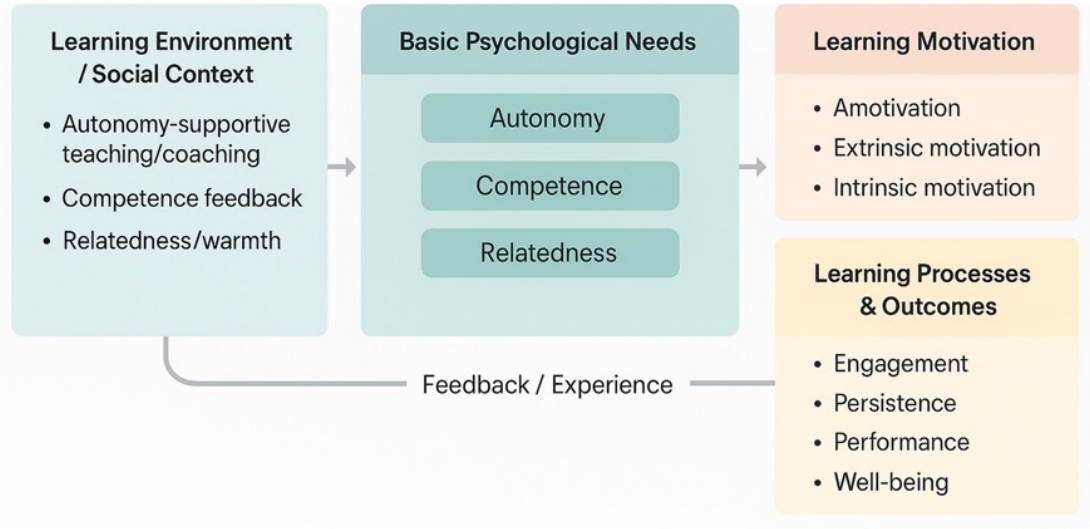


Figure 1: Conceptual framework of learning motivation grounded in Self-Determination Theory (SDT).

Self-Determination Theory (SDT) proposes that when the three basic psychological needs—autonomy, competence, and relatedness—are satisfied, individuals' motivation is enhanced [7, 9]. Within SDT's framework of motivational regulation [10], intrinsic motivation reflects an individual's inherent psychological tendency to engage in learning for interest and satisfaction rather than for external rewards or pressures [2, 6, 7]. SDT theory positions motivation along a continuum from amotivation to intrinsic motivation, which emphasizes that extrinsic motivation can be gradually internalized—progressing from external regulation through introjected and identified regulation to integrated regulation [10–12], ultimately internalizing as intrinsic motivation, illustrated in Figure 2 below. Hence, focusing on intrinsic motivation as the key psychological construct provides a more direct and stable indicator of students' genuine learning drive and overall motivational quality. This research emphasizes the assessment of intrinsic motivation as the key construct of overall learning motivation.

Accurately assessing students' motivation level is a prerequisite for deploying timely support strategies that are aimed at enhancing motivation and related engaging behaviors in real learning environments. Though various studies have proposed assessment methods for learning motivation, most rely on traditional self-report instruments (e.g., questionnaires) [13–17], Experience Sampling Method (ESM) protocols [18, 19], or expert evaluations (e.g., observations, interviews) [20, 21]. These approaches can capture important aspects of students' experiences, but they are often time-consuming to administer and score, lack real-time feedback that could be used for in-the-moment support, and are prone to biases that result from manual interpretation, retrospective recall, or evaluator subjectivity. Though traditional sensor-based Ecological Momentary Assessment (EMA) methods enable more immediate assessment to some extent [22, 23], the frequent prompting can disrupt participants' ongoing activities, increase burden, and induce behavioral

reactivity and measurement bias [24–26].

While some neuroscience methods (e.g., EEG) [27, 28] can be used to measure learning motivation by linking neural activity to motivational states, their broader applicability, scalability, and generality are constrained by the need for specialized equipment, controlled lab conditions, and expert interpretation [29]. As a result, such methods are difficult to deploy at scale in everyday educational settings, such as regular classrooms or dormitories, and they are less suitable for continuous, low-burden monitoring of motivation across large and diverse student populations.

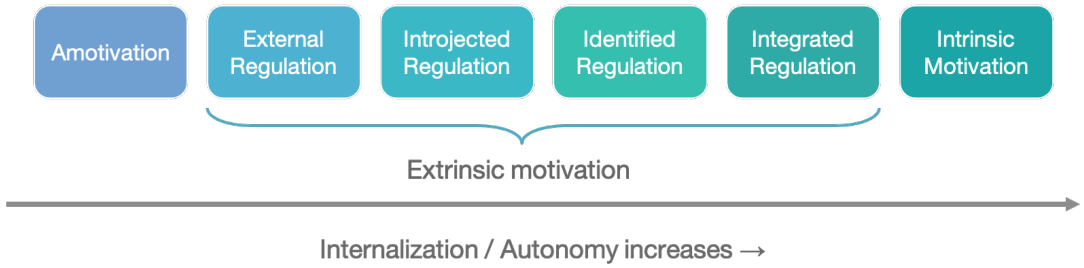


Figure 2: The motivational continuum in Self-Determination Theory (SDT).

From a longitudinally point of view, motivation levels are relatively stable psychological constructs that change gradually rather than abruptly. This attribute makes periodic and continuous assessment feasible and meaningful. A sensing framework that can repeatedly observe these temporal fluctuations can therefore provide more reliable and ecologically valid evidence for learning motivation assessment.

Mobile sensing technologies hold notable potential for capturing internal psychological constructs [30–35], and prior work has demonstrated their feasibility in contexts such as schizophrenia detection [36–38] and long-term monitoring of student mental health [39, 40]. These studies suggest that combining passive sensing with traditional assessment tools opens new directions for psychological assessments. However, only few research specifically target the sensing of learning motivation [41], and there is a notable absence of hybrid approaches that integrate mobile sensing for real-time and continuous motivation assessment [31, 32]. Although individual differences influence behavioral expression, under similar learning environments certain behavioral patterns may consistently correspond to similar learning motivation level [2, 25, 42]. This provides a basis for identifying shared behavioral indicators of learning motivation while still allowing person-specific baselines to account for individual variability. To address this gap, this dissertation proposes an innovative hybrid approach that leverages smartphone sensing technologies to capture students' dynamic learning motivation levels.

Despite its potential, employing mobile sensing to assess learning motivation poses a range of challenges that require further investigation. Specifically, what kind of sensing infrastructure should be used to effectively collect students' behavioral data? How should the experimental protocol be designed to enable reliable and ecologically valid data collection? How can raw sensor data be translated into interpretable behavioral features and patterns? What analytic strategies are needed to examine the relationships between

these behavioral data and motivation level? Finally, what are the target outcomes in terms of motivation that should be assessed or predicted? These key questions serve as the foundation for the studies presented in this dissertation.

1.2 Problem Statement

This section delineates the problem space that motivates the present dissertation by contrasting traditional assessments of learning motivation with emerging mobile sensing approaches and by highlighting their respective limitations. It first reviews how conventional self-reports and expert observations, despite their widespread use and practical value, suffer from sparse sampling, recall and social desirability biases, and limited scalability, which constrain their temporal resolution, ecological validity, and granularity. It then summarizes existing mobile sensing studies that leverage smartphone-based behavioral traces to infer psychological constructs, noting their promise but also the scarcity of work that directly targets learning motivation or explicitly aligns sensing features with Self-Determination Theory. Building on this, the section articulates three main research gaps, namely the lack of integrated, theory-aligned hybrid methodologies, the absence of a reusable motivation-related feature space, and the limited development of individual-level modeling. It also discusses key methodological challenges for a hybrid route, including latent construct inference, personalization, temporal dynamics, integration protocols, and unobtrusive, ethically governed sensing. Together, these issues motivate the need for a systematic SDT-grounded framework that combines traditional measures with passive smartphone sensing and thus set the stage for the research scope and questions formulated in Section 1.4.

1.2.1 Limitations of Traditional Learning Motivation Assessment

Traditional methods for assessing learning motivation, mostly non-invasive and highly operable such as self-report questionnaires [13–16, 43] and expert interviews [20], have long been dominant in educational psychology [44]. They are widely used in both research and practice to capture students' self-perceived reasons for studying, levels of interest, and satisfaction of basic psychological needs, and they can be administered with relatively low cost within existing curricular and assessment routines. As a result, they have become a standard reference point for describing motivational profiles and evaluating the impact of instructional interventions.

While these tools (including standardized instruments like the Academic Motivation Scale or custom surveys based on SDT) [45] provide valuable insights, they are inherently limited in several ways. In particular, they primarily offer retrospective and self-reflective accounts, rather than direct traces of behavior in context, and therefore struggle to capture fine-grained temporal dynamics, situational influences, and day-to-day fluctuations in learning motivation. These constraints become especially salient when the aim is to move from occasional diagnostic use toward continuous monitoring or the design of timely, data-informed interventions.

- First, such methods are typically administered at fixed time points, which fails to capture the dynamic and fluctuating nature of motivation throughout the learning process [46]. In most educational settings, questionnaires or interviews are scheduled only at a few pre-defined occasions (e.g., beginning, middle, or end of a semester), so they provide isolated snapshots rather than a continuous record of change. This coarse sampling overlooks short-term rises and drops in motivation that occur across days, weeks, or specific learning activities, and therefore cannot adequately represent trajectories, turning points, or rapid deteriorations that may be critical for intervention.
- Second, self-reports are subject to memory biases [47, 48], social desirability effects [49, 50], and participants' interpretations of questions [47, 51, 52], all of which can lead to inaccuracies. When students are asked to summarize how they "usually" feel or behave, they tend to rely on incomplete recall or on a few salient episodes, which distorts the frequency and intensity of their actual experiences. At the same time, they may respond in ways that present themselves as diligent, engaged, or autonomous learners, or they may simply misunderstand key terms in the items, thereby introducing both systematic and random error into the measurement of motivation.
- Third, observational assessments by experts are time-consuming and resource-intensive [53], which also do not scale to large populations. Reliable observation protocols require trained raters, careful scheduling of classroom visits or video recordings, and subsequent coding and consensus procedures, all of which demand substantial time and financial investment. As class sizes and program cohorts grow, it becomes increasingly infeasible to provide fine-grained observational coverage for every student, resulting in sparse data that cannot support continuous or large-scale monitoring of learning motivation.

Collectively, these drawbacks restrict the timeliness, ecological validity, and granularity of motivational data that can be obtained through traditional approaches. These constraints suggest the need for a complementary approach that improves ecological validity and temporal resolution; passive smartphone sensing provides such a route by enabling unobtrusive, fine-grained observation in real-world learning contexts.

1.2.2 Existing Mobile Sensing Studies for Learning motivation and Related Psychological Constructs

In recent years, mobile sensing technologies have emerged as promising tools for unobtrusively capturing behavioral patterns that are potentially linked to psychological constructs [54–57]. Prior studies have explored the use of smartphones to monitor students' mental health [58, 59], academic performance [39, 59], and well-being in university students [60, 61]. For instance, several projects have employed data from GPS, smartphone screen usage, and ambient sensors to measure and infer psychological phenomena like depression [60, 62, 63], stress [64], or other mental states [61]. These studies have demonstrated that passive sensing can capture subtle behavioral data, such as mobility routines, phone

interaction patterns, or social proximity, that correlate with important psychological variables in academic contexts.

However, only a limited number of studies have directly targeted learning motivation itself [41]. The majority of mobile sensing research focuses on broader psychological constructs (e.g., general mental health, stress, depression, anxiety), often treating motivation only as a secondary outcome or proxy. Additionally, few studies explicitly align their behavioral sensing efforts with established psychological theory such as SDT [65], leaving a gap in the integration of theory-driven motivational constructs with sensing-based features. This lack of interdisciplinary hybrid approaches, combining passive sensing with well-founded learning motivational frameworks, represents a critical research gap that the present dissertation aims to address.

Taken together, current work reveals the promise of passive sensing but ignores critical gaps for learning motivation in particular—namely, (i) a systematic, theory-aligned hybrid approach, (ii) a reusable learning motivation associated behavioral feature space for replication and comparison, and (iii) individualized modeling.

1.2.3 From Traditional Measures to Mobile Sensing: Research Gaps and Challenges in Motivation Assessment

While both traditional assessments and passive sensing offer valuable perspectives, a clear research gap remains in how to combine these approaches to assess learning motivation in a way that is theoretically grounded and practically useful. In particular, there remains no widely accepted framework capable of integrating traditional self-report measures and smartphone-sensed behavioral data for assessing learning motivation grounded in Self-Determination Theory.

- **Research Gap**
  - **Lack of Integrated, Theory-Aligned Methodology:** The field lacks an integrated, theory-aligned methodology that enables unobtrusive, continuous, and real-time assessment of learning motivation [66]. Existing work is often fragmented across isolated stages of the pipeline, for example focusing only on particular sensing modalities, specific preprocessing techniques, or stand-alone prediction models, rather than on a coherent end-to-end framework. As a result, few current studies articulate a systematic approach that jointly covers data collection, preprocessing, behavioral identification, alignment with SDT constructs, and association and prediction analyses tailored to this outcome. This absence of a unified methodological blueprint makes it difficult to design comparable studies, to interpret results in a cumulative way, and to translate findings into scalable monitoring tools for real-world educational settings.
  - **Absence of A Reusable, Cross-Study Motivation Feature Space:** There is a shortage

of a reusable learning motivation–associated behavioral feature space (e.g., a feature dictionary and shared behavioral clusters) that supports replication and cross-study comparison. Most existing studies define their own ad hoc sets of features and behavioral markers, which are tightly bound to a particular dataset, context, or technological stack, and are seldom documented in a way that facilitates reuse. Consequently, it is challenging to compare effect sizes, to re-implement feature extraction in new cohorts, or to assess whether similar behavioral regularities emerge across institutions and student populations. Without a shared, theory- and behaviorally grounded feature space, the field lacks a common reference for mapping digital traces to motivational constructs and for building cumulative evidence.

  - **Individual-Level Modeling and Estimation Gap:** Individual-level modeling and estimation remains under-explored; most existing studies stay at the group level, with limited attention to between-group comparisons, person-specific baselines, and temporal alignment with students' daily study routines. In many cases, analyses are based on aggregated or cross-sectional data, which tends to obscure intra-individual variability and short-term motivational fluctuations that are central to SDT-aligned perspectives on learning. This limits the ability to capture how each student's motivational states evolve over time in response to contextual factors, and how these trajectories differ between subgroups (e.g., courses, programs, or institutions). More systematic individual-level modeling would enable the estimation of personalized baselines, the detection of meaningful deviations from those baselines, and the examination of how motivational dynamics unfold within the fine-grained temporal structure of students' real-world study routines.

- **Methodological Challenges within The Hybrid Route**

  - **Latent Construct Inference:** Inferring a latent motivation level from raw behavioral data requires explicit theoretical alignment and reliable labels. Intrinsic learning motivation is not directly observable since it is a trait-like psychological construct grounded in the humanistic SDT theory. It must be inferred from externally manifested behaviors. To reduce construct ambiguity, sensor-derived features should be aligned with SDT constructs and linked to self-report measures by taking into account label quality and noise control.

  - **Personalization:** Motivational expression varies across students, so assessment must account for individual differences while retaining shared structure under comparable learning environments. The same behavior may reflect different states for different learners; individual-level baselines and simple calibration help correct such offsets. At the same time, when contexts are broadly similar, certain behavioral patterns can reliably map to similar learning motivation level, supporting generalization and fairness.

- **Temporal Dynamics:** Learning motivation level is dynamic and changes over time. Learning motivation fluctuates within days and weeks in response to contextual factors such as workload, feedback, stress, and social interactions, even though it shows relative stability across longer periods (for example, across a semester or academic year). Static or one-off measures miss this temporal structure and can only provide snapshots that average over these fluctuations. As a consequence, they risk obscuring short-term rises and drops that are theoretically important for understanding self-regulation, engagement, and risk states, and they limit the possibility of examining lead–lag relations between motivational shifts and subsequent behavior.
- **Integration Protocols:** Combining traditional assessments with sensing features calls for clear rules about when and how fusion occurs. In practice, this includes procedures for time alignment between questionnaire administrations and sensing windows, normalization of heterogeneous feature scales, reliability weighting of different information sources, and principled handling of missing or noisy data as part of the protocol to ensure validity and robustness. Without such explicit integration protocols, hybrid approaches remain ad hoc, are difficult to reproduce across studies, and become vulnerable to biases introduced by irregular sampling, unequal measurement quality, or arbitrary choices made at the fusion stage.
- **Unobtrusiveness and Ethical Governance:** Assessment integrating continuous sensing must remain unobtrusive and privacy-preserving. Intrusive data collection methods can alter participants' natural behavior or raise privacy concerns, ultimately compromising ecological validity. Developing sensing frameworks that operate seamlessly in the background without interrupting the learning experience is a crucial design consideration. Purpose limitation, informed consent, data minimization, secure storage, and controlled access are required; specific safeguards and governance practices are outlined in Section 1.7.

Together, these research gaps and challenges define the problem space addressed in this dissertation. In response, Section 1.4 clarifies the scope of this dissertation, states the research problem, and sets out the research questions that guide the subsequent work.

1.3 Research Problem and Objectives

This section refines the overall research problem and translates it into a set of research questions and objectives that structure the dissertation. It starts from the challenge of assessing learning motivation in ecologically valid ways by combining conventional psychological assessments with passive smartphone sensing under the Self-Determination Theory framework, and it formulates a research problem focused on aligning raw behavioral data with SDT-grounded learning motivation in an interpretable, reliable, and individually useful manner. Building on this, the section introduces three research questions that guide the

empirical work: an exploratory question that examines how university students' daily behaviors captured via passive sensing relate to their learning motivation; a confirmatory question that seeks refined, temporally sensitive behavioral indicators of learning motivation; and a predictive question that investigates to what extent individual motivation levels can be estimated through context-aware integration of sensing data. For each question, specific objectives are outlined, including the design of an unobtrusive sensing framework, the construction and refinement of SDT-aligned behavioral feature spaces, and the development of individualized, interpretable, and scalable modeling approaches that respect temporal dynamics, inter-individual variability, and ethical and privacy-preserving data practices.

1.3.1 Research Problem

This dissertation addresses the research problem of assessing learning motivation in ecologically valid ways by combining conventional psychological assessments with passive smartphone sensing under the SDT framework. The core challenge is to align raw behavioral data with SDT-grounded learning motivation so that results are interpretable, reliable, and useful at the individual level.

Within this problem space, this work advances from analyzing behavior-learning-motivation associations, to optimizing behavior identification for refined learning-motivation-aligned behavioral indicators with reliability and temporal sensitivity, and finally predicting individualized learning motivation level with context-aware modeling. Throughout, the approach recognizes inter-individual variability, short-term fluctuations versus longer-term stability, and the need for unobtrusive and privacy-preserving data practices.

1.3.2 Research Questions

Given this research problem framing, we articulate the following research questions that guide the research and provide a focused agenda for the work, structuring the subsequent empirical studies and methodological developments presented in this dissertation:

Research Question 1 (Exploratory): How do university students' daily behaviors relate to their learning motivation levels in real-world settings captured via passive smartphone sensing?

Research Question 2 (Confirmatory): What passively sensed behavioral indicators can effectively characterize the temporal dynamics of students' learning motivation?

Research Question 3 (Predictive): To what extent can individual motivation level be estimated by integrating passive smartphone sensing with context-aware inference?

RQ1: Exploring Behavioral Correlates of Learning Motivation in University Students. Learning motivation plays a pivotal role in academic performance and engagement, shaping how persistently students study, how deeply they process course material, and how they respond to challenges; however, the behavioral patterns that reflect students' learning motivation level in real-world settings, across days, courses, and study contexts, remain insufficiently understood. Traditional self-report and observational methods do not capture these dynamics in real-time, because they are typically administered at sparse

time points and rely on retrospective judgments, which leads to missed opportunities for timely support, such as the early detection of motivational declines and the provision of responsive, low-stakes interventions.

Objectives: To characterize daily behavioral data, the first part of this work combines passive smartphone sensing with SDT-based measures of intrinsic motivation. A lightweight sensing application collects mobility, phone interaction, and ambient data aligned with concurrently collected self-report measures of learning motivation.

- *Design a sensing framework suitable for continuous use in daily educational settings.* This framework should operate unobtrusively in the background, minimizing disruption to teaching and learning activities while still capturing rich, ecologically valid behavioral data over extended periods. It must also respect practical constraints such as device heterogeneity, battery consumption, data security, and privacy regulations, so that it can be realistically deployed in real-world classrooms and university environments.
- *Identify which behavioral features and patterns are informative of intrinsic learning motivation level.* This involves systematically deriving candidate features from multiple sensing modalities (e.g., mobility, device interaction, app use, social rhythms) and examining in detail how these candidate signals relate to theoretically grounded indicators of intrinsic motivation derived from established frameworks. The goal is to distinguish signal from noise by iteratively testing, refining, and pruning the feature set, so that we isolate stable and interpretable behavioral indicators that can reliably reflect motivational states across time and contexts, rather than incidental fluctuations, measurement artifacts, or highly context-specific behaviors that do not generalize beyond a particular course, cohort, or setting.
- *Examine whether mobile sensing can serve as a viable enhancement and complement to self-report in assessing learning motivation.* This entails systematically comparing sensing-based indicators with traditional self-report measures to evaluate not only convergent validity, but also the added predictive value that each source contributes and the robustness of these indicators over time under real-world conditions. In doing so, we seek to determine whether hybrid models that combine self-reports with passively sensed data can offer more continuous, fine-grained, and scalable assessments than self-report alone, while still preserving interpretability and psychometric quality at a level that is comparable to, or better than, conventional self-report–only approaches.

RQ2: Optimize Behavior Identification for Dynamic Learning Motivation Levels. While the preliminary exploration mainly reveals associations between passive sensing data and learning motivation level, it also provides the initial version of robust, generalizable behavioral indicators of learning motivation. This version requires refinement. This part improves behavior identification, introduces context-aware segmentation, and validates whether refined behavioral indicators can reliably distinguish students at

different levels of learning motivation.

Objectives:

- *Optimize the behavioral identification workflow by encompassing feature engineering and behavioral pattern identification to refine behavioral indicators of learning motivation, while aligning them with Self-Determination Theory (SDT) constructs to enhance interpretability and comparability.* Concretely, this involves systematically constructing, selecting, and transforming raw sensing signals into higher-level features, grouping them into meaningful behavioral profiles, and iteratively pruning or adjusting them based on their theoretical relevance and empirical performance. By explicitly mapping these refined behavioral indicators onto SDT dimensions (e.g., autonomy, competence, relatedness, intrinsic motivation), the workflow aims to produce a structured feature space that is both psychometrically grounded and reusable across studies, thereby supporting more consistent interpretation and comparison of results.
- *Compare behavioral differences between groups at different learning-motivation levels and behavioral changes within individuals as their motivation fluctuates, refine behavioral features that exhibit temporal sensitivity, stability, and strong informativeness.* This entails conducting both between-group analyses (e.g., contrasting high- versus low-motivation profiles) and within-person longitudinal analyses (e.g., tracking each student over time) to observe how specific behavioral patterns co-vary with measured motivation. Through these comparisons, we seek to identify which features consistently differentiate motivational levels across students, which features are responsive to short-term motivational shifts within the same student, and which subset of features provides a robust, temporally sensitive, and practically useful basis for modeling and estimating learning motivation.

RQ3: Towards Real-Time Assessment of Learning Motivation with Data-Driven Context-Aware Inference. Despite progress in refining behavioral indicators, translating complex behavioral data into individualized real-time estimations remains challenging; conventional models often lack interpretability or adaptation across students. In practice, turning high-dimensional, noisy, and context-dependent digital traces into moment-to-moment motivational estimates for each learner requires models that can handle person-specific baselines, heterogeneous routines, and shifting contexts without becoming opaque black boxes. However, many commonly used modeling approaches are trained at the group level, assume that all students follow similar patterns, and prioritize overall predictive accuracy over transparency, which limits their ability to explain predictions, to generalize across cohorts, and to support trustworthy, student-tailored monitoring in real educational settings.

Objectives:

- *Modeling learning motivation as a latent variable via multimodal integration.* In this work, learning motivation is treated as an unobserved construct that cannot be measured directly but must instead

be inferred from multiple observable indicators, including self-report scores, passively sensed behavioral traces, and contextual information. By integrating these heterogeneous modalities within a unified latent-variable framework, we aim to obtain more robust, theory-consistent estimates of students' motivational states than would be possible from any single data source alone.

- *Individualized real-time assessment based on historical and behavioral context with high interpretability.* The goal is to generate, for each student, moment-to-moment estimates of learning motivation that take into account their personal history, typical routines, and current behavioral signals, rather than relying solely on group-level averages. At the same time, the assessment models are designed to remain transparent and explainable, so that the contribution of specific features, past patterns, and contextual factors to each estimate can be inspected and communicated in an interpretable way.
- *Movement toward scalable, adaptive, and generalizable learning motivation assessment in educational settings.* Beyond proof-of-concept demonstrations, the objective is to contribute methods and workflows that can be deployed across larger cohorts, different courses, and diverse institutional environments without extensive re-engineering. This involves developing approaches that can adapt to new student populations and changing contexts while maintaining acceptable levels of validity and reliability, thus paving the way for practical, system-level integration of continuous motivation assessment into everyday educational practice.

1.4 Research Scope, Boundaries and Limitations

This section defines the scope, boundaries, and limitations of the dissertation by specifying what the research does and does not cover, and by clarifying the conditions under which its findings should be interpreted. It focuses on a hybrid psychometric assessment of learning motivation in full-time university students, combining SDT-based self-report questionnaires with passive smartphone sensing in authentic semester-based learning contexts, and aligning temporally aggregated behavioral features with motivational measures for association, feature refinement, and individualized, context-aware estimation tasks. At the same time, it sets explicit boundaries by restricting the work to non-invasive behavioral metadata (excluding neuroimaging and content-level data), to voluntary student samples observed over a single semester, and to a non-interventional design with strong privacy safeguards and strict adherence to Institutional Review Board (IRB)/ethics and data-protection principles. Finally, it acknowledges key limitations related to sampling and compliance, label quality for learning motivation, sensing coverage and missingness, non-stationarity in academic calendars, modeling and interpretability challenges, and privacy-driven constraints on data obtained from students' everyday smartphone use. Specifically, it aims to combine theory-grounded, questionnaire-based measures with ecologically valid behavioral traces collected via smartphones, so that learning motivation can be assessed more continuously, with richer contextual information, and with stronger links to actual day-to-day study behavior. By fusing these two complementary sources of information within a single framework, the research seeks to move beyond isolated,

snapshot-like measurements toward a more dynamic and context-sensitive characterization of learning motivation. The primary areas of research include the following:

- **Population & Setting:** Full-time university students in real-world learning contexts (coursework and daily study routines). The target population consists of students who are enrolled in regular higher-education programs and who follow typical semester-based schedules, including lectures, assignments, exams, and self-regulated study periods. Data collection is embedded into these authentic settings rather than controlled laboratory environments, ensuring that observed behaviors and motivational states reflect genuine academic pressures, constraints, and opportunities.
- **Data Modalities:** Passive smartphone sensing (mobility/activity, phone interaction, app categories and usage duration, environment metadata) aligned with concurrent SDT-based self-report measures of learning motivation. The sensing layer continuously records how students move, interact with their devices, and use different categories of applications, as well as selected contextual or environmental indicators, without requiring additional manual input. In parallel, SDT-based self-report scales are administered at planned intervals to capture intrinsic learning motivation and related constructs, providing a theory-informed ground truth against which the behavioral data can be aligned and interpreted.
- **Temporal Alignment:** Fixed aggregation windows to align behavioral features with learning motivation, with analyses at both within-person and between-group levels. Behavioral logs are segmented into predefined time windows (e.g., daily or weekly), and features derived from these segments are matched to the closest or overlapping self-report assessments, creating synchronized behavior–motivation pairs. This alignment enables the study of temporal dynamics within individual students over time, as well as systematic comparisons between subgroups that differ in overall motivation profiles, course enrollment, or other relevant characteristics.
- **Analytic Tasks:** Behavior–motivation associations; optimization of learning-motivation-aligned behavioral indicators (refined behavioral features with reliability and temporal sensitivity); individualized, context-aware concurrent estimation with person-level interpretability. First, we examine how specific behavioral features correlate with or predict measured learning motivation, identifying meaningful behavior–motivation links. Second, we iteratively refine the feature space to retain indicators that show stable associations, sensitivity to motivational fluctuations, and theoretical coherence with SDT. Third, we develop and evaluate models that estimate each student's motivational state in a concurrent and context-aware manner, while preserving interpretability at the person level so that the contribution of particular behaviors and historical patterns to each estimate can be understood and explained.

1.4.1 Research Boundaries

This subsection delineates the main boundaries within which the dissertation is conducted, specifying what is deliberately excluded from the scope of the work and under which conditions data are collected, processed, and interpreted throughout the empirical studies. It clarifies that only non-invasive assessment modalities based on behavioral metadata are used, with no invasive psychometric methods, neuroimaging procedures, or content-level data collection (e.g., no audio recordings, message text, or screen captures), and that participation is restricted to voluntary university students who are observed over a single academic semester within their usual study environment. The overall design is explicitly non-interventional and relies on passive background data collection rather than experimentally manipulating conditions or providing targeted interventions, so the resulting findings are interpreted as associational patterns rather than causal effects, even when relationships appear robust or temporally ordered. Finally, the subsection emphasizes strict privacy safeguards, including the use of only coarse location information, non-identifying representations of social interaction (for example, counts and durations instead of identities or content), and full adherence to IRB requirements and data-protection principles that govern storage, access, and secondary use of the collected data.

- **Assessment Modalities:** Exclude all invasive psychometric methods and neuroimaging techniques(e.g.,EEG); no collection of content-level data (e.g., message text, within-app content usage). This constraint ensures that all measurements remain strictly non-invasive and focus on behavioral metadata rather than the semantic content of communications or learning materials. The study therefore relies exclusively on surface-level interaction logs and sensor summaries, avoiding any direct inspection of what students read, write, or share, which in turn reduces both the intrusiveness of data collection and the ethical risks associated with accessing highly personal information.
- **Participants & Timeframe:** Only voluntary university student samples; longitudinal observation is limited to a semester scale, without long-term intervention evaluations. Participation is restricted to students who freely consent to take part in the research, with no compulsory or high-stakes involvement, and without extending the sampling frame to other populations such as school pupils or working adults. The observation period is deliberately confined to a single academic semester to capture meaningful medium-term dynamics in learning motivation while keeping the burden on participants manageable and avoiding claims about long-term developmental or intervention effects that the available data cannot justify.
- **Design Stance:** Unobtrusively designed, non-interventional studies are used throughout, and we make no causal claims beyond the level of associational evidence supported by robustness checks. Data are collected in a passive, background manner without manipulating students' learning environments, without changing course structures, and without delivering targeted motivational interventions, so the researcher's role remains observational rather than experimental at all stages of the work. Consequently, all analyses focus on identifying and characterizing robust associations

and patterns in the data, examining how behaviors and contexts co-vary with motivation, and using sensitivity and robustness checks to support these findings, but they do not extend to strong causal conclusions about specific behaviors or contextual factors, and any causal language is deliberately avoided in the interpretation.

- **Privacy:** No precise location tracking or inference of personal social content; strict adherence to IRB/ethics and data-protection principles. Location information, when used at all, is restricted to coarse or semantically abstracted representations that cannot reveal exact home addresses, specific room locations, or detailed movement trajectories, and is instead limited to higher-level zones such as "dormitory," "library," or "teaching building." Likewise, social interaction is represented only through non-identifying metadata (such as counts, durations, or temporal rhythms of communication) rather than any message content, contact identities, or conversation topics, so that the social graph and personal details remain protected. All procedures are reviewed and approved through appropriate IRB/ethics channels, and data handling follows established data-protection regulations from recruitment through analysis, including strict data minimization, secure storage on protected servers, controlled and role-based access to de-identified datasets, and the systematic use of anonymization or pseudonymization where applicable to reduce re-identification risk while preserving analytic utility.

1.4.2 Research Limitations

This following subsection critically examines the main limitations that qualify the evidence and generalizability of this work. It first notes that convenience sampling, attrition, and heterogeneous compliance restrict external validity and may introduce self-selection biases. It then highlights that the "ground truth" labels of learning motivation are based on self-reports that remain vulnerable to response styles, transient moods, and imperfect measurement invariance. Next, it discusses how operating-system policies, device differences, and permission changes generate nonrandom missingness and coverage gaps in sensing data, and how academic-calendar events and unmeasured contextual factors introduce non-stationarity that can destabilize learned relations over time. The subsection further points out that complex multimodal models risk overfitting and yield fragile feature attributions, so individual-level estimates must be interpreted cautiously. Finally, it stresses that strong privacy constraints, cross-site differences, and the impossibility of openly sharing raw logs limit reproducibility, making transparent reporting of procedures, feature definitions, and analysis pipelines essential.

- **Sampling & Compliance:** We employ convenience sampling (all participants are voluntary), and the passive sensing studies experience attrition as well as heterogeneous compliance. These factors limit external validity. Besides, the findings may not generalize to other institutions or cultural contexts. In particular, students who choose to participate may differ systematically from non-participants in terms of motivation, technology use, or attitudes toward monitoring, which can introduce self-selection bias. Attrition over time further skews the sample toward individuals who

are more compliant, more interested in the study, or less affected by technical burdens, thereby reducing representativeness as the study progresses. Heterogeneous compliance with app installation, permissions, and daily phone use also means that data quality and coverage vary across participants, making it difficult to extrapolate results to broader student populations or to educational systems with different technological or cultural infrastructures.

- **Label Quality (learning motivation "ground truth"):** Self-reports are vulnerable to participant response styles and emotions/moods, and measurement invariance[1] across days while persons may be imperfect. Though using short scales, unobserved bias can persist and may distort behavior–motivation associations. For example, students' ratings may be influenced by transient stress, fatigue, or social comparison on the day of answering, rather than by their typical learning motivation level. In addition, some respondents may show systematic tendencies such as acquiescence, extreme responding, or impression management, which are not fully corrected by scale length or item design. Even if formal tests of reliability and basic measurement invariance are conducted, subtle shifts in item interpretation or response styles across time or subgroups can remain, meaning that the so-called "ground truth" labels used to train or evaluate models still carry non-trivial noise and bias.

- **Sensing Coverage & Missingness:** Operating-system policies, background permissions, and device differences can cause nonrandom missing data and sensor noise. Though employing pre-registered rules for imputation and anomaly handling while making features more robust to noise, systematic coverage gaps may remain and may bias time-series estimates. For instance, newer operating-system versions or specific manufacturers' power-saving strategies may aggressively throttle background processes, leading to under-sampling of certain behaviors for subsets of users. Connectivity issues, storage limits, or users temporarily disabling permissions can also introduce patterns of missingness that correlate with stress, busyness, or other unmeasured states, rather than being random. Even with careful preprocessing, these structured gaps may distort apparent temporal trends or weaken the observed relationships between behavior and learning motivation, especially when trying to model fine-grained dynamics.

- **Non-Stationarity:** Academic-calendar events (e.g., examinations, pubic holidays) can change data patterns and make learned relations unstable. And residual confounding may remain since we cannot fully rule out other unmeasured factors, especially during exams weeks. In practice, behaviors such as study intensity, sleep patterns, mobility, and social interactions often shift markedly around key deadlines or holidays, altering the behavioral meaning of particular features over time. A feature that signals high motivation in regular weeks may instead reflect exam-related stress or overload during peak periods. Without explicitly modeling such non-stationarity and calendar

[1] "Measurement invariance"means the instrument functions equivalently across time and across persons. If this condition is not met, observed score differences may reflect changes in how the "ruler" works (e.g., item interpretation or response styles) rather than true differences in learning motivation).

structure, models trained on one phase of the semester may generalize poorly to another, and apparent associations may partly reflect unobserved changes in context rather than true motivational dynamics.

- **Modeling & Interpretability:** Complex models face risks of overfitting and instability of feature attributions, especially under multimodality. Individual-level estimation and inference may vary widely across contexts and should be interpreted with caution. Models that integrate multiple sensing streams and historical information can easily capture spurious correlations or cohort-specific artifacts, yielding high apparent accuracy but limited external validity. Moreover, feature attribution techniques (e.g., importance scores or local explanations) may be sensitive to model specification, data splits, or minor perturbations, making it difficult to draw firm substantive conclusions about which behaviors matter most. Consequently, even when models generate plausible individualized estimates, their outputs should be treated as probabilistic indicators rather than precise diagnostic scores, and any interpretation must acknowledge these modeling uncertainties.
- **Privacy & Reproducibility:** Strong privacy constraints limit data sharing and thereby reproducibility. We detailed data dictionaries where appropriate. Still, differences across sites and contexts limit cross-site reproducibility. Because raw sensing logs often contain potentially identifiable patterns and sensitive information about daily routines, strict privacy protections are necessary, which typically prevent open release of full datasets. As a result, other researchers must rely on derived features, synthetic examples, or high-level documentation, which may not fully capture all nuances of the original data. Furthermore, institutional policies, technical infrastructure, and student cultures differ across universities and countries, so even if methods are shared, replicating the same sensing setup and student behaviors elsewhere is challenging. These constraints mean that reproducibility efforts must focus on transparent reporting of procedures, feature definitions, and analysis pipelines, while accepting that exact data replication is often infeasible.

1.5 Methodological Overview

This section details how the hybrid psychometric approach is operationalized through an integrated methodological design that bridges conventional psychological assessment and data-driven mobile sensing analytics. It first outlines an overall design that combines a vertical three-stage progression from association analysis, through behavior identification optimization, to concurrent estimation with a horizontal five-layer architectural framework covering perception, preprocessing, representation, analysis, and reporting. It then describes how SDT-aligned self-report measures and passive smartphone sensing via the iSense app are collected and processed through a standardized pipeline into a subject–window dataset suitable for learning motivation sensing and measurement. Building on this foundation, the section introduces three progressively staged studies that establish feasibility, refine behavioral indicators, and achieve individual-level estimation using both explicit numeric modeling and textual LLM-augmented inference. Finally, it summarizes the overall data flow through the five-layer architecture and clarifies how information moves

from raw logs and questionnaires to behavioral representations, association analyses, concurrent estimation, and SDT-aligned reporting in real-world educational contexts.

1.5.1 Comprehensive Research Design

The overall methodological design of this dissertation is built upon a hybrid psychometric approach that bridges conventional psychological assessment and data-driven mobile sensing analytics. Methodologically, this dissertation adopts a vertical three-stage research progression intersecting a horizontal five-layer framework that provides a reusable, end-to-end backbone for all studies. Specifically, we employ the sequential three-stage research design that proceeds from behavior-learning-motivation association exploration, through behavior identification optimization, to learning-motivation concurrent estimation. In doing so, the previous three research questions are addressed, thereby demonstrating the feasibility of the overall approach. To conduct this research design, we develop a five-layer architectural framework comprising a staged workflow of perception, preprocessing, representation and behavior identification, analysis, report and explanation.

1.5.2 Data Collection and Preprocessing

Data originate from two complementary sources are processed through a standardized workflow to ensure quality, comparability, and interpretability for learning motivation sensing and measurement. For sensing data, we used a self-developed native Android app (iSense) to passively log behavioral data during daily study routines (see Chapter 3 for details); for learning motivation, we administered periodic online questionnaires to obtain concurrent self-reports aligned with the study timeline.

Data Collection:

- **Passive Smartphone Sensing:** Behavioral data of students in their daily real-world learning contexts. In this research, such data are collected through passive logging on students' own smartphones, without requiring additional devices or extra actions beyond normal phone use. Unobtrusive logging during everyday learning contexts captured a range of behavior- and context-related signals, allowing learning-related routines to be observed in situ rather than in artificial laboratory environments. Concretely, the sensing layer recorded the following categories of information:
  - *Mobility* (e.g., location changes and routine stability). These indicators summarize how students move across their typical environments (such as home, campus, and other frequently visited places) and how stable or variable their daily patterns of presence are over time. Rather than storing precise trajectories, we focus on aggregated descriptors of movement and regularity, which can reflect, in a coarse way, the structure and consistency of students' study-related routines.
  - *Screen interaction patterns* (e.g., unlocks, session lengths, overnight use). These features capture how often and when students interact with their phones, including the frequency

of unlock events, the duration of continuous usage sessions, and the presence of extended or fragmented use during late-night hours. Together, such interaction patterns provide a behavioral proxy for attentional rhythms, possible interruptions, and general engagement with digital devices across the day and night.

  - *Ambient context where permitted* (e.g., coarse light/noise indicators). When allowed by permissions and ethics constraints, we derive high-level summaries of the surrounding environment, such as approximate light levels or coarse proxies of ambient noise, without recording raw audio or identifiable environmental content. These contextual cues offer a lightweight indication of whether students are, for example, in quieter versus noisier settings or in brighter versus dimmer environments, which may co-vary with study situations.

  - *Usage of different applications* (e.g., application name, duration of application usage). The sensing system logs which applications are in use and for how long, focusing on app identifiers and usage durations rather than in-app content. This allows us to distinguish, at a coarse level, between time spent in different categories of apps (such as education, communication, or entertainment) and to examine how patterns of app usage relate to students' learning-related activities and motivational states in daily life.

- **Self-Report Measures:** The psychological scales, Academic Self-Regulation Questionnaire (SRQ-A) [13, 45] which derived from Self-Determination Theory (SDT), are adopted to measure learning motivation administered periodically and synchronously during behavioral data collection, obtaining concurrent motivation levels aligned with behavioral data for subsequent analyses.

Preprocessing: We ingest passive smartphone sensing data and conventional psychometric measures, assign unique participant IDs and convert all timestamps to a single time zone, and de-identify data under controlled access. At this initial integration stage, all raw logs and questionnaire responses are routed through a common pipeline in which identifiers are replaced by stable pseudonyms, direct personal identifiers are removed, and access is restricted to authorized researchers, ensuring that subsequent processing operates on a privacy-preserving dataset. We perform temporal alignment by aligning all streams to a fixed-length analysis window and predefine coverage metrics with exclusion or weighting rules, so that each behavioral and psychometric record is consistently mapped to comparable time slices. This step enforces a uniform temporal structure across heterogeneous data sources and encodes, in advance, how to treat windows with insufficient coverage (e.g., down-weighting or excluding them) rather than deciding case by case.

We then conduct event-level cleaning, including deduplicating records, correcting timestamp order and clock drift, and schema/unit harmonization across sources. Concretely, repeated entries caused by logging quirks are removed, events that appear out of sequence are reordered when warranted, and systematic offsets in device clocks are adjusted so that cross-stream comparisons remain meaningful. At the

same time, field names and measurement units are harmonized across different app versions or devices, producing a coherent event schema. Missingness is characterized with indicators, and conservative imputation is applied only when justified, with missing-data flags retained so that later analyses can explicitly account for gaps rather than silently ignoring them. Then we standardize features to a common scale, allowing downstream models and descriptive analyses to compare magnitudes across heterogeneous variables in a stable way. Finally, we aggregate to the analysis window and join sensing with psychometrics into a versioned subject–window dataset with a complete audit trail, so that every derived record can be traced back to its original inputs and preprocessing decisions for transparency and reproducibility.

1.5.3 Progressively Staged Study Design

The hybrid approach is realized and validated through the following three-staged studies, forming a cohesive progression from association to optimization to estimation. These three stages are designed to build on one another: each study uses the insights and outputs from the previous one, so that the overall framework moves step by step from exploratory linkage between signals and constructs toward more refined procedures and finally toward individualized estimation. In this way, the sequence of studies does not represent three independent experiments, but rather a cumulative empirical pathway that operationalizes the proposed hybrid psychometric approach.

First, we establish feasibility by mapping associations between passively sensed behavior and learning-motivation measures (association). In this initial stage, the primary goal is to examine whether behavioral traces derived from passive smartphone sensing exhibit systematic, theoretically meaningful relationships with conventional learning-motivation scores, and to identify which broad categories of features show promising links. This provides foundational evidence that the hybrid assessment idea is empirically plausible and that the sensing signals carry information relevant to the targeted motivation constructs, justifying more fine-grained investigation in subsequent stages.

Next, we refine feature sets and procedures to enhance robustness, comparability, and temporal sensitivity (optimization). Here, we iteratively adjust and prune behavioral indicators, aggregation choices, and preprocessing rules so that the resulting feature space is more stable across time, more interpretable, and more easily reused across studies or cohorts. The focus is on improving the reliability of behavior–motivation relations, strengthening alignment with theoretical constructs, and ensuring that the features can capture meaningful temporal changes in learning motivation rather than incidental fluctuations or noise.

Finally, we conduct individual-level, theory-consistent concurrent estimation that links behavioral traces to learning-motivation constructs with interpretable outputs (estimation). In this last stage, the optimized feature space and procedures are used to build models that generate concurrent estimates of learning motivation for each student, while remaining explicitly anchored in the underlying motivational theory.

The emphasis is on producing person-level outputs that can be explained in terms of concrete behavioral patterns and SDT-aligned constructs, thereby demonstrating that the hybrid approach can deliver individualized, interpretable assessment rather than only aggregate-level associations.

- **Association Analysis: Explore the Associations between Daily Behaviors and Learning Motivation in Daily Real-World Settings**
  - **Goal:** *Test the feasibility of parallel mobile sensing and self-report collection, and explore and establish associations between behavioral data and concurrent learning motivation.* Concretely, this involves examining whether students can reliably provide questionnaire responses while a sensing app runs in the background, whether the resulting data streams are of sufficient quality and coverage, and whether the behavioral indicators derived from these streams show systematic, interpretable links with self-reported learning motivation levels measured in the same time periods.
  - **Design:** *A lightweight sensing app iSense is deployed in parallel with periodic online questionnaires.* The sensing streams are preprocessed and segmented into time windows from which behavioral features are extracted; behavioral patterns are then summarized and identified via PCA, and subsequently related to the concurrent self-reported learning motivation levels via association analysis. In practice, this means that raw logs from iSense are transformed into a structured feature set for each time window, dimensionality is reduced to reveal broader behavioral profiles, and these profiles, together with individual features, are statistically examined in relation to the motivation scores collected in overlapping or matching windows.
  - **Outputs:** *An integrated and longitudinal mobile sensing study that continuously assessed learning motivation; preliminary feasibility evidence for using passive smartphone sensing integrated with self-report measures for learning motivation assessment; an initial version of behavioral feature dictionary; guidance for methodological refinement.* Together, these outputs comprise a proof-of-concept demonstration that parallel sensing and self-report collection is workable in a real student population, a first empirically grounded catalog of candidate behavioral indicators of learning motivation, and concrete recommendations on how to improve subsequent study designs, preprocessing steps, and analytic choices in order to build a more robust hybrid assessment pipeline.
- **Behavior Identification Optimization: Refining Behavioral Indicators via Investigating Differences Between Motivation-Level Groups and Within-Person Fluctuations Over Time**
  - **Goal:** *Optimize behavior identification to effectively capture dynamic levels of learning motivation.* In this stage, the focus is on improving how raw sensing signals are transformed

into higher-level behavioral indicators so that they can more sensitively and reliably reflect both stable differences between students and short-term fluctuations in their motivation over time. The optimization aims to make these indicators more robust, theoretically meaningful, and suitable for subsequent modeling and interpretation.

  - **Design:** *Conduct a larger-sample, longer-observation, integrated* longitudinal *mobile-sensing study; refine behavior identification by (i) enhancing feature engineering, (ii) formalizing fusion with explicit rules for time alignment, reliability weighting, and missing-data handling, and (iii) extracting behavioral patterns via clustering.* Compared with the initial exploratory study, this design increases both the number of participants and the duration of observation, providing richer data to examine temporal dynamics. Feature engineering is extended and systematized to derive more nuanced and interpretable descriptors from the raw sensing streams, while fusion is governed by pre-specified rules that make the combination of sensing and self-report data transparent and reproducible. Clustering is then used to discover recurring behavioral profiles or patterns across the feature space. Compare behavioral-feature differences (a) between motivation-level strata (between-person) and (b) within-person fluctuations over time to guide feature selection and refinement.In doing so, only features that show meaningful and stable contrasts across and within students are retained as core indicators.
  - **Outputs:** *A blended, larger-scale, and comprehensive mobile-sensing study enabling continuous assessment of learning motivation; a refined behavioral-features dictionary and behavioral patterns; and evidence of reliability and temporal sensitivity.* Concretely, this stage yields a more mature catalog of behavioral features and clusters that can be reused in later work, together with empirical demonstrations that these indicators behave consistently across time windows while still responding to genuine changes in motivation. The resulting feature dictionary and pattern set, supported by longitudinal analyses, provide a stronger foundation for subsequent individualized estimation of learning motivation in the final stage of the framework.

- **Concurrent Estimation: Learning Motivation Modeling with LLM-Augmented Inference**
  - **Goal:** *Achieve individual-level concurrent estimation of learning motivation and strengthen interpretability via theory-consistent context encoding.* In particular, the aim is to move from group-level associations toward person-specific estimates that are produced in the same time windows as the observed behavior, rather than retrospectively or only at aggregated scales. By embedding Self-Determination Theory (SDT) constructs directly into the way behavioral and contextual information is represented, the goal is not only to predict learning motivation levels for each student, but also to ensure that the resulting estimates and explanations can be interpreted in terms of familiar motivational dimensions and real-world study contexts.

- **Design:** *Leverage LLM-augmented inference to (i) encode contextual data and align behavioral traces with higher-order constructs of learning motivation; (ii) temporally align SDT-aligned self-reports with time-ordered, refined behavioral features; and (iii) textualize the combined dataset to prompt LLMs under zero-shot and few-shot settings with chain-of-thought rationales for individual-level concurrent estimation and inference.* Concretely, contextual and behavioral information is transformed into structured textual descriptions that explicitly reference relevant motivational constructs, and these descriptions are paired with their corresponding self-report labels within fixed analysis windows. The resulting textualized subject–window records are then used as prompts to large language models, which are instructed to produce both point estimates and short, theory-grounded rationales, allowing the models to integrate historical context, current behavior, and SDT-based concepts when inferring each student's momentary learning motivation.
- **Outputs:** *Interpretable mappings from behavioral data to learning-motivation constructs; and evidence of the feasibility of LLM-based concurrent estimation of learning motivation integrated with smartphone-based behavioral sensing.* The study delivers concrete examples of how refined behavioral features and contextual descriptions can be translated into SDT-aligned explanations at the individual level, showing which patterns of smartphone use and study context are linked to higher or lower motivational states. In addition, it provides empirical results on the performance and stability of LLM-augmented estimation under realistic data constraints, thereby demonstrating that this approach can serve as a practical step toward scalable, interpretable hybrid assessment in everyday educational settings.

Across above three-staged studies, the methodology advances in a structured architecture: association (identifying robust associations and establishing feasibility), optimization (refining behavioral indicators, improving temporal sensitivity, and validating fusion), and estimation (individual-level concurrent estimation with LLM-augmented inference and interpretable alignment to psychological theory). This progression completes the realization of the hybrid approach for learning motivation sensing and assessment in real-world educational contexts.

1.5.4 Overall Architectural Framework and Data Flow

Figure 3 given below illustrates the overall data flow through the five-layer architecture that maps the SDT-aligned construct of learning motivation to passively sensed behavioral data in a structured, stepwise manner. It schematically shows how information moves through each layer of the framework, from initial data acquisition and preprocessing to feature representation, analysis, and final reporting, while preserving the link to the underlying motivational constructs at every stage of the pipeline and making explicit which inputs and outputs are handled in each layer. In doing so, the figure provides a compact yet informative visual overview of how theory-grounded measures of learning motivation are operationalized and connected to concrete behavioral traces within the proposed hybrid assessment architecture, helping

readers see how psychological theory, sensing design, and computational procedures are integrated into a single coherent workflow that can guide both implementation and replication in future studies.

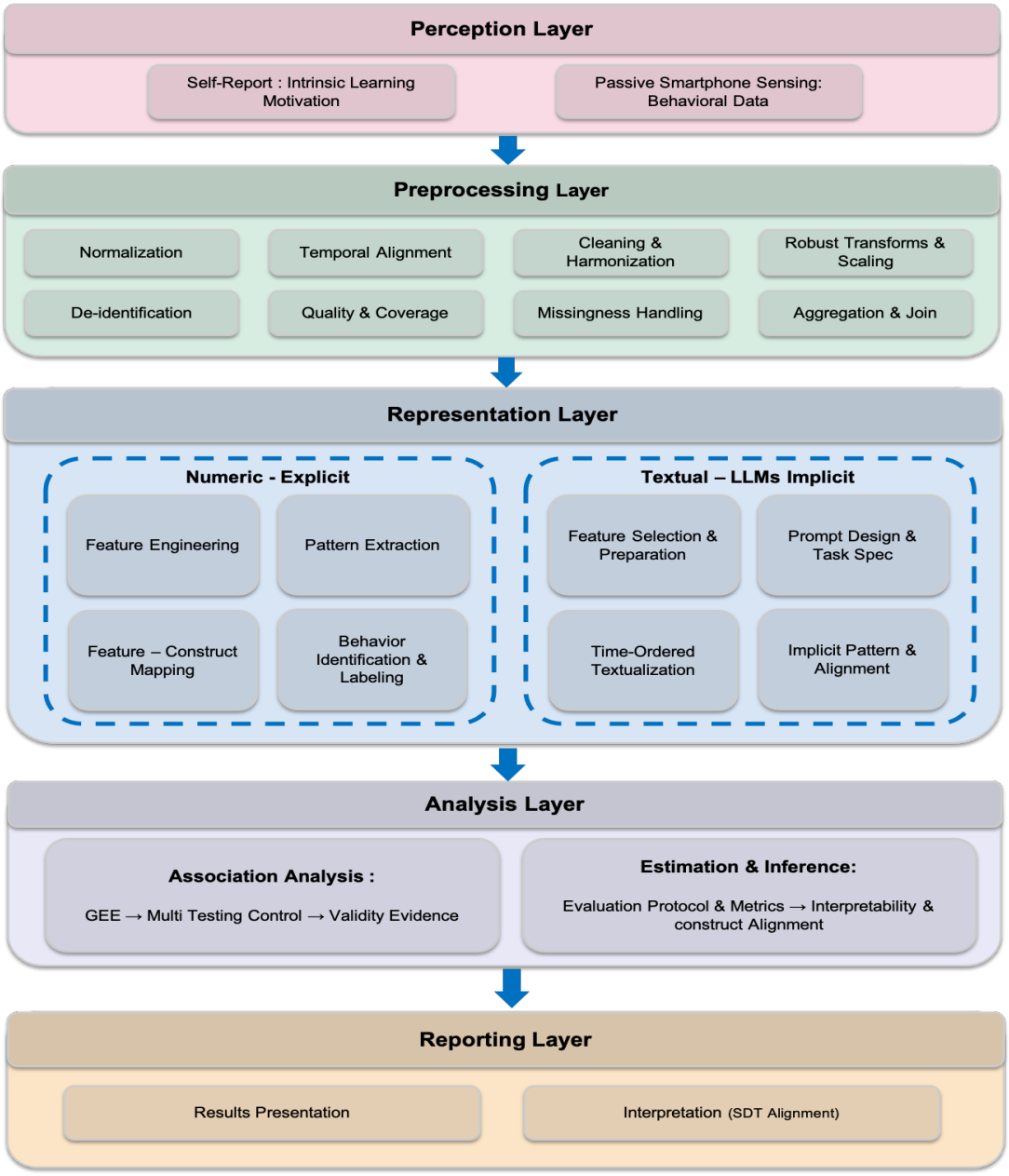


Figure 3: A five-layer architectural framework of the hybrid approach

Layers:

- **Perception Layer - Data Collection of Passive Sensing Data and Self-Report Measures:** This layer collects concurrent input data in real-world educational settings along two streams.
  - **Self-report:** periodic SDT-aligned measures of learning motivation. These measures are administered at pre-defined time points using theory-grounded questionnaires or short scales that operationalize key SDT construct intrinsic learning motivation. They provide a structured,

psychometrically informed summary of students' motivational states, which serves both as a primary outcome for traditional analyses and as a reference signal for validating and calibrating sensing-based indicators.

  - **Passive Smartphone Sensing:** behavioral data and ambient context data. These data streams are captured unobtrusively from students' smartphones, including interaction logs, mobility-related information, app usage patterns, and coarse environmental indicators (e.g., approximate light or noise levels where permitted). All sensing runs in the background during everyday life, yielding ecologically valid traces of study-related routines and contexts that can be temporally aligned with the self-report measures to support hybrid assessment.

- **Preprocessing Layer - Data Intake & Normalization, De-Identification, Temporal Alignment, Quality & Coverage, Cleaning & Coverage, Missingness Handing,Robust Transforms & Scaling:**

  - **Data Cleaning & Normalization:** We ingest passive smartphone logs and psychometric scores and harmonize subject identifiers and time zones for consistent processing. In practice, this means that all raw records from different sources are brought into a single pipeline where each entry is tagged with a unified participant ID and converted to a common time reference, so that subsequent steps can treat the combined data as one coherent, chronologically comparable dataset.

  - **De-Identification:** We de-identify personally identifiable information, restrict access to raw logs. Concretely, direct identifiers are removed or replaced with pseudonymous codes before analysis, and access to the original, unprocessed logs is limited to controlled environments and authorized personnel only, ensuring that all downstream processing operates on privacy-preserving data.

  - **Temporal Alignment:** We anchor analysis to a common, fixed-length analysis window and align sensing windows to their corresponding psychometric measures windows under a pre-specified rule. This provides a uniform temporal structure in which all behavior and questionnaire scores are mapped onto the same type of windows, so that comparisons and associations across participants and days are based on consistently defined time segments.

  - **Quality & Coverage:** We compute per-window coverage metrics and predefine exclusion or weighting rules based on data completeness and device uptime. These coverage indicators quantify how much valid data each window contains and are then used to decide whether a window should be kept, down-weighted, or excluded from analysis, thereby making quality control an explicit and reproducible part of the preprocessing.

  - **Cleaning & Harmonization:** We remove duplicates, correct timestamp ordering and clock

drift, and standardize event schema. These steps reduce technical artifacts and inconsistencies between devices or app versions, ensuring that all events follow a coherent format and time sequence before they are transformed into higher-level features.

  - **Missingness Handling:** We characterize missingness mechanisms, add missing-data indicators, and apply conservative imputation only when methodologically justified. By explicitly labeling where and when data are missing and carefully limiting imputation to clearly motivated cases, we preserve information about gaps in the data while avoiding the introduction of artificial patterns that could bias later analyses.

  - **Scaling:** We standardize features to a common scale. This rescaling removes the influence of differing units and ranges across variables, allowing subsequent statistical models and descriptive analyses to compare and combine features in a stable and numerically well-behaved manner.

  - **Aggregation & Join:** We aggregate events to the analysis window and join sensing and psychometric data into a versioned subject–window dataset. In this final step, all cleaned events are summarized within each window and merged with the corresponding psychometric scores, producing an analysis-ready table where each row represents a given participant in a given window, together with explicit versioning to track preprocessing choices over time.

- **Representation Layer - Feature Engineering, Feature-Psychological Construct Mapping, Behavioral Pattern Identification:** Convert the subject–window dataset into interpretable behavioral representations and produce behavior identification outputs for downstream analysis.

  - **Numeric (explicit):**

    * Feature engineering (behavioral aligned). Compute numeric features with aggregation, proportioning, transformations, and standardization to form a clean feature matrix. Starting from cleaned event-level logs, we summarize raw signals into interpretable behavioral indicators (e.g., counts, rates, ratios, variability measures) and place them in a structured matrix that is directly usable for statistical analysis and modeling.

    * Feature and psychological-construct mapping. Record each feature's target construct, any context gating, and possible lag, creating testable mapping rules. For every engineered feature, we explicitly document which SDT-aligned construct it is intended to reflect, under which contextual conditions it should be interpreted (e.g., only during study days), and whether it is expected to precede or follow changes in motivation, so that these mappings can later be empirically evaluated.

    * Pattern extraction. Association analysis stage: apply PCA on the feature matrix for behavior pattern extraction; *Optimization stage:* apply clustering on features or factor

scores for behavior pattern extraction and refinement. In the first stage, PCA is used to condense correlated features into broader behavior factors, whereas in the optimization stage, clustering is applied to these features or factor scores to reveal recurring groupings of behaviors, allowing us to iteratively refine and stabilize the set of behavior patterns.

* Behavior identification and expert-informed labeling. Combine quantitative criteria with psychology expert interpretation to name factors or clusters and summarize behavior patterns, then formalize labeling rules. Quantitative outputs (e.g., loadings, centroids) are thus linked to expert knowledge to assign meaningful labels and descriptions, and these labels are translated into explicit rules that define how each numeric pattern is identified in subsequent analyses. In this way, data-driven factors or clusters are not treated as purely statistical artifacts, but become interpretable behavioral categories whose identification is governed by transparent, reusable criteria that can be consistently applied across time, samples, and related studies.

– **Textual (LLMs implicit):**

* Feature selection and preparation. Use the refined behavior features and preserve their temporal order. We first select the subset of numeric indicators that have been validated in the explicit pipeline and then arrange them chronologically, ensuring that the sequence fed to the LLM reflects the actual evolution of behavior over time.

* Textualization with time ordered representation. Convert features into clear English sentences that include sensor, value, and start and end time, maintaining a time ordered sequence. Each window or event is rendered as a concise, human-readable description that encodes which sensor was involved, what value was observed, and over which time span it occurred, producing a temporally ordered "behavioral narrative" for the model.

* Prompt design and task specification. Define instructions and rating scales and configure zero shot or few shot prompts, optionally requiring CoT rationales. In this step, we specify exactly what the LLM should output (e.g., ratings of learning motivation), how those outputs are scaled, and whether example cases or chain-of-thought explanations are provided, so that the model's task and response format are clearly constrained.

* Implicit pattern extraction and alignment.Let the LLMs learn representations and patterns implicitly from the textualized behavioral data without imposing hand-crafted scoring rules at the outset, and then use a separate, post hoc coding step to align the resulting outputs with psychological constructs. The LLM processes the textualized sequences and internally derives latent representations of behavior and its relation to

motivation; afterwards, we systematically code and interpret its outputs using a theory-informed scheme so that predicted scores, rationales, and qualitative attributions can be mapped back onto the predefined motivational constructs and directly compared with explicit numeric indicators such as scale scores and engineered behavioral features.

- **Analysis Layer - Association Analysis, Estimation and Inference:** This layer links behavioral representations to psychometric measures for association and validation and conducts estimation with inferences and evaluation.
  - **Association Analysis:**
    - Modeling approach (GEE-based association). Regress the psychometric measures on factor scores or cluster labels (or features) with a within–between decomposition. In other words, we model learning-motivation scores as outcomes and include behavior-derived variables as predictors, explicitly separating between-person differences from within-person fluctuations over time. Estimate population-averaged effects via Generalized Estimating Equations (GEE) [67, 68], so that the resulting coefficients reflect average associations at the population level while accounting for the correlated, longitudinal structure of repeated measures.
    - Multiple testing control. Apply the Two-Stage Benjamini–Hochberg (TSBH) [69] procedure to control FDR within defined test families and report adjusted q values. Concretely, association tests are grouped into coherent families, the adaptive step-up procedure is used to limit the expected proportion of false discoveries, and all inferential summaries are presented using these q values rather than raw p values, ensuring that inferences respect the multiple-comparison structure.
    - Validity evidence. Summarize convergent and discriminant validity, report reliability where applicable, and consider equivalence tests for practically null effects. This involves organizing results so that associations consistent with theoretical expectations are highlighted, potential overlaps between constructs are examined, internal consistency metrics are documented where relevant, and formal tests are used to support claims that some effects are small enough to be considered practically negligible rather than merely non-significant.
  - **Estimation and Inference:**
    - Evaluation protocol and metrics. Use held-out evaluation with subject-wise or window-wise splits and repeat random splits to reduce sampling variance; summarize performance with MAE as the primary metric and compare results across different shot settings. That is, models are trained and tested on disjoint subsets of subjects or time windows, the splitting is repeated multiple times to obtain stable estimates, and MAE

is used as the central indicator of predictive performance when contrasting various prompting or data conditions.

  * Interpretability and Psychological construct alignment. Analyze CoT rationales using post hoc coding to summarize behavior–motivation patterns and provide construct-alignment evidence. After obtaining CoT-based explanations from the models, these textual rationales are systematically coded, grouped, and interpreted so that recurring patterns can be linked back to theoretical constructs, thereby offering qualitative support that the estimated relationships between behavior and learning motivation are consistent with the intended psychological framework.

- **Reporting Layer - Reports and Explanations:** We report the results in clear text, tables and figures, and explain how they address the study questions.
  - Results Presentation. We present the behavioral, cluster results, behavioral patterns, and GEE coefficients with confidence intervals and q values, and LLMs performance measured by MAE with example outputs. In doing so, we report both the numerical summaries and the corresponding uncertainty ranges, together with illustrative LLM predictions that show how estimated scores relate to concrete behavioral sequences. We visualize the results using variance plots, coefficient plots, and selected CoT explanations, so that the variability of features, the direction and magnitude of estimated effects, and the qualitative reasoning captured in CoT can be inspected in a clear and integrated way.
  - Interpretation (SDT alignment). We interpret associations and inferred scores (estimation) within the Self-Determination Theory framework, linking observed behavioral patterns to the relevant psychological constructs. In this interpretive step, effect directions, strengths, and exemplar patterns are read through the lens of SDT, so that changes in behavior can be meaningfully related to shifts in learning motivation, and the estimated scores are understood as reflections of the underlying motivational processes rather than purely statistical outputs.

Data Flow:

- Perception layer: SDT-aligned self-reports and passive smartphone logs (behavior and context) are collected in parallel. That is, theory-grounded questionnaires are administered at scheduled intervals while, at the same time, background sensing on students' smartphones continuously records high-level behavioral and contextual signals, ensuring that both modalities cover the same real-world learning periods.
- Preprocessing layer: Data are de-identified, temporally aligned to a fixed analysis window, checked for quality and coverage, cleaned, handled for missingness, scaled, and aggregated, building a standardized subject-window dataset. In this step, all raw inputs are transformed into a privacy-preserving, quality-controlled, and consistently structured table in which each row corresponds to

a given participant in a given analysis window.

- Representation layer: Numeric route: behavioral features with an explicit feature to construct mapping and PCA/clustering for behavior identification, so that high dimensional logs are reduced to interpretable patterns grounded in psychological theory. Textual route: time ordered textualization for LLM prompting, in which the same behavioral information is expressed as natural language descriptions that preserve temporal structure. Concretely, the numeric branch encodes behavior as explicit feature vectors that are linked to psychological constructs and then summarized into higher level patterns via PCA and clustering, providing compact, human interpretable indicators. In parallel, the textual branch converts the same temporally ordered information into readable sentences that describe routines and intensities over time, and these sentences can be directly processed by LLMs to learn latent representations and generate predictions from purely text based input.
- Analysis layer: Numeric path uses GEE association models with TSBH control of FDR to estimate longitudinal relationships between behavioral indicators and learning motivation in a statistically controlled and transparent way, so that population level associations are quantified while properly accounting for multiple testing. Textual path relies on LLM augmented concurrent estimation that is evaluated by MAE and interpreted via CoT, so the model produces both numerical predictions and accompanying explanations for each observation. Thus, the explicit route quantifies population level associations under multiple testing control and provides interpretable effect summaries, whereas the implicit route uses LLMs to generate concurrent motivational estimates whose accuracy is summarized by MAE and whose reasoning is examined through chain of thought explanations that can be inspected for psychological plausibility and alignment with SDT constructs.
- Reporting layer: Results are presented in text, tables, and figures, with SDT aligned interpretation that explicitly maps behavioral patterns to psychological constructs and provides reproducible outputs that can be checked and extended by other researchers. Final reports therefore integrate numerical summaries that quantify effects, visualizations that highlight trends and group differences, and theory based explanations that situate the findings within the broader motivation literature, accompanied by sufficient detail on procedures, parameter settings, and generated outputs to support replication, secondary analysis, and re use in future studies or applied systems.

1.6 Research Contributions

Smartphone sensing technologies offer promising opportunities for unobtrusively and continuously assessing students' learning motivation, opening new avenues for educational psychology and for non-disruptive, routine-preserving assessment in everyday study contexts. Building on this premise, this dissertation contributes a hybrid, SDT-grounded approach that links theory-based measures with passively sensed behavioral features and, where appropriate, with LLM-augmented estimation. The contributions are articulated as follows.

- **Self-Determination Theory: Theoretical Framing and Ecologically Valid Feedback.** We map SDT-based intrinsic learning motivation onto interpretable behavioral patterns derived from passive smartphone sensing, demonstrating learning motivation as periodically assessable in real-world contexts while accommodating short-term fluctuations. In addition, the hybrid assessment approach provides ecologically valid, time-resolved evidence that complements expert-based, episodic observation. This additional perspective helps refine the construct's nomological placement[2], and validate SDT theoretical models using empirical behavioral evidence.
- **Methodology: An Innovative Hybrid Psychometric Approach and An Operationalizable Framework.** We propose an innovative hybrid psychometric approach that integrates conventional psychological self-reports with passive smartphone sensing. The approach provides a systematic methodological pipeline spanning data collection, preprocessing, behavioral feature extraction/identification, association analysis, and construct modeling with concurrent estimation. Methodologically, we establish behavior–motivation linkages and refine SDT-aligned behavioral indicators that map onto (intrinsic) learning motivation. Then, we incorporate LLM-augmented inference to enable individualized predictive modeling and continuous, concurrent estimation of motivation levels from sensor traces. We also provide an feasible framework that operationalizes this hybrid approach in practice and is reusable and extensible to the continuous assessment of other SDT constructs. Taken together, these elements form a transparent, reproducible workflow that enhances behavioral granularity, temporal sensitivity, and interpretability in the unobtrusive, continuous assessment of students' learning motivation.
- **Integrated Mobile Sensing for Learning Motivation: An Initial Longitudinal Study Serie.** To the best of our knowledge, this is the first comprehensive and interdisciplinary mobile sensing study integrating passive smartphone sensing and conventional psychometric methods to assess SDT-grounded learning motivation. This longitudinal mobile sensing program consists of two sequential sub-studies. Association Analysis Study quantifies links between objectively sensed social-behavioral patterns and SDT-based learning motivation, and identifies initial set of behavioral patterns that show meaningful associations with learning motivation. Doing so, it furnishes interpretable, theory-consistent data and provides baseline benchmarks for assessing motivation from mobile sensing. Optimization Study refines these behavioral indicators, and derives more precise mappings between sensed behaviors and motivation via clustering. And it also compares trajectories across motivation strata, revealing systematic differences in sleep regularity and smartphone use. Taken together, this series of mobile-sensing studies provides a practical template for the design and implementation of hybrid smartphone-sensing experiments and lays the groundwork for future modeling and estimation research on learning motivation.

[2]Using *intrinsic learning motivation* as an example, "nomological placement" refers to the construct's position in the broader theory network—its expected relations to antecedents (e.g., basic psychological need satisfaction), outcomes (e.g., behavioral engagement, persistence, achievement), and related/distinct constructs (e.g., interest)—used to establish construct validity in real learning contexts.

- **Practice: Guidelines for Unobtrusive, Privacy-Preserving Use.** We distill actionable guidelines for educational settings based on: (1) which behavioral data are informative for learning motivation assessment, (2) how continuous assessment can remain unobtrusive and privacy-preserving, and (3) where the approach is not appropriate. These guidelines outline implications for light-touch continuous perception and formative assessment and emphasizes boundaries that avoid high-stakes or punitive uses.

Collectively, these contributions provide a comprehensive methodological paradigm to continuously assess learning motivation, which combines passive smartphone sensing with advanced analytics and theoretical grounding from educational psychology. Our research considerably advances the understanding and assessment of learning motivation in real-world educational settings, highlighting the transformative potential of unobtrusive, continuous, and context-aware assessment methodologies beyond traditional methods. This work specifies how psychological theory, mobile sensing, and modeling can be integrated to move from association to optimization to learning motivation concurrent estimation, while preserving privacy safeguards and clarifying limits of use. Subsequent chapters operationalize and evaluate these elements in detail.

1.7 Human Subjects' Protection

The human subjects' protection is a major concern of our work. This section details the human-subjects safeguards for learning motivation sensing and assessment, including our data privacy and security. The whole research is approved by the Institutional Review Board (IRB) of the conducting institution Jilin University in China. Since the work involves passive smartphone sensing with student participants, ethics is primarily treated as a design constraint.

- **Approvals and Support:** All procedures strictly comply with Jilin University's IRB requirements and with all applicable data-protection regulations governing human-subject research throughout the entire study. Our studies are supported by the National Natural Science Foundation of China (No.62077027), the Education Department of Jilin Province, China (No.JJKH20200993K), the Ministry of Education of the People's Republic of China (No.2021180010), n nthe National Natural Science Foundation of China (No.62476111), the Department of Science and Technology of Jilin Province, China (No.20230201086GX), and the Humanity and Social Science Youth Foundation of Ministry of Education of China (No.20YJC190034).
- **Consent and Rights:** Participation is voluntary and based on explicit consent after clear explanations of what is collected and why, the risks/benefits, and withdrawal options. Participants were recruited on a compensated basis for their time and were expressly informed about the potential risks; compensation was not contingent on performance or data patterns, and participation could withdraw the study at any time without penalty. Participants may request access, correction, or deletion as permitted by policy. Retention is limited to the minimum necessary for analysis and

verification, with documented, auditable deletion procedures. Participants were free to discontinue participation and withdraw from the study at any time, and their data will be deleted within two weeks after confirmation of their withdraw. And we limit data retention to 6 months for raw sensing data.

- **Security and Access Control:** Data are encrypted in transit and at rest and stored on privately managed, non-public servers with role-based access control and audit logging. We purchased a professional high-privacy private-cloud Aliyun server from Alibaba Cloud. Each login requires multi-factor authentication with real-name identity verification (government-issued ID) and live facial verification, and sessions are protected by strict access controls. Access is strictly limited to authorized researchers for approved purposes, if any third-party attempts to log in, we receive immediate phone call of notification then conduct a detailed authorization check. During the data-collection period, we download collected data every day to local storage and create offline backups on dedicated hard drives (retained exclusively by the research team), thereby reducing any risk of exposure of participants' data. Any external sharing requires governance approval.
- **Data Minimization and De-Identification:** Only variables essential to the research aims are collected; passive sensing focuses on behavioral data (e.g., usage of different applications: only the usage timestamp, duration and application name are collected). We collect no message/photo/-text content or any within-app content, and location uses low precision where possible. Given that participants' daily living and learning activities were mostly on campus, we assigned each location with a functional label (e.g., library, classroom). Personal identifiers are not collected, and datasets are irreversibly anonymized at source with no re-identification keys retained; analyses use anonymous study codes that cannot be mapped to individuals. Public reporting relies on aggregated or de-identified statistics.

1.8 Structure of the Dissertation

This section outlines how this dissertation advances the hybrid psychometric approach integrating self-reports and passive smartphone sensing for continuous assessment of learning motivation. It provides a brief roadmap of the dissertation, explaining how the subsequent chapters implement, refine, and evaluate this hybrid approach, from initial association analyses through optimization of behavioral indicators to individualized, context-aware estimation.

- **Chapter 1 - Introduction.** Articulates the research aims, defines research scope, boundaries and limitations, states research problem and formulates research questions and specify research objectives, specify the research scope, boundaries and limitations, presents the methodological overview, summarizes contributions, and finally notes the human subject protections and presents the structure of this dissertation.
- **Chapter 2 - Background and Related Work.** Reviews learning motivation under SDT and

conventional assessments in a structured manner, outlining core constructs, typical measurement tools, and how learning motivation is usually operationalized in existing work. Surveys passive smartphone sensing for learning motivation and its relevant constructs, mapping which behavioral indicators and modalities have been used, how they relate to SDT informed variables, and where current approaches succeed or fall short in ecological validity and scalability. Systematically analyzes and synthesizes these research gaps into a coherent study trajectory, specifying which questions remain open, how the subsequent studies in this dissertation address them step by step, and how the overall program moves from conceptual clarification to methodological development and empirical validation.

- **Chapter 3 - Passive Smartphone Sensing Design, Data Collection, and Reprocessing.** Articulates the shared technical core of this dissertation. It specifies the passive smartphone sensing design, the parallel collection of passive smartphone sensing data and psychometric scales, and a common data-collection protocol. Then it details preprocessing principles from temporal alignment, de-duplication, missing-data handling, normalization, and leakage prevention, as well as family-level behavioral feature construction. Unless explicitly stated in the study, all subsequent chapters inherit the core shared procedures defined here.
- **Chapter 4 - Association Analysis of Daily Behaviors and Intrinsic Learning Motivation.** Employ passive smartphone sensing and periodic self-reports in parallel, extracts behavioral features and identifies behavioral patterns, and conducts association analysis with learning motivation. Provides feasibility evidence and delivers an initial set of behavioral feature dictionary and behavioral patterns.
- **Chapter 5 - Optimizing Behavior Identification by Refining Indicators from Behavioral Changes.** Optimizes behavior identification, enhances feature engineering and behavioral pattern extraction with clustering. Conducts comparative analyses of behavioral feature differences at the between-group and within-person levels to derive refined behavioral indicators aligned with learning motivation.
- **Chapter 6 - Concurrent Estimation of Learning Motivation with LLM-Augmented Modeling.** Leverages LLM-augmented inference to encode contextual data and to align behavioral data with higher-order constructs of learning motivation. SDT-aligned self-report measures are temporally aligned with time-ordered refined behavioral features; the combines dataset are textualized and used to prompt LLMs under zero-shot and few-shot settings with Chain-of-Thought rationales for individual-level concurrent estimation and inference of learning motivation.
- **Chapter 7 - Conclusion and Future Work.** Summarizes the dissertation as a whole, synthesizes the key findings with respect to each research question and situates the main contributions

across the different empirical studies, and analyzes the limitations in a structured way while underscoring how this dissertation outlines concrete future directions in learning motivation assessment and learning analytics. It ultimately closes by answering the research questions explicitly, drawing together evidence from all chapters, and articulating the broader implications for both research and educational practice, including how the proposed approaches can inform future methodological work and real world deployment.

1.9 Summary

This chapter introduces the research background and problem statement, formulates the leading research problem, and states the concrete research questions and objectives. It outlines the research scope, boundaries, and limitations, and summarizes the study's contributions. The chapter then presents a methodological overview along with the overall research design and the architectural framework that operationalizes the proposed hybrid approach. Most importantly, we prioritize participants' rights and privacy ethical considerations and human-subjects protections guides our entire research design and are thoroughly described. Finally, the chapter concludes by outlining the structure of the dissertation. Taken together, the chapter provides an integrated overview of the research and lays the theoretical and practical foundations for the subsequent chapters.

1.10 List of Publication

The work developed in this dissertation has directly and indirectly contributed to the following publications:

Y. Feng, T. Zhao, J. Li, Y. Zu, A. Tavares, T. Gomes, and H. Xu, "Social behavioral sensing: An exploratory study to assess learning motivation and perceived relatedness of university students using mobile sensing," in Proc. 45th Annu. Meeting Cogn. Sci. Soc. (CogSci), vol. 45, no. 45, 2023.

Y. Feng, T. Zhao, Y. Zhang, Y. Zu, A. Tavares, T. Gomes, and H. Xu, "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 Proc. 46th Annu. Meeting Cogn. Sci. Soc. (CogSci), vol. 46, 2024.

Y. Feng, T. Zhao, Y. Zu, A. Tavares, T. Gomes, and H. Xu, "An explorative investigation into leveraging LLMs to predict university students' learning motivation," in Proc. 47th Annu. Meeting Cogn. Sci. Soc. (CogSci), vol. 47, 2025

2

Background and Related Work

This chapter lays the conceptual and methodological groundwork for continuous assessment of learning motivation in higher education. We first review Self-Determination Theory, describing the motivational continuum and regulation spectrum, and explain why intrinsic motivation is adopted as the primary target construct, with attention to its temporal dynamics across micro, meso, and macro timescales. Building on this, we develop a conceptual bridge from SDT constructs to observable behavioral indicators, clarifying how sensed behaviors can be organized into feature families and mapped to motivation levels. On the technical side, we introduce passive smartphone sensing and digital phenotyping, outline an operational flow for transforming raw logs into behavioral features and unsupervised behavior patterns, and discuss privacy, compliance, and data-quality considerations. We then present LLM-augmented modeling as a complementary route to traditional machine learning for estimating intrinsic motivation from time-aligned behavioral summaries. Finally, we review related work on traditional assessment methods and smartphone-based approaches, compare methods in a structured way, and identify key research gaps that directly motivate the hybrid, SDT-aligned assessment framework developed in the remainder of this dissertation.

2.1 Theoretical Background

This section develops the conceptual foundation for using SDT to guide continuous assessment of learning motivation with passive sensing. It first revisits the SDT motivational continuum and the internalization process, clarifying how forms of regulation range from external control to intrinsic motivation and why intrinsic motivation is adopted as the primary target construct for this dissertation. It then examines the temporal dynamics of intrinsic learning motivation by distinguishing stable, trait-like baselines from state-level fluctuations and by describing how motivation unfolds across micro, meso, and macro timescales in everyday academic life, together with the implications for person-centered, continuous assessment. Finally, it builds a conceptual bridge to sensing by specifying observable behavioral feature families, outlining how time-aligned behavioral data and motivational measures can be mapped onto each other, and defining validity checks that ensure SDT-consistent, interpretable mapping rules for subsequent analyses.

2.1.1 Motivational Continuum, Internalization, and Assessment Implications

This subsection clarifies how SDT conceptualizes learning motivation along a continuum from amotivation to intrinsic motivation and how different forms of external, introjected, identified, and integrated regulation fit within this spectrum. It outlines how these regulatory styles differ in autonomy, volition, and subjective experience in everyday learning contexts, and how they can coexist and fluctuate within the same student. Building on this continuum, the subsection then explains why intrinsic motivation is adopted as the primary target construct in this dissertation, emphasizing its theoretical centrality, ecological relevance for daily academic life, robust links to key outcomes, and methodological advantages as a clear scalar anchor for the hybrid assessment pipeline.

Motivational Continuum: Within SDT, learning motivation is organized along a continuum from amotivation to intrinsic motivation [6, 9, 70]. Originally articulated by Deci and Ryan and refined over subsequent work, the continuum places a set of extrinsically regulated forms between these endpoints: external regulation, introjected regulation, identified regulation, and integrated regulatio [6, 71]. Intrinsic motivation sits at the rightmost end and refers to engaging in an activity out of interest, curiosity, or challenge, accompanied by experiences of autonomy and competence [72, 73]. The continuum captures a gradation in autonomy/volition and clarifies that multiple forms can co-exist and fluctuate within the same person across contexts [6, 71, 74]. In academic settings, intrinsic motivation represents the "ideal endpoint", typically associated with voluntary engagement, deeper processing, persistence, and well-being [6, 75, 76]. Accordingly, we treat intrinsic motivation as the core construct and primary outcome for characterizing learning motivation, and we use it as an anchor for interpreting dynamics in other regulatory forms.

Regulation Spectrum: External → Internal (learning-context rendering): In SDT, the regulation spectrum describes how extrinsic motivation becomes progressively internalized—from external to introjected to identified to integrated regulation—eventually approaching intrinsic motivation [9]. This pathway was first outlined by Deci and Ryan and systematically developed under Organismic Integration Theory (OIT) [6]. It reflects a stepwise increase in autonomy and volition, while acknowledging that different forms can co-occur and vary within learners' everyday study routines.

- *External Regulation:* Behavior is controlled by external consequences (rewards, punishments, grading, attendance rules, others' expectations) [77]; the perceived locus of causality is outside the self [71]. In practice, compliance is contingent on the presence of external prompts and sanctions, so effort often rises when monitoring is salient and falls when it is absent. Learners commonly report pressure, tension, and a focus on "meeting requirements" rather than personal interest. *Learning context:* attending class to avoid attendance penalties, drilling problem sets to hit a scholarship threshold, last-minute cramming to avert failing [2, 78]. Additional examples include completing assignments solely to satisfy minimum grade criteria or responding to frequent reminder messages to avoid penalties. *Typical cues:* strong dependence on reminders/surveillance; performance tracks

the intensity of external pressure and drops when pressure eases [79–81]. In day-to-day patterns, behavior may spike around deadlines and then quickly decay, with limited carryover once the external contingency is removed.

- *Introjected Regulation:* External demands are "swallowed" into the self and enacted to avoid guilt/shame to protect self-worth [82–84]; the control is internalized but still felt as pressure [85]. Students work to satisfy an internal "should," and engagement often feels compelled rather than chosen. Effort can be intense but brittle, with bursts of activity followed by fatigue or self-criticism. *Learning context:* studying to avoid embarrassment or disapproval from peers/instructors; high-pressure bursts driven by a self-imposed "I must not slack" rule [78, 86]. Learners may postpone until anxiety rises and then overcorrect with short sprints of high effort. *Typical cues:* anxiety-driven effort, perfectionism; the dominant feeling post-completion is relief rather than satisfaction or interest [87, 88]. After tasks, students often report reduced tension but little curiosity, and the cycle can repeat when self-worth is again perceived as at risk.

- *Identified Regulation:* The learner endorses the personal value or goal relevance of the task; action is more voluntary and purposeful [2]. Value acceptance supports flexible persistence, planned practice, and strategic allocation of effort across weeks. It remains sensitive to context but is less contingent on external prompts than controlled forms. *Learning context:* recognizing that statistics or programming is essential for the thesis or internship, then planning practice and breaking goals into milestones [89, 90]. Students typically track progress, schedule study blocks, and revise plans when obstacles arise. *Typical cues:* clear articulation of "why this matters to me," emergence of self-set milestones and forward-looking time/resource allocation [91]. Day-to-day, learners show steady engagement tied to self-endorsed aims and a willingness to adjust strategies to maintain progress.

- *Integrated Regulation:* Task value is integrated into one's identity and long-term aspirations, consistent with other values; choices feel self-congruent [6]. Actions align with a broader narrative of "who I am," and decisions across courses and projects tend to reinforce that identity. Engagement generalizes across settings and remains stable even when demands fluctuate. *Learning context:* acting from an identity such as "I am a data scientist/educational innovator," which guides autonomous course selection, cross-course transfer, and sustained project work [92, 93]. Learners proactively seek experiences that fit their trajectory and integrate new skills into existing commitments. *Typical cues:* narrative consistency between learning choices and identity; natural cross-course connections; self-directed, stable time management. Over time, patterns show durable routines, long-horizon planning, and coherent links between daily tasks and long-term aims.

- *Intrinsic Motivation:* Behavior is energized by interest, curiosity, and challenge, with ongoing experiences of autonomy and competence (a qualitative shift at the end of internalization) [6, 73]. Activities are pursued for their inherent appeal, and engagement is sustained without external

prompts. Learners often experiment, explore alternatives, and treat difficulties as interesting problems to solve. *Learning context:* iterating on a model out of curiosity, searching literature to test an idea, experiencing "flow" while experimenting/coding/deriving [94, 95]. Sessions may extend naturally when insight emerges, and attention stays focused without supervision. *Typical cues:* engagement is driven by the inherent appeal of the activity; supervision is unnecessary to maintain focus and deep processing [75, 80]. Learners report satisfaction and interest after completion, and they readily re-engage in similar tasks because the activity itself is rewarding.

Internalization is gradual rather than a jump. Each phase has characteristic psychological markers (locus of control, value endorsement, self-congruence) and subjective experiences (tension/guilt → purpose/volition → interest/absorption). Environments that support autonomy, competence, and relatedness facilitate movement along this pathway [96, 97]. Intrinsic motivation is therefore the most autonomous form and the meaningful endpoint for "why the student learns," making it a natural focal variable.

Motivation for Adopting Intrinsic Motivation as the Target Construct: Grounded in the SDT motivational continuum and the regulation spectrum, we select intrinsic motivation as the primary scalar for assessing learning motivation. This choice reflects theoretical clarity, fit with everyday academic life, and practical value for analysis and feedback. Below we state five complementary reasons and explain how each reason supports our design and our hybrid pipeline.

- *Construct's Centrality:* Intrinsic motivation directly reflects the inherent value of an activity and captures interest, enjoyment, and voluntary engagement. It is closely linked to curiosity, deep processing, and sustained effort without the need for external control. Because it sits at the autonomous end of the continuum, it offers a clear, theory-based anchor for interpreting how students invest time and attention in study-supportive contexts.
- *Reference Point for Interpretation:* Treating intrinsic motivation as a scalar gives a common reference point for reading changes in other forms of regulation. In practice, this means we can interpret movement along the path from control to autonomy and summarize within-person deviations around a personal baseline. The scalar target also supports simple aggregation and clear textualization, which then helps the LLM align examples, scales, and rationales in a consistent way.
- *Ecological Validity:* Within-person variation in intrinsic motivation is sensitive to everyday academic cues such as task clarity, perceived difficulty, competence feedback, and social support. These cues change over days and weeks, so intrinsic motivation can be monitored at a time scale that matches passive sensing and low-burden self-reports. The construct varies enough to be informative, yet remains interpretable for students and instructors in authentic settings.
- *Links to Outcomes:* Intrinsic motivation shows reliable associations with conceptual understanding, persistence, self-regulated learning, and well-being. These links increase the predictive value of the target and make any estimate actionable for formative feedback. By focusing on a construct

with established outcomes, we improve the chances that changes in the estimate will align with meaningful academic behavior and support timely, low-stakes interventions.

- *Method Benefits:* Centering intrinsic motivation provides a clear target for the full workflow, from feature design and textual representation to LLM-augmented inference and reporting. It simplifies feature naming and directionality, supports participant-level interpretability, and enables transparent evaluation with calibration and error checks. Because the construct is well defined and theory-aligned, it also travels better across cohorts and sites, which helps with reuse, comparison, and future cross-cultural validation.

2.1.2 Temporal Dynamics of Intrinsic Learning Motivation

This subsection focuses on the temporal dynamics of intrinsic learning motivation and how they matter for assessment design. From a SDT perspective, intrinsic motivation combines a relatively stable, trait-like baseline with systematic within-person fluctuations driven by need-supportive or need-frustrating contexts. These fluctuations unfold across multiple time scales, from momentary classroom episodes to weekly course rhythms and semester-long trajectories, and they are mirrored in observable learning behaviors. On this basis, the subsection motivates a person-centered, continuous assessment strategy that estimates individualized baselines, tracks deviations over time, and time-aligns behavior and motivation, treating them as mutually influencing processes rather than as static, one-directional links.

- *Within-Person Variability & Trait Baseline:* From an SDT perspective, intrinsic learning motivation has both trait-like stability and state-level variability [71, 74]. The trait-like part reflects a learner's degree of internalization and the long-term satisfaction of the three basic psychological needs (autonomy, competence, and relatedness) [96, 97]. Over years of socialization and study, these experiences build a relatively stable individual baseline: some students tend to engage voluntarily and persist even with limited prompts, while others show lower spontaneous engagement [98–100]. This baseline summarizes a person's typical quality of regulation across tasks and contexts. At the same time, intrinsic motivation varies within the same person over time around that baseline [2, 101, 102]. These short-term changes are driven by need-supportive or need-frustrating cues in the immediate context [96, 103]. Autonomy support (choice, meaningful rationale, voice) [104–107], competence support (clear structure, optimally challenging tasks, informative rather than controlling feedback) [108–110], and relatedness support (respect, care, belonging) [6, 111–113] tend to raise intrinsic motivation in the short run. In contrast, control signals (threat, pressure, contingent regard), competence threats (unclear goals, repeated failure without scaffolding), and social exclusion cues tend to lower it [113–115]. In practice, a student can move from controlled effort to willing engagement within a single afternoon if a teacher reframes an assignment with choice, clarifies expectations, and provides timely, informational feedback [108, 116, 117]. This two-level view, a stable orientation grounded in internalization plus context-driven deviations, helps explain both between-person differences (different baselines) and within-person fluctuations (ups

and downs around one's own baseline) [118]. For assessment, this implies that raw scores are harder to compare across persons than changes relative to each person's baseline. A small rise for a low-baseline learner may indicate strong contextual support; a similar absolute score for a high-baseline learner may indicate a dip [119–121]. SDT therefore promotes person-centered, continuous assessment that tracks deviations from the individualized baseline rather than only group norms. In short, intrinsic motivation is not random noise; it is structured variation around a meaningful set point shaped by internalization and by ongoing need support in the learning environment.

- *Timescales & Rhythms:* Within-person fluctuations in intrinsic motivation unfold across multiple timescales and show rhythmic patterns [2, 102, 122, 123], and these patterns map onto observable learning behaviors [75, 108, 124, 125].
  - *Micro Scale (minutes to hours):* Need-supportive events in class—choice over methods, a clear scaffold for a difficult problem, or informative feedback on an attempt—can produce immediate gains in willing engagement [104, 105, 126, 127]. Behaviorally, this appears as faster task start, longer time-on-task, fewer app switches and distractions, more in-class questions and help-seeking, and more attempts at higher-difficulty items [128–131]. Digital traces include Learning Management System (LMS) dwell time and page depth, steadier quiz pacing, keystroke/edit bursts that reflect focused work, and Integrated Development Environment (IDE) activity with fewer idle gaps [132–134]. Conversely, controlling cues or competence threats can lead to slower starts [135, 136], more task switching, and avoidance of challenging items [137–139].
  - *Meso Scale (days to weeks):* Course structure and collaboration cycles shape regular highs and lows [140–142]. Assignments, labs, and group meetings create a weekly rhythm that is visible in study-window regularity and density [143–145], timely versus last-minute submissions (early submissions versus deadline spikes) [146, 147], attendance stability [148–151], and peer interaction rates (forum posts, repository commits, shared document edits, and platform interactions) [152, 153]. Need support at this scale often comes from predictable structure, transparent criteria, and reliable peer support [154, 155]; need thwarting comes from volatile scheduling, unclear rubrics, or group dysfunction [156].
  - *Macro Scale (months to a semester):* Gradual changes in value endorsement and integration can be observed [127, 157]. When students find coherence between tasks and their goals or identity, they tend to make more autonomous choices of courses and projects [93, 158, 159], engage in spontaneous cross-course transfer of skills [160–162], show greater persistence on long tasks [163–165], and stabilize their self-regulation rhythms (more regular planning and reflection) [166, 167]. In exam periods, activity volume can rise (more hours logged, more submissions), but autonomy and competence may fall if pressure and evaluative threat dominate, which could produce a "high quantity, lower quality"pattern typical of controlled

regulation [114, 168–170].

Importantly, behavior can lead or follow motivation [124, 171–173]. A single high-quality practice session may raise later competence feelings and interest, whereas controlled last-minute studying can produce high activity with low intrinsic motivation [174–177]. Analyses should therefore allow synchronous and associated relations between motivation and behavior. For modeling, this means aligning measures within the same time windows and also testing lead–follow structures (e.g., behavior at t estimating motivation at t, and vice versa).

- *Design Implications for Continuous Assessment:* Given the trait baseline plus state fluctuations and the multi-scale rhythms, continuous assessment of intrinsic motivation is feasible and informative. Beyond group-level assessment, use a person-centered design: estimate an individualized baseline and use it as a reference point. Track the direction and magnitude of deviations from that baseline rather than only absolute scores. For sampling, align high-frequency, low-burden assessment with daily behavior patterns. Examples include: the distribution and density of daily study windows, focus duration and micro-breaks, use rhythms of learning apps/platforms, within-day and across-day submission patterns for assignments and practice, entries into library/class/self-study spaces, and peer collaboration or help-seeking events. Use continuous assessment so that data cover micro (event/hour), meso (day/week), and macro (month/semester) levels. Then time-align and jointly analyze intrinsic motivation and behavior (e.g., test synchronous and lagged associations within the same or sliding windows). In modeling and reporting, separate effects at the within-person, between-person, and between-group levels (e.g., section, course, major). For interpretation, link short-term increases or decreases in motivation to subsequent changes in observable learning or daily behavior patterns. After a rise in motivation, one may see faster task start, longer time-on-task, fewer app switches, more questions/help-seeking, and upward shifts in platform time and submission rhythms [104, 178, 179]. Slow trends in motivation align with long-term habit change—more regular study windows, higher persistence on long tasks, more autonomous course/project choices, and more frequent cross-course transfer [163, 165]. In our analysis, we treat behavior and motivation as two way related in a tightly coupled manner: observable behavioral patterns at a given moment indicate the current level of motivation, motivation in turn drives and shapes behavior, and gradual or abrupt changes in behavior feed back to adjust motivation over time. We explicitly keep the SDT meaning visible by interpreting short term changes in motivation related labels as echoes of need support or need frustration in the student's environment. By viewing slower, longer term trends as adjustments in value endorsement and the progressive integration of learning goals into the self. These modelling choices produce time series labels that better reflect the psychological meaning of SDT constructs, rather than treating motivation as a purely static score. They provide a stable, interpretable target for mapping from behavioral representations in both the numeric and textual routes.

2.1.3 Conceptual Bridge to Sensing

This subsection builds a conceptual bridge between SDT-grounded learning motivation and passively sensed behavioral data. It first shows how continuous sensing streams and platform logs can be organized into everyday behavioral feature families, such as temporal rhythms, daily routines, contextual environments, and social versus solitary patterns. It then outlines how these observable indicators are time aligned with intrinsic motivation measures across multiple time scales and analyzed relative to individualized baselines to obtain interpretable mapping rules and time series labels. Together, these steps provide the foundation for treating sensed behavior as an SDT-consistent proxy space for learning motivation that can be empirically tested for convergent validity, robustness, and transferability.

Observable Behavioral Indicators: In a continuous assessment framework, observable behavior data can be collected continuously via sensing (e.g., passive sensing and platform/device logs). After grouping by attributes, these data form *behavioral feature families*, defined in everyday contexts as:

- *Temporal Rhythm Features:* When activities happen, how long they last, how they are spaced, and how regular they are (e.g., within-day/within-week distributions of activity windows, session length, usage intervals, sleep/nap periods) [59, 132, 180–183].
- *Daily Patterns and Habits:* Typical ways and stable routines for handling daily tasks (start–switch–end sequences, resumption after interruptions, long-term persistence, planning ahead vs. last-minute handling, single-task focus vs. multitasking) [140, 177, 184–188].
- *Context and Environment Features:* Where behavior occurs and environmental characteristics when permitted (share and stability of time at home/classroom/campus, location-switch frequency, and when available, noise/light or network context) [55, 189–193].
- *Social vs. Solitary Features:* Behaviors with social interaction vs. solitary behaviors [39, 190, 194–196].

Foundations for Mapping Sensed Behaviors to Learning Motivation Levels: Using passive mobile sensing for continuous assessment, we time-align continuous changes in observed behaviors across micro (minutes/hours), meso (days/weeks), and macro (months/semester) timescales with learning motivation levels (especially intrinsic motivation), and then build and test a mapping from behavior to motivation. With passive sensing and platform logs, we continuously observe behavioral features; these external behaviors show systematic links to motivation level, and some are driven by changes in motivation [39, 59, 197].

To quantify this relation, we time-align behavior and motivation series, standardize them relative to an individualized baseline, and perform association analyses. These include within-person correlations, cross-correlations and tests of precedence, and hierarchical models that separate within-person, between-person, and between-group effects. We also allow both directions in interpretation: behavior can indicate

and influence motivation, and motivation can drive changes in behavior.

For validity, we focus on two checks. *Convergent validity:* do the behavior series and the motivation indicators show time-aligned agreement? *Robustness and transferability:* are results stable across window sizes, course/assignment cycles, and term rhythms, supported by resampling or cross-validation? Because mapping sensed behavior to internal learning motivation is a relatively new approach, its effectiveness must be verified through these tests. The outcome is a set of SDT-consistent, time-aligned, and interpretable mapping rules and time-series labels.

2.2 Technical Foundations

This section introduces how passive smartphone sensing, digital phenotyping, behavioral representation, and LLM-augmented inference jointly support the hybrid assessment of learning motivation. It first defines passive smartphone sensing and clarifies its role as a low-burden data acquisition channel, situating it within a digital phenotyping pipeline that cleans, windows, and transforms multi-modal logs into interpretable behavioral phenotypes. It then describes how these heterogeneous, irregular, and privacy-sensitive data streams are preprocessed, aligned, and converted into behavioral features and unsupervised behavior patterns, with explicit attention to interpretability and ethical compliance. Finally, the section outlines how LLM-based modeling operates on textualized behavioral summaries to estimate intrinsic learning motivation, contrasts this approach with traditional machine learning and deep learning, and highlights its implications for within-person sensitivity, unobtrusive assessment, cross-context adaptability, and theory-consistent interpretation.

2.2.1 Passive Smartphone Sensing and Digital Phenotyping

This subsection clarifies how passive smartphone sensing and digital phenotyping provide the technical backbone for our hybrid assessment approach. We first define passive sensing as a low-burden data acquisition channel that continuously collects multi-modal behavioral and context signals from students' own devices and situate it within a digital phenotyping pipeline that cleans, segments, and transforms raw logs into interpretable behavioral phenotypes. We then outline the operational flow from ethically governed acquisition and preprocessing to windowing and feature/phenotype construction, and summarize key data characteristics such as multimodality, irregular timing, missingness, device heterogeneity, and privacy/security constraints. Together, these elements specify both the opportunities and the methodological challenges of using smartphone-derived traces as a basis for continuous, SDT-aligned assessment of learning motivation.

Passive Smartphone Sensing: Collect behavior and context data without extra user actions, using built-in sensors and system/app logs with low burden over time [55, 59, 189, 198]. In practice, the smartphone functions as a ubiquitous sensing platform that runs in the background while students go about their usual activities, so that data are gathered continuously without requiring additional effort or active input. This passive mode of collection yields ecologically valid traces of everyday life and study

routines, making it possible to observe behavioral regularities and changes over extended periods with minimal disruption. Typical sources include:

- *Motion:* Accelerometer, gyroscope, magnetometer [199, 200]. These motion sensors capture fine-grained information about movement, orientation, and acceleration of the device, which can be used to infer activity levels, rest–activity cycles, and aspects of physical routines (e.g., sedentary versus more active periods) at different times of day.
- *Location & Proximity:* Global Positioning System (GPS), cell towers, Wi-Fi scans, Bluetooth [198, 201]. Together, these signals provide coarse-to-fine information about where a person spends time and how their proximity to places, networks, and other devices changes, enabling the identification of key locations (such as home, campus, or library) and patterns of movement or co-location relevant to learning contexts.
- *Device Interaction:* Screen on/off and unlock, foreground app sessions, app switches, notifications and responses, input counts [202–206]. Such interaction logs describe how frequently and in what ways the phone is used, including bursts of activity, persistent engagement with certain apps, and responsiveness to notifications, offering a window into attention rhythms, digital multitasking, and possible interruptions during study.
- *System Context:* Battery/charging, network type and changes [207]. These system-level indicators reflect aspects of the device's operating environment, such as when it is charging, how connectivity fluctuates, and which network types are available, all of which can influence or constrain sensing quality and also indirectly signal aspects of daily routine (for example, typical charging times or places).

Sampling strategy (periodic, event-triggered, duty cycling, adaptive) to balance battery, resolution, and timeliness, plus on-device preprocessing (de-identification, compression, noise reduction) [208–211]. In designing a sensing system, these choices jointly determine how often sensors are queried, how much detail is captured, and how strongly the system impacts battery and storage, while local preprocessing reduces data volume and protects privacy before transmission or analysis.

Digital Phenotyping: A framework to turn multi-source digital traces into high-resolution profiles of behavior and functioning. The pipeline is from clean to segment to feature to pattern, producing interpretable phenotypes and linking them to target variables. By imposing this staged transformation, digital phenotyping offers a structured way to move from heterogeneous raw logs to compact, meaningful descriptions of behavioral tendencies that can then be statistically associated with outcomes of interest such as learning motivation.

Relation: Passive smartphone sensing is the data acquisition channel; digital phenotyping is the analysis pipeline from raw streams to interpretable patterns. Together they provide the route "sense to represent to interpret." [55, 212, 213] In other words, sensing provides the continuous, real-world input signals,

while digital phenotyping organizes, summarizes, and connects those signals to psychological constructs, so that the combined framework can support theory-driven, data-rich assessment.

Operational Flow for Behavioral Data

- *Acquisition & Compliance:* Informed consent, data minimization, on-device aggregation and de-identification [211, 214, 215]. Data collection begins only after participants have been clearly informed about the study procedures and risks and have provided formal consent, ensuring that participation is voluntary and compliant with ethical requirements. Throughout the pipeline, data minimization principles are applied so that only information strictly necessary for the research questions is collected, stored, and processed. Wherever possible, aggregation and de-identification are performed directly on the device, reducing the amount of raw, potentially identifying data that ever leaves participants' phones and aligning the acquisition process with contemporary privacy and compliance guidelines.

- *Preprocessing & Quality Control:* Time alignment, missing handling, artifact filtering, unified time zones and landmarks [180, 216–218]. In the preprocessing stage, heterogeneous logs from different sensors and sources are first aligned in time so that events can be meaningfully compared across streams. Missing data are identified and handled according to predefined rules, and obvious artifacts (e.g., duplicate bursts, impossible values) are filtered out to avoid contaminating downstream features. All timestamps are converted to unified time zones and anchored to consistent temporal landmarks (such as day boundaries or academic calendar events), providing a coherent temporal reference frame for quality control and subsequent analysis.

- *Windowing & Feature Extraction:* Fixed/sliding windows at minute/hour/day/week scales and event-aligned windows [211, 219–221]. The cleaned event streams are segmented into fixed or sliding windows of predefined length (for example, minute-, hour-, day-, or week-scale windows), as well as windows aligned to salient events when appropriate. Within each window, raw samples are summarized into features such as counts, durations, variability measures, or ratios, turning irregular, high-frequency logs into structured, window-level descriptors. This windowing and feature extraction step preserves temporal resolution at relevant scales while producing a tractable representation for modeling.

- *Phenotype Construction:* Group related features into behavioral phenotypes (e.g., temporal rhythm, place/context, social–solitary structure) for downstream analysis [213, 222–225]. After feature extraction, conceptually related indicators are organized into broader behavioral phenotypes, such as patterns that characterize daily and weekly rhythms, dominant places and contexts, or the balance between social and solitary activity. These grouped phenotypes provide more interpretable units than individual features and can be used as higher-level predictors or descriptors in subsequent statistical and psychometric analyses. In this way, the feature space is structured around meaningful

behavioral dimensions rather than a flat list of isolated variables.

Data Characteristics

- *Multi-modal and source heterogeneity* (sensor streams + system/app logs). This means that the data originate from multiple, fundamentally different channels, such as continuous sensor readings and discrete application or system events, each with its own format, frequency, and noise characteristics. As a result, they must be carefully synchronized and transformed before they can be analyzed together in a single framework.
- *Irregular time structure* (continuous streams and discrete events; gaps due to power saving and permissions). The data arrive at non-uniform intervals, mixing high-frequency continuous measurements with sporadic events, and containing holes where logging was paused. Such gaps can be caused by battery-saving mechanisms, revoked permissions, or the operating system stopping background processes, and they complicate any attempt to reconstruct a smooth time series.
- *Resolution trade-offs* (sampling interval, duty cycle, location accuracy vs. battery). Decisions about how often to sample sensors, how long to keep them active, and how precise location estimates should be directly affect both data richness and device burden. Higher temporal and spatial resolution improves detail but consumes more energy and resources, so the sensing design must balance accuracy with acceptable battery and performance costs.
- *Missing data, noise, and drift* (device differences, indoor GPS, permission loss, background kills, changing routines). Missing segments, spurious values, and slow changes in signal properties are common, arising from hardware and firmware differences, poor indoor GPS quality, users temporarily revoking permissions, apps being killed in the background, or shifts in daily behavior patterns. These issues must be recognized and handled explicitly to avoid biased estimates and misleading conclusions.
- *Device/Operating System (OS) heterogeneity* (brands, OS versions, permission models). Different manufacturers, operating-system versions, and permission policies implement sensing and background execution in diverse ways, so identical code may produce different data quality and coverage across devices. This heterogeneity introduces systematic variation that must be accounted for when comparing participants or combining datasets.
- *Proxy semantics* (signals reflect externalized behavior, not the latent variable directly; mapping needs theory and evidence). The collected signals are only indirect manifestations of the constructs of interest, capturing observable behavior rather than the underlying psychological states themselves. Interpreting them as indicators of a latent variable therefore requires explicit theoretical reasoning and empirical validation, rather than assuming a one-to-one correspondence.
- *Context dependence and confounds* (the same pattern can mean different things across tasks or

exam periods). A given behavioral signature may have different interpretations depending on the surrounding context, such as whether it occurs during regular coursework or during high-stakes exam weeks. Without considering these contextual factors and potential confounders, apparent associations may be misread or attributed to the wrong underlying processes.

- *Time consistency* (time zones, daylight saving, clock skew). Ensuring that all timestamps are comparable requires correcting for differences in time zones, handling daylight-saving changes, and adjusting for device clock drift or misalignment. If these issues are not properly resolved, temporal patterns and alignments with external events can become distorted.
- *Privacy and compliance* (informed consent, minimization, on-device aggregation/blur, possible differential privacy or derived features). Data collection and processing must follow strict ethical and legal standards, including obtaining informed consent, limiting data to what is necessary, and using techniques such as on-device aggregation or blurring to reduce identifiability. Where appropriate, methods like differential privacy or reliance on derived features instead of raw logs can further strengthen privacy guarantees.
- *Security and compute* (edge–cloud split, encryption, access control, processing audit). The system must protect data in transit and at rest through strong encryption, enforce strict role-based access control on servers and devices, and clearly separate what is processed locally on the device from what is transmitted to the cloud for aggregation, storage, or modeling. Compute-intensive operations should be scheduled in ways that respect device constraints, for example by prioritizing on-device preprocessing and deferring heavier workloads to periods of charging or stable Wi-Fi connectivity, so that battery drain and performance impact remain acceptable. All processing stages should be logged with timestamps, component identifiers, and configuration versions, enabling end-to-end audits that can reconstruct data flows, detect anomalies or misuse, and demonstrate compliance with institutional security policies and external regulatory requirements.

2.2.2 Behavioral Representation

Behavioral Features & Behavior Identification: Convert multi-modal, time-stamped streams into comparable, interpretable features and patterns. This work uses *unsupervised* approaches only (rule/prototype methods and clustering). Supervised methods are out of scope [226, 227].

Preprocessing and Time Alignment

- *Time unification:* unify time zones/Daylight Saving Time (DST), deduplicate, fix clock drift; map continuous streams and discrete events to one timeline [180, 216, 228]. This step ensures that all records from different devices and sources share a single coherent temporal reference, so that events can be compared and merged without ambiguity. By first removing repeated entries and correcting for device-specific clock offsets or drifts, we obtain a consistent, ordered sequence in which both continuous measurements and discrete events are placed on the same unified timeline

for downstream processing.

- *Missing and noise:* resample (e.g., 1s/1min/5min), reasonable imputation, artifact filtering (e.g., GPS drift, background heartbeats) [229–232]. After unifying time, streams are resampled at predefined intervals to create regular time grids, which facilitates comparison across sensors and subjects. Within this framework, we apply carefully chosen imputation strategies only where appropriate and remove obvious artifacts such as implausible GPS jumps or synthetic "heartbeat" pings from background services, thereby reducing noise while preserving genuine signal structure.
- *Window design:* fixed/sliding windows (e.g., 15 min / 1 h / 24 h) or event-aligned windows (e.g., 1 h after unlock; ±30 min around a submission); set overlap [219, 220, 228]. Based on the regularized timeline, we segment the data into analysis windows whose length and position are chosen to match the temporal scale of interest. These windows may be fixed or sliding, and can also be anchored to specific events, with controlled overlap to capture temporal context while avoiding excessive redundancy, thus shaping the basic units for feature extraction.
- *Individualized baseline:* within-person standardization (z-scores/percentiles/deviation from baseline) to separate within-person change from between-person differences [119, 120, 233]. For each individual, we transform raw window-level values into personalized metrics that express how much a given observation deviates from that person's own typical level. This within-person standardization allows us to distinguish changes over time within the same student from stable differences between students, aligning the data structure with longitudinal and person-centered analyses.
- *Multi-scale versions:* prepare aligned data at micro/meso/macro scales for reuse in later steps. Using the same underlying timeline and windows, we derive parallel datasets at different temporal resolutions (for example, minute-level, hour-level, and day- or week-level aggregates). These multi-scale versions remain mutually aligned, so that later analytic stages can flexibly choose the most appropriate scale without repeating early preprocessing, and can relate patterns observed at fine-grained levels to broader temporal summaries.

Feature Extraction

- *Temporal rhythm and regularity:* session length, start latency, spacing, day/week periodic strength (e.g., simple spectral energy), regularity indices, burstiness. Together, these indicators describe how long typical interaction episodes last, how quickly they are initiated after key events, how evenly they are spaced across time, and how strongly behavior repeats with daily or weekly cycles. They capture whether activity unfolds in a smooth, predictable rhythm or in short, intense bursts separated by long pauses, providing a compact summary of temporal structure and regularity.
- *Foreground/app interaction:* counts and duration of foreground sessions, app switch rate, foreground/background ratio, notification-to-response delays, keystroke/edit bursts [202, 234–236]. These features quantify how frequently and for how long users actively engage with their devices,

how often they move between different apps, and how quickly they respond to incoming notifications. They also reflect short intervals of intense typing or editing, thereby characterizing the intensity and fragmentation of focused interaction with the smartphone.

- *Mobility and places:* number and duration of stays, place-switch rate, radius of gyration, location entropy, commute/entry–exit rhythms [194, 237, 238]. This set of indicators summarizes how many distinct places a person visits, how long they remain in each location, and how frequently they move from one place to another. Measures such as radius of gyration, entropy of visited locations, and regular entry–exit patterns capture the spatial spread, diversity, and routine structure of mobility behavior.

- *Communication/social (metadata, compliant):* counts and duration of sessions, reciprocity, within-day distribution, collaboration event counts and spacing [59, 239]. These variables focus on high-level properties of social and collaborative activity, based on metadata rather than content, such as how often and how long interactions occur and how balanced exchanges are between parties. They also describe when during the day such interactions tend to cluster and how collaboration-related events are spaced over time, offering a coarse-grained view of social engagement patterns.

- *Device/system context:* battery curve, charging cycles, network type/switches, screen on/off counts. These features characterize the operational environment of the device, including how battery level changes over time, when and how often the phone is charged, and which network connections are in use. They also include counts of screen on/off events, providing additional context about overall device availability and usage conditions during the observation period.

Behavior Identification (unsupervised)

- *Rule/prototype methods (lightweight, interpretable):* threshold and logic rules to mark typical segments (e.g., “long session + low switching = focused segment”). These approaches rely on simple, human-readable conditions that directly encode intuitive behavioral criteria, so that each detected segment can be easily explained in terms of the underlying rules. Prototypes derived from such rules serve as canonical examples of particular behavioral modes, providing a straightforward way to label and communicate patterns without requiring complex models.

- *Unsupervised clustering:* k-means [240] / GMM [241, 242] / Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [243] / spectral clustering [244] on window-level features to group behavioral states and obtain centroids and prototypes. In this case, window-level feature vectors are partitioned into clusters so that segments with similar behavioral profiles are grouped together, and each cluster is summarized by a centroid or representative prototype. These unsupervised methods reveal recurring state types without pre-specified labels, allowing researchers to interpret each cluster post hoc as a distinct behavioral state and to use its centroid as a concise description of that state in downstream analyses.

Interpretation and Robustness

- *Interpretability:* In our study, the extracted behavioral patterns will be co-interpreted with collaborating psychology experts to align them with SDT constructs and to check face/content validity.
  - *Feature contributions:* quantify how each behavioral feature pushes the estimated score up or down (e.g., coefficients/SHapley Additive exPlanations (SHAP)) [245], enabling both local and global explanations.
  - *Cluster prototypes:* representative centroids/medoids of discovered behavioral states that summarize salient feature levels and anchor human-readable labels.
  - *Exemplar visualizations ("why this state"):* time-aligned traces and session timelines that show typical windows assigned to a state and the cues that drove the assignment.

Privacy and Compliance

- *Data minimization and on-device aggregation/anonymization; informed consent, revocation, purpose limitation.* Concretely, only the smallest set of signals required for the research questions is collected, and wherever possible they are aggregated or anonymized directly on the device before upload, reducing the risk of identifying individual participants. Participants are clearly informed about what is collected, why it is collected, and for how long it will be stored, and they retain the right to withdraw or revoke their participation, ensuring that data use remains tightly bound to the stated and approved purposes.
- *Versioned processes* (sampling, preprocessing, feature configs) to support reproducibility and audit. Each major step in the pipeline—such as sensing configurations, preprocessing scripts, and feature-definition files—is given an explicit version label and documented so that analyses can be exactly reproduced at a later time. This versioning also enables systematic auditing of changes over time, making it possible to trace which configuration produced which results and to compare outcomes across different pipeline variants in a transparent way.

2.2.3 Learning Motivation Modeling (LLM Modeling)

Large Language Models can perform instruction-guided inference on textual summaries of behavior and context. When they are given a window level summary that condenses the relevant signals into natural language, they can output an estimate of the target psychological variable (for example, an intrinsic learning motivation level, a numeric score, or a qualitative direction of change) together with a short, human readable rationale that explains how this estimate was reached. Strengths of this approach include (1) the ability to work in few shot and even zero shot settings without additional task specific training, (2) the capacity to semantically integrate heterogeneous signals once they have been expressed in a unified textual summary, and (3) the production of readable explanations with structured outputs that can be inspected, logged, and linked back to the underlying indicators for further analysis.

From Labeled Behavioral Data to A Motivation Estimate

- *Time alignment and textualization:* aggregate features in a same-window unit (e.g., 1 h / 24 h / 1 week) and convert to a compact, human-readable summary (include deviation from the individualized baseline). In practice, this means that all numeric indicators within a given window are combined into a single description that reflects both their absolute level and how far they lie above or below that person's typical pattern. The resulting textual snippet acts as a concise, window-level "behavior profile" that preserves temporal structure while remaining directly interpretable by humans and LLMs.
- *Prompt design:* give task instructions (estimate motivation level/direction), add a few labeled examples, and define a structured output format (e.g., JSON with {level/score/rationale}). The prompt therefore specifies exactly what the model is expected to do, what the possible response scale or categories are, and how the answer should be organized. By including a small number of illustrative examples and a fixed output scheme, we reduce ambiguity for the model and make the downstream parsing and comparison of its responses straightforward.
- *Model inference and calibration:* run the LLM, then map outputs to the individual baseline and perform temporal splits/rolling checks for stability and calibration [246–248]. After obtaining raw estimates from the model, we re-express them relative to each student's own baseline to ensure that person-specific levels are taken into account. We then evaluate these baseline-adjusted outputs under different temporal partitions or rolling windows to examine whether performance and score distributions remain stable over time, and to detect possible drift or miscalibration.
- *Same-window measurement* (optionally with precede/follow checks): focus on same-window estimates; optionally test whether behavior precedes or follows motivation in adjacent windows for mechanism probes [249–251]. The primary analysis treats behavior and motivation that fall in the same window as concurrent indicators, providing a direct snapshot of their co-occurrence. When needed, we additionally compare windows just before and just after a given motivation measurement to explore whether particular behavioral changes tend to come earlier or later, offering tentative clues about temporal ordering and possible mechanisms.

Brief Contrast with Traditional Machine Learning/Deep Learning

- *Traditional machine learning/deep learning:* trains on numeric features with larger labeled sets; strong numerical stability and mature deployment, but higher cost for cross-context transfer and semantic explanation [252, 253]. In practice, these models typically rely on carefully engineered feature vectors and benefit from substantial amounts of annotated data, which allows them to achieve robust performance under well-defined training conditions. However, adapting them to new institutions, populations, or sensing configurations often requires additional retraining or fine-tuning and careful revalidation, and their internal representations are not naturally expressed in

psychologically meaningful terms, so extra work is needed to provide human-understandable explanations.

- *LLM inference:* instruction-driven, works with few aligned examples; good at semantic integration and explanation; needs strict structured outputs, consistency checks, and calibration. They can be used together. In this setting, LLMs can operate directly on textualized feature descriptions and prompts, integrating heterogeneous signals through natural-language reasoning and producing both quantitative estimates and verbal rationales in a single step. To make these outputs usable in a psychometric workflow, the prompts must enforce a clear schema, and additional procedures are required to check reliability over time and align the model's responses with numeric baselines, allowing LLM-based inference to complement traditional models rather than replace them.

Implications for Assessing Learning Motivation

- *Within-person sensitivity:* combine deviation from baseline with context cues to detect small changes. In practice, this means expressing each estimate relative to the student's own typical level rather than to a group average, and jointly considering situational information such as time of day, calendar phase, or known study periods. This combination helps the method pick up subtle shifts in motivation that might otherwise be obscured by stable between-person differences or by routine fluctuations in behavior.
- *Unobtrusive and explainability:* same-window inference lowers sampling burden and yields checkable rationales. Because behavior and motivation are estimated within the same analysis window, there is no need for dense additional self-report sampling, keeping the overall burden on students low. At the same time, the method can return short, human-readable explanations for each estimate, allowing researchers and practitioners to inspect and verify the reasoning behind the inferred motivational level.
- *Cross-context adaptability:* adapt by editing instructions and examples across courses, weeks, or cohorts. Instead of retraining a full numerical model, the approach can be tuned to new teaching formats or institutional settings by adjusting the prompt wording, the reference scales, and the illustrative examples used. This makes it easier to reuse the same basic framework across different courses, semesters, or student groups while still respecting local context.
- *Method boundaries:* avoid assuming "more activity = higher motivation"; keep SDT-consistent meaning; audit outputs for structure, calibration, and stability. The approach must explicitly guard against overly simplistic interpretations of behavioral intensity, ensuring that estimated motivation remains grounded in SDT constructs rather than in raw activity levels alone. Regular checks on output format, distribution, and temporal behavior are therefore required to verify that results stay well-calibrated, structurally consistent, and theoretically coherent over time.

2.3 Review of Related Work: Assessing Learning Motivation in Higher Education

This section reviews how learning motivation has traditionally been assessed in higher education and positions passive smartphone sensing as a complementary, continuous assessment route. It first summarizes standard methods such as self-report scales, structured classroom observations and semi-structured interviews, and ecological momentary assessment (EMA/ESM), outlining their strengths and limitations in terms of construct clarity, contextual richness, temporal resolution, burden, and coverage. It then turns to existing work on passive smartphone sensing and related constructs, highlighting how background logging of behavioral signals can unobtrusively track students' routines over time and how selected studies in higher education (e.g., StudentLife, SmartGPA, long phone+EMA designs) link sensed behavior to stress, engagement, and academic outcomes. Building on a structured, side-by-side comparison of these approaches, this section mainly identifies three key research gaps, namely that: (1) there is a lack of direct, continuous SDT-based learning motivation assessment via sensing, (2) there is an absence of reusable SDT-aligned behavioral indicators and explanation rules, and (3) there is a need for an integrated, scalable, and privacy-conscious continuous assessment pipeline. These gaps directly motivate and align with the hybrid approach developed in this dissertation.

2.3.1 Traditional Assessment Methods for Learning Motivation in Higher Education

Self-Report Scales: The most common approach is to use structured questionnaires (Likert scales) that directly measure motivation [43]. Typical tools include Self-Regulation Questionnaire (SRQ)/SRQ-A (SDT-based self-regulation styles: external, introjected, identified, intrinsic; used in academic contexts and adapted for Higher Education (HE)) [13, 45], Academic Motivation Scale (AMS) (intrinsic, extrinsic, amotivation; many higer education adaptations) [16], Intrinsic Motivation Inventory (IMI) (SDT-based experience for a specific activity: interest/enjoyment, perceived competence, effort) [15], and Motivated Strategies for Learning Questionnaire (MSLQ) (motivation and learning strategies, with an official manual) [43]. These are usually delivered online by instructors, assessment offices, or research teams. Timing is often pre/post (start/end of term) or three waves (start–mid–end) [157, 254]. Some studies use short weekly forms to lower burden and match learning logs [173, 255]. *Strengths:* clear constructs, easy to interpret, low cost, scalable. *Limits:* self-report bias and low time resolution; cross-course or cross-culture use needs local validation and tests of measurement invariance [49, 256].

Structured Classroom Observation & Semi-Structured Interviews: Trained observers record student participation and teacher–student interactions in real classes (e.g., Classroom Observation Protocol for Undergraduate STEM (COPUS) with 2-minute time slices) [257, 258]. Researchers also run semi-structured interviews at key points (e.g., midterm) to capture mechanisms [259]. *Strengths:* rich process

and context evidence; good for triangulation with scales and logs [260–262]. *Limits:* training and inter-rater work, high human cost, limited coverage, weak representation of out-of-class study, and higher ethical/data-governance demands [263, 264].

Ecological Momentary Assessment (EMA/ESM): Short prompts in real life, triggered by time, random events, or specific situations, to capture "right-now" motivation, engagement, stress, fatigue, etc [265, 266]. Typical settings in student samples are ≈6 prompts/day for ≈7 days with ≈70–80% compliance [267, 268]. *Strengths:* high real-world validity and high time resolution; shows within-person change and links to context [121, 269]. *Limits:* still active self-report, with burden and reactivity; busy periods (e.g., exams) often have selective missingness [268, 270, 271].

Summary: Traditional methods complement each other: scales give clear constructs; observation/interviews give context and mechanisms; EMA/ESM gives time resolution. For continuous assessment, we still need approaches that (a) track within-person change aligned with the academic rhythm and (b) to a large extent, do not disturb daily activities.

2.3.2 Using Passive Smartphone Sensing to Assess Learning Motivation and Related Constructs

To the best of our knowledge, no prior work has directly conducted continuous measurement of intrinsic learning motivation within the SDT-framework using smartphone-based sensing. Nevertheless, studies that combine sensor-based assessments with psychological variables in other higher education contexts may offer useful insights. Passive smartphone sensing uses phone logs/sensors (location, foreground app sessions, screen unlock/time, notifications and checking, communication/activity rhythms, mobility and dwell, activity patterns) to continuously record behavior in the background [189, 272]. Students do not need to fill repeated forms like EMA; data collection, to a large extent, does not disturb daily activities [267, 273]. This keeps time order and supports tracking slow drifts in internal states. In public work, direct, continuous estimation of SDT learning motivation itself is still rare [223, 225, 274, 275]. Most studies target related constructs (stress/affect [224, 276–278], engagement/persistence [132, 279–281], academic outcomes [39, 282–284]) and align passive signals with light labels (short scales/EMA) in time windows to study "behavior patterns ↔ psychological/learning states." For example, StudentLife shows links among location, conversations, activity, sleep, and phone use with stress, affect, and academic outcomes over a semester [59]; longer studies link phone signals and EMA to examine co-change and lead–lag [40, 285, 286].

Three SDT-relevant Themes:

- *Relatedness:* Passive "social sensing" uses call/message frequency and spread, response rhythms, sustained conversation density, and approximations of face-to-face co-presence (e.g., long co-location) [287]. These traces show time-linked associations with loneliness, stress, and negative affect [288–291]. These can act as behavioral proxies of relatedness need satisfaction or frustration.
- *Competence (engagement, control, effective effort):* Foreground session structure, app switching rate, sustained task time, attendance/mobility patterns, and stable routines help explain or predict engagement quality and grades (e.g., SmartGPA) [39].
- *Autonomy (self-direction and choice):* Daily context choice and switching patterns (mobility diversity, predictability changes, self-scheduled foreground use) link to perceived control, functioning, and affect stability [292, 293], suggesting ecological proxies of autonomy need satisfaction (to be validated against scales/EMA) [118].

Representative Lines in Higher Education:

- *StudentLife:* Semester-scale sensing (location, activity, social, sleep) + light labels to track stress, affect, participation, and outcomes [59]. In this study, students' phones continuously recorded where they spent time, how physically active they were, how much they slept, and how often they engaged in conversations, while brief surveys and academic records provided information on psychological states and performance. Together, these streams enabled the reconstruction of longitudinal profiles over an entire semester, showing how everyday behavioral patterns co-varied with stress, mood, class attendance, and academic results.
- *SmartGPA:* Links study/social time blocks, app switching, and mobility patterns to Grade Point Average (GPA); "regularity–low distraction–persistence" works as measurable engagement signals. By segmenting time into blocks characterized by sustained study-related app use, limited distracting switches, and stable movement routines, the system identified behavioral signatures associated with better academic outcomes. These findings illustrate how phone-derived indicators of regular study habits and reduced fragmentation of attention can serve as quantitative proxies for engagement and are predictive of term GPA [39].
- *Passive Social Sensing Reviews:* Summarize how phone logs can, to a large extent, not disturb daily activities, to track social trajectories and infer relatedness, loneliness, or emotional difficulty. Such reviews highlight that metadata from calls, messages, and co-location patterns can reveal how students' social networks evolve over time and how densely or sparsely they interact. On this basis, researchers can approximate constructs like perceived closeness, social withdrawal, or emerging emotional strain, while keeping the sensing burden low and avoiding direct intrusion into message content [34, 290].

- *Long Phone and EMA Studies:* Over weeks/months, rises and falls in stress/negative affect or engagement often pair with detectable behavior changes, enabling observation of slow within-person drifts. By repeatedly collecting momentary self-reports alongside phone-based behavior logs, these studies show that gradual shifts in internal states tend to be mirrored by systematic changes in sleep, mobility, interaction, or phone use. This coupling allows researchers to trace how individuals' trajectories of well-being or engagement unfold across longer periods, rather than relying only on isolated snapshots [224, 294].

Strengths: Passive smartphone sensing enables *continuous* collection of behavioral traces in the background and is largely non-disturbing for participants. Data can be windowed to align naturally with the academic rhythm (e.g., days, weeks, modules, exam periods), which simplifies comparison across courses and terms. Because temporal order is preserved, analysts can test both directions of change—"motivation/need change → behavior change" and "behavior change → motivation/need change"—rather than relying only on static snapshots. The pipeline is scalable: once permissions are granted, the marginal burden per student is low and long time series can be gathered with minimal maintenance. Repeated measures within the same person support within-person analysis, allowing estimation of personal baselines, detection of short-term deviations, and separation of between-person differences from within-person fluctuations. These properties make the approach suitable for large cohorts and for n-of-1 views of individual learning dynamics, while keeping participant burden modest [295].

Limits: Phone signals are *behaviors* and contexts rather than the latent construct itself, so any inference about SDT needs or motivation must be theory-guided and used with caution. The mapping from sensed features to SDT needs/motivation should be specified in advance, then tested for convergent validity against established scales before being generalized to new sites or populations. Light labels are still required (e.g., brief scales or EMA) to anchor estimates, check calibration, and monitor drift over time. Stability can also be affected by platform, permission, and battery policies; operating system updates, vendor power management, and app background restrictions may change sampling rates or produce missing windows. These factors introduce variability that must be tracked, documented, and, where possible, mitigated in preprocessing and analysis. Finally, direct, continuous estimation of SDT learning motivation in authentic settings is still rare in published work, so careful replication, cross-site validation, and transparent reporting are needed before high-stakes use [223, 296, 297].

2.3.3 Method Comparison and Research Gaps

This subsection compares the main approaches used to assess learning motivation and related constructs in higher education and clarifies how they complement one another. It first summarizes self-report scales, structured classroom observation and interviews, ecological momentary assessment, and existing passive smartphone sensing in terms of their definitions, key features, strengths, and limits, using a structured table as a compact, side-by-side overview. It then draws out three key research gaps that emerge from

this comparison and explains how the present dissertation is aligned with these gaps by proposing SDT-guided, time-aligned mappings, reusable behavior indicators and explanation rules, and an integrated, largely non-disturbing continuous assessment method.

Comparative Analysis

The table below summarizes each method's Definition/Key features/Strengths/Limits/Example references, and connects to the gaps that follow. It provides a compact, side-by-side overview that makes explicit how the methods differ in scope and design, and how they complement or overlap with one another. By organizing this information in a structured way, the table serves as a bridge between the descriptive review of existing approaches and the subsequent identification of unresolved gaps that motivate the present work.

Self-Report Scales: Standardized and construct-aligned (e.g., SRQ/SRQ-A/AMS/IMI/MSLQ). These instruments provide comparable scores across classes and terms, work well for large samples, and are straightforward to administer online or on paper. They are useful for staged tracking (e.g., start/mid/end of term), establishing baselines, and anchoring other data sources such as logs or sensor features. Typical formats use Likert items with clear subscales, allowing reliability checks (e.g., coefficient alpha/omega) and simple score aggregation. At the same time, limits remain: self-report bias (recall, social desirability), low time resolution for day-to-day changes, and the need for local validation before cross-context use. Cross-course and cross-lingual deployment should test measurement invariance and check for differential item functioning to avoid unfair comparisons. Short forms can reduce burden but should keep coverage of the intended construct. In practice, scales work best as a backbone for periodic assessment and as labels for convergent validity with behavioral indicators. *Examples:* SRQ-A/AMS adaptations in HE; official SDT IMI materials; MSLQ manual.

Structured Observation & Semi-Structured Interviews: Give process and contextual evidence (e.g., COPUS time-slices; interview mechanisms) to document not only what happens in a class but also how it happens moment by moment and under which conditions it occurs. These methods reveal how engagement unfolds over time, which cues and instructional moves are present at different points, and why students act as they do in response to tasks, peers, and instructors in the local setting. They are strong tools for explanation, for refining or revising hypotheses that were originally generated from scales or log data, and for triangulation when results from different sources disagree or seem ambiguous. Observations can use short, repeated time windows (e.g., COPUS 2-minute slices) with trained coders who apply a shared scheme, while interviews use a semi-structured guide that covers goals, challenges, and supports but still allows flexible follow-up probes to clarify meaning and explore unexpected themes. Limits include the need for training and reliability work (e.g., inter-rater checks and periodic calibration meetings), the high human cost in terms of time and expertise, limited coverage of very large cohorts or many course sections, weaker representation of out-of-class behavior that occurs away from observed sessions, and tighter ethics and governance requirements for recording, storing, and using rich qualitative data. Best practice is to

sample classes and weeks strategically rather than attempting full coverage, link observation notes or interview excerpts to non-identifying context descriptors instead of names, and combine these materials with lightweight artifacts such as syllabi and assignments to interpret observed behavior in relation to course design. *Examples:* COPUS original paper and university teaching-center COPUS resources, which provide coding schemes, training materials, and guidance for interpreting classroom observation data.

EMA/ESM: High real-world validity and fine time resolution make this approach well suited to capturing within-person dynamics and linking fluctuations in experience to specific contextual factors. Prompt schedules can be signal-contingent, time-based, or event-based, and items are kept brief so that each assessment can be completed quickly with low burden, which allows repeated sampling across the day without overwhelming participants. These data support analyses of day-to-day change, context effects, and lead and lag relationships with behavior streams, for example examining how motivation shifts before and after particular activities or contexts. However, EMA/ESM is still an active self-report method that imposes effort on participants and can elicit reactivity, since repeated questioning may itself influence how people feel or behave. Busy periods such as exams or deadlines tend to show selective missingness, and prompts can cluster around these times in ways that distort the picture of everyday life. Compliance can be improved by using very short questionnaires, scheduling prompts within fixed daily windows, giving clear and simple instructions, and offering small incentives that encourage continued participation without becoming coercive. Practical factors such as battery life, notification settings, and examination rules also affect whether prompts are delivered and seen. When analyzing EMA/ESM data, it is important to plan explicitly for missing data, to record prompt coverage and delivery, and to report compliance (percentage of prompts answered) so that findings are properly qualified and readers can judge data quality. *Examples:* reviews and meta-analyses of EMA/ESM studies often report about six prompts per day over approximately seven days with overall compliance rates around 70–80%.

Current Existing Passive Smartphone Sensing: Background logging records phone data continuously in the background with little or no user interaction, so day-to-day routines are, to a large extent, not disturbed. This mode of collection aligns naturally with the academic rhythm (e.g., weekly courses, assignments, exams) and can be paired with "light labels" such as brief scales or short EMAs that mark specific windows in time. Because data are gathered over long periods, the approach supports extended time series and within-person analysis, allowing researchers to study how the same individual changes around their own baseline across weeks or semesters. With time-aligned windows, it is also possible to test lead–lag relations between behavior and psychological states, examining whether particular behavioral patterns tend to precede, coincide with, or follow measured changes in learning-related constructs. Typical signals include app-usage sessions, screen on/off and unlock events, notification flow and timing, coarse location together with Wi-Fi/Bluetooth proximity, and motion/activity summaries from inertial sensors. Each of these streams provides complementary information about daily routines (e.g., phone interaction rhythms, mobility/co-presence, periods of activity or rest). Taken together, they deliver near-continuous coverage while keeping response burden low, since participants do not need to repeatedly

report details of their behavior for the system to maintain a rich record.

However, there are important limits. Observed behavior is not identical to the underlying construct, and in other words behavior ≠ construct, so mapping sensed streams to SDT must rely on explicit, theory-guided feature design. In practice, this means that features need to be deliberately constructed to reflect autonomy-related contexts, competence-related contexts, or relatedness-related contexts, instead of being treated as direct measurements of motivation. These theoretically motivated features then require convergent validity checks against established SDT-based scales, so that any apparent association between behavior and learning motivation can be evaluated against an independent benchmark. At the same time, stability can be affected by platform differences, because operating system versions and permission regimes vary across devices, and because iOS and Android expose different types and granularities of logs. Vendor-specific power policies can further throttle sampling or defer background tasks, which changes when and how often events are recorded. Taken together, these technical factors can alter data availability and timing characteristics across devices or releases, and they can make an apparently stable mapping between behavior and motivation less robust than it seems if such heterogeneity is not accounted for in preprocessing and analysis.

Privacy and battery constraints further shape the sensing design, because duty cycling is needed to reduce sensing frequency when appropriate, and on-device preprocessing is required to limit how much raw data is transferred off the phone. Minimal retention policies are also needed so that only necessary artifacts are kept and only for as long as they remain essential for the planned analyses. These design choices reduce privacy risk and energy cost for participants, since less detailed and less persistent data are stored and transmitted. At the same time, they require careful advance planning of sampling schedules, preprocessing pipelines, and feature definitions, so that essential information for downstream analysis is preserved in aggregated or transformed form. In other words, privacy protection and battery preservation must be balanced with analytic needs, and the sensing configuration must be designed in such a way that the resulting dataset still supports valid, interpretable modeling of learning motivation while respecting ethical and technical constraints.

Because direct, continuous assessment of SDT motivation is still rare, careful labeling procedures are essential, for example using brief, well validated scales that are explicitly aligned to specific behavioral windows in time and then repeating this process across multiple cohorts or academic terms to test whether patterns replicate. Such repeated labeling and replication help ensure that the inferred links between sensed behavior and motivational states are not artifacts of a single sample or semester, but instead reflect stable relationships that can support robust psychological inference. In practice, passive sensing functions best as a low burden background layer that complements brief self-report scales or EMA. Doing so, provides dense behavioral coverage that can be fused with sparse labels and used to build person centered models that focus on individualized baselines and deviations from each student's typical routine rather than on one time, purely cross sectional group snapshots. *Examples:* StudentLife; SmartGPA; passive social sensing

reviews. A structured, side-by-side comparison across approaches (columns: *Definition*, *Key features*, *Strengths*, *Limits*, *Example references*) is provided in Table 1.

Table 1: Method comparison across approaches used to assess learning motivation and related constructs.

Method	**Key Features**	**Strengths**	**Limits**	**Example References**
Self-report scales (SRQ/SRQ-A, AMS, IMI, MSLQ)	Standardized; mostly online; low-frequency waves (start/mid/end); supports cross-group comparison	Clear constructs; interpretable; low cost; scalable	Self-report bias; low time resolution; needs invariance tests and local adaptation	[13, 45], [16] , [15], [43]
Structured classroom observation & semi-structured interviews	2-min time slices; short training; key sessions; audio/ transcripts/coding	Rich context and mechanism insight; good for triangulation	High human cost; limited coverage/frequency; weak outside-class coverage; higher ethics demands	[257, 258], [259]
EMA/ESM	≈6/day × ≈7 days; ≈70–80% compliance	High ecological validity; high time resolution; within-person focus	Active self-report; burden/reactivity; selective missingness in busy periods	[265, 266], [267, 268]
Passive smartphone sensing (smartphone-only)	Largely non-disturbing; aligns with weeks/ assignments/ exams; often window-aligned with light labels	Long time series; within-person analysis; scalable; can test lead–lag	Behavior ≠ construct; need SDT mapping & convergent validity; platform/permission differences; privacy/compliance; few direct SDT-motivation studies	[59], [39], [225, 275, 288], [224, 276–278], [132, 279–281], [39, 282–284]

Key Research Gaps

- *Lack of direct, continuous assessment of SDT learning motivation in HE using passive smartphone sensing.* We need SDT-guided observable proxies, time alignment, within-person tests, and checks

of stability and interpretability.

- *Lack of reusable SDT-aligned behavior indicators/feature families, model prototypes, and explanation rules.* Different studies define features and pattern labels differently, reducing comparability, replication, and transfer. The field needs shared SDT-grounded "behavior → motivation"resources.
- *Lack of an integrated continuous assessment method that is passive, to a large extent non-disturbing, and scalable (collection → time alignment → modeling → explanation → calibration), with systematic tests of convergent validity, robustness/transferability, and privacy compliance.* Most studies focus on related constructs or partial steps; end-to-end, reusable, compliant solutions are still missing.

Alignment with this dissertation: We first build SDT-guided time-aligned mappings and same-window estimation from behavior to motivation; then we produce reusable behavior indicators/feature families, pattern prototypes, and explanation rules for cross-study use; and finally we design an integrated, largely non-disturbing method (collection → alignment → modeling → explanation → calibration) with systematic checks for convergent validity, robustness/transfer, and privacy minimization.

2.4 Chapter Summary

This chapter set the base for continuous assessment of learning motivation in higher education. In SDT, we focus on intrinsic motivation as the target scalar. It is trait-like but still changes within the same person over time. We described rhythms across multiple timescales and argued for time-aligned, continuous assessment. We also built a conceptual bridge from SDT constructs to observable behavior so that later chapters can map behavior to theory.

On the technology side, we introduced passive smartphone sensing (smartphone-only) and showed how to turn raw logs into behavior feature families and behavior patterns. Data can be collected in the background and, to a large extent, without disturbing daily activities. We outlined a simple modeling primer and noted how LLM-augmented inference can estimate motivation in the same time window, with straightforward splits and metrics.

The related work shows that traditional methods (scales, observation/interviews, EMA) and passive sensing approaches are complementary. Phone-based behavior provides ecological proxies for SDT needs (relatedness, competence, autonomy) and for engagement, yet direct, continuous assessment of SDT learning motivation itself is still rare.

Three gaps remain. First, the field lacks direct, continuous SDT-aligned assessment using passive sensing in higher education. Second, there is no reusable set of SDT-aligned behavior indicators/feature families, model prototypes, and explanation rules. Third, we lack an integrated, largely non-disturbing continuous assessment method that covers collection, time alignment, modeling, explanation, and calibration and that includes systematic checks of convergent validity, robustness/transferability, and privacy compliance.

3

Passive Smartphone Sensing Design, Data Collection, and Preprocessing

This chapter establishes the shared technical core that is used throughout the dissertation to collect passive smartphone sensing data and psychometric measures in parallel within authentic educational settings. It specifies the common architecture, configuration, and procedures that make the studies comparable and reproducible across cohorts and time. The material is organized as follows. Section 3.1 presents the device-side design, describing the Android-native *iSense* application together with its supporting backend components, and clarifying how the client and server interact during routine operation. Section 3.2 details the data-collection workflow, in which weekly questionnaires are administered with a stable cadence while passive sensing runs over the same periods, ensuring that self-reports and behavioral traces cover matching windows. Section 3.3 explains the preprocessing and reprocessing steps that transform raw logs into an analyzable *subject* × *window* dataset suitable for subsequent modeling and analysis.

Unless a later chapter explicitly states a deviation, the methods, parameter choices, and operational settings defined here function as the defaults for all studies in the dissertation. Any departures (e.g., alternative settings or adjustments introduced for a specific research question) are documented in the corresponding chapter; in the absence of such notes, readers should assume the procedures in this chapter govern the implementation and interpretation across the work.

3.1 Passive Smartphone Sensing Design

This section details the concrete implementation of our passive sensing infrastructure, from on-device modalities and sampling strategy to backend storage, version compatibility, and protection measures. We first describe how *iSense* uses event-driven capture on Android (with low-frequency polling only where callbacks are insufficient), and how its installation, project-join, permission granting, setup checks, and battery-optimization exemption are guided step by step to secure stable background logging while still

collecting metadata only. We then outline the console and data stack based on Vue 2, Vite, MDUI, MongoDB, and Aliyun OSS, including project/user path organization, per-project databases, questionnaire push logs, and SDK synchronization. Finally, we explain how power, privacy, and security are treated by design: event-driven capture, coarse location, batch uploads with offline retry, data minimization and de-identification, encrypted transport and isolated storage, strict role-based access and auditing, and explicit retention and deletion rules that ensure only de-identified or aggregated derivatives are used for analysis.

3.1.1 Modalities and Sampling Strategy

We prioritize event-driven sensing and resort to low-frequency polling only when a reliable callback is unavailable or insufficient (for example, for coarse location updates or certain motion sensors). Concretely, triggers include screen unlock/lock events, foreground-app transitions, notification posts, network connectivity changes (e.g., Wi-Fi, cellular, offline), and power/charging state changes. To minimize privacy risk and energy cost, we collect only timestamps, counts, categories, and state indicators, and we do not collect any content-level data (message bodies, audio, images, or typed text). This design reduces radio/sensor use and CPU wakeups while preserving the behavioral signals needed for analysis. In parallel with the behavioral logs, we record a small set of quality signals: time-resolved coverage (whether sensing is active as scheduled), basic health checks (e.g., service alive, permissions intact), and lightweight anomaly flags that mark gaps, impossible values, or abrupt shifts. These quality indicators help detect issues early, guide re-collection when feasible, and qualify downstream analyses without altering the privacy-preserving scope of the data.

3.1.2 *iSense* (Android-native) App: Functional Flow and Screens

This subsection documents the participant-facing onboarding flow for the Android-native *iSense* app, using step-by-step instructions tied to concrete interface screens. It first describes how students download and install the package and then join the project by logging in with their Chinese name and student ID. It then details how they grant the necessary app permissions, activate the notification listener, and run the setup check so that all required services are authorized. Finally, it explains how to whitelist *iSense* in the system's battery-optimization settings, ensuring stable background sensing while still collecting metadata only and preserving privacy.

Step 1: Download and Install

We provide the installation package to participants in advance and make sure they know where it is stored so that they can easily locate the file on their device without searching. Clearly indicate the file name and, if possible, give a brief reminder of where it will appear (for example, in the "Downloads" folder or in a designated project folder). Then give clear, step-by-step guidance that walks them through extracting (unzipping) the files and completing the installation process from start to finish, so that even participants with limited technical experience can follow along. Explain precisely where they should save the package (for example, in a specific folder created for the study), how to open the compressed archive and view its

contents, and which file they need to tap or click in order to launch the installer. Instruct them to follow the on-screen prompts carefully and to continue running the installer until the setup finishes successfully and the application is fully installed on their device. Once installation is complete, they should be able to locate and recognize the application on their home screen or app list; the icons of our applications are shown in Figure 4.

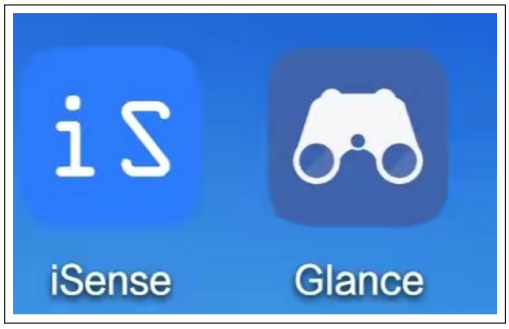


Figure 4: Icon of *iSense*.

Step 2: Join the Project (Chinese Name + Student ID)

On the *iSense* app's join-project login screen shown in Figure 5, participants first locate the two text inputs and type their Chinese name in the field labeled "Name" and their student ID in the field labeled "Student ID." At the top of the page, the header clearly reads "Log in to iSense," indicating that this screen is for joining the project by logging in. After entering both pieces of information, participants briefly review the two entries to make sure each field has been completed. Once they have confirmed that the "Name" and "Student ID" inputs are both filled, they tap the "Log In" button. This action submits the information and advances them to the next step in the join-project flow, continuing the process of accessing *iSense* through the specified login screen.

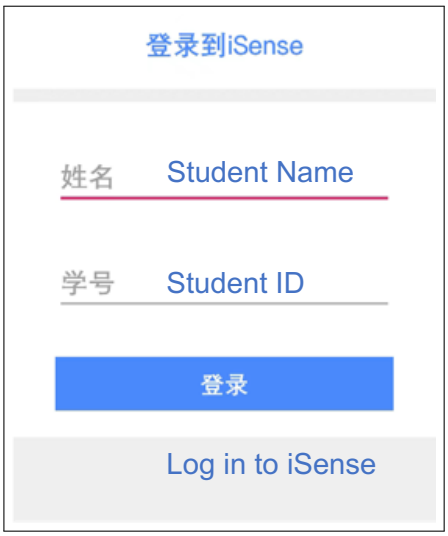


Figure 5: User interface for joining the project.

Step 3: Log in and Join the Target Project

After login, confirm the project name `jlu`, then tap *Search project* → *Join this project* → *Confirm*. Left to right: (1) "Search Project" screen with a query field ("jiu"), a "Search Project" button, and an empty results area; (2) project details showing Project ID "jiu," a brief description, and a "Join this project" button; (3) a confirmation dialog asking to confirm joining, with "Cancel" and "Confirm" actions. The interfaces are shown in Figure 6.

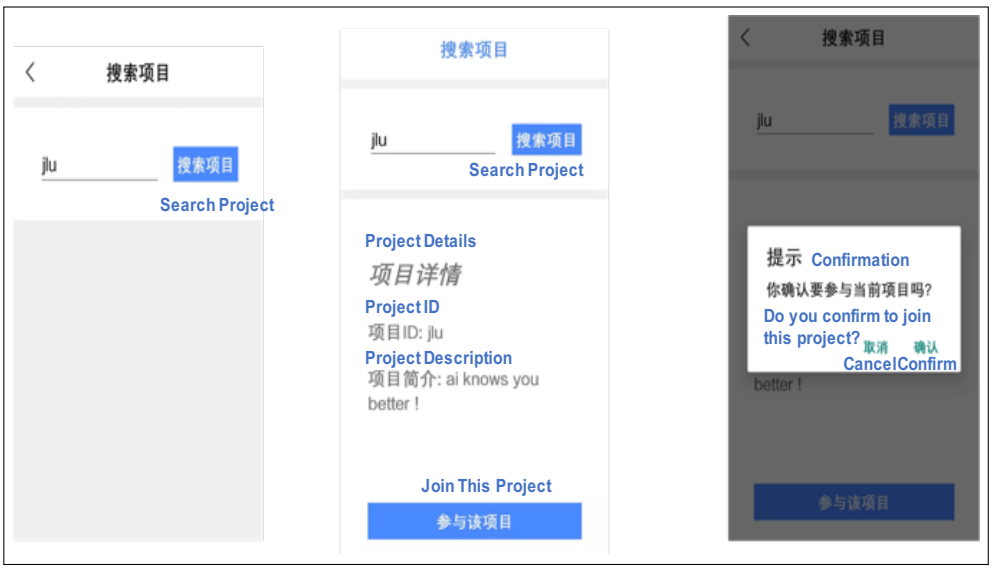


Figure 6: User interface for logging in.

Step 4: Grant Permissions

From *App Permission Management*, open the relevant system pages and enable the required permissions for the app: notification access, usage access, coarse location, and activity recognition. Turn on each item one by one until all four switches show as allowed. After confirming that these four permissions are enabled, return to *App Permission Management* to verify their authorized status. The interfaces of granting application permissions are shown in Figure 7 and the interfaces of granting notification access are shown in Figure 8 in the following parts.

Left Interface of Figure 7 - iSense Home ("Welcome to iSense"):

- *The screen displays the participant alias `mzy-30` in the account header, together with a hashed device/participant code.* This pair identifies the current user within the app while keeping personal details abstracted behind the hash, and it appears whenever the session or device needs to be referenced inside the system.
- *A vertically arranged menu presents the available actions in a clear list with one item per line: Search project*, *Sign out*, *App self-check*, *App Permission Management* (visually highlighted in red to draw the user's attention as the next required step), and *Real-time behavior recognition*. This

layout groups the core operations in a single place and makes it straightforward for the user to scan the menu and quickly navigate to the required action.

- *At the bottom of the screen, the toggle labeled Notification listener service is shown in the OFF position and marked "not authorized".* After the user grants the necessary usage access and notification access in system settings, this control can be enabled; once permissions are in place, the toggle will switch on and reflect the authorized state.



Figure 7: Granting required permissions (overview).

Right Interface of Figure 7 - App Permission Management (system settings): Pink switches are *ON* (granted); grey are *OFF*.

- *Basic permissions:* Basic permissions are configured as follows: *Send notifications* is set to *ON* to support the foreground service and status prompts; *Camera* is set to *ON*, but only metadata are used if needed and no image content is collected; *Storage* is set to *ON* to allow local buffering of data before secure upload; *Phone state* is set to *ON* so that coverage and uptime checks can be performed; *Motion sensors* are set to *ON* to enable activity recognition; and *Location* is set to *ON*, restricted to coarse or functional-zone information. By contrast, *Record audio* is set to *OFF*, meaning audio is not collected in this setup, and *Short Message Service (SMS)* is also set to *OFF*, so message content is not collected.

- *Special permissions:* Special permissions are configured as follows. The *System alert window* permission is set to *ON* so that the app can display critical prompts or overlays on top of other screens whenever user attention is required, for example when a mandatory setting is missing or a

background service needs to be restarted. The *Install apps* permission is set to *ON* to allow in-app updates when applicable, which means that newer, protocol-compatible versions of the client can be installed without asking participants to download and install a fresh package manually. *App usage access* is set to *ON* because this permission is required to collect app-usage statistics that summarize which apps are in the foreground and for how long, thereby providing the basic usage indicators needed in the study. The *Modify system settings* permission is set to *ON* to support operations such as placing the app on required whitelists and configuring background rules that would otherwise restrict continuous sensing. *Ignore battery optimizations* is set to *ON* in order to prevent the operating system from stopping or throttling background sensing through its battery-management routines during normal use. Finally, *Listen to system notifications* is set to *ON* so that the notification listener can operate as intended and record notification metadata, while still respecting the constraint that no message content or payload is accessed or stored.

Action Flow in Figure 7:

- *Open the iSense app and navigate to the home screen, then locate the left-side panel and tap the entry labeled App Permission Management (highlighted with a red box).* This opens the dedicated permissions area for the application. Use this option to proceed to the system pages where the required access can be granted.
- *On the permissions page, review the lists under both Basic and Special permissions, and toggle every required item to ON exactly as shown.* Confirm that each switch changes to the enabled state and that the interface indicates the permissions are allowed. Ensure no required permission remains off before leaving this page.
- *Return to the home screen and locate the Notification listener service control at the bottom of the interface, then switch the toggle to ON.* After enabling it, the status text should update from "not authorized" to "authorized". This change confirms that notification listening is active and the setting has been successfully applied.

Notes for Figure 7: Only variables necessary for the study are collected; no message, photo, or within-app content is stored. Location is processed at a coarse/functional-zone level. All data are de-identified and stored on private, access-controlled servers.

Left Interface of Figure 8—iSense Home:

- *The main menu displays the following options in a single, easy-to-scan list: Search project*, *Sign out*, *App self-check*, *App Permission Management*, and *Real-time behavior recognition*. These entries appear in the primary navigation area of the app and provide direct access to project lookup, account sign-out, diagnostic checks, permission settings, and live behavior functions, respectively. Users can tap any item to open the corresponding screen and proceed with the intended action.

- *At the bottom of the interface, a dedicated tile labeled Notification listener service shows the current status as (authorized), and its pink switch is clearly toggled to ON.* This layout makes it obvious that the service is enabled, with the colored switch confirming the active state. The "authorized" text and the *ON* position together indicate that notification listening is currently operational.
- *This status readout directly reflects whether the system-level Notification access permission for iSense has been granted.* When that Android permission is allowed, the app displays "authorized"; if the permission is revoked, the status would update to indicate it is no longer authorized. In short, the indicator mirrors the underlying operating-system authorization rather than an independent, app-only setting.



Figure 8: Granting notification access.

Right Interface of Figure 8—System "Notification Access":

- *In the Android system settings for notification access, the screen presents a list of applications that are allowed to read notification metadata.* On this page, the user can review which apps currently hold this authorization and manage the permission directly. The list functions as the system-level control panel for viewing and adjusting notification-access rights.
- *The toggles shown on this screen are: Desktop Launcher* set to *ON*, *Health* set to *ON*, *Google Play Services* set to *OFF*, and *iSense* set to *ON*. These *ON/OFF* positions indicate each app's current authorization status for reading notification metadata. In particular, *iSense* being *ON* confirms that notification access has been granted for the app.

Action Flow in Figure 8:

- *From the iSense home screen, locate the tile labeled Notification listener service and tap it to proceed.* This action opens the Android system *Notification access* settings page via a direct link from the app, taking you to the operating system's panel where notification-reading permission is managed. Use this system page to review and adjust which apps are allowed to access notification

information.

- *On the system Notification access page, find iSense in the list of applications and switch its toggle to ON.* If the system displays a confirmation dialog, read the prompt and select the option to allow or confirm the change. After accepting, ensure that the *iSense* entry clearly shows the enabled state so that the permission has been fully granted.
- *Return to the iSense app.* The tile now displays *Notification listener service (authorized)* with the switch visibly set to *ON*, confirming that the permission is active. This status indicates that the listener is connected and functioning, reflecting the successful authorization you just completed in the system settings.

Notes for Figure 8: Granting notification access allows *iSense* to read notification metadata only, and this access is strictly limited to technical fields such as the originating app or package identifier together with the timestamps associated with each notification event. In other words, the system can see which application generated a notification and at what time it was posted or updated, and these elements are then used to derive coverage and usage signals, for example to determine when notifications occur during the day and from which app they are sent. This information is processed at an aggregate, behavioral level and serves solely to characterize notification timing and frequency, rather than to inspect or reconstruct any personal communication content. No message content is collected at any point: the app does not access or store the bodies of messages, the textual content, or any attached payloads such as images, audio, or embedded media. The notification itself is treated only as a source of non-semantic metadata, and the substantive content that might appear to participants on the screen is neither read nor saved by *iSense*.

All records are de-identified and linked only to a study-specific research ID, so that the stored data cannot be directly associated with participants' real-world identities within the analysis environment. These de-identified records are kept under restricted access and are stored exclusively in controlled environments that are technically and administratively protected, and access to them is granted only to authorized personnel who are involved in the research and who operate under the applicable ethical and data-governance requirements.

Step 5: Complete the Setup Check

Carefully verify that all required service toggles within the app are set to *ON* and therefore fully enabled for all participants. To do this, participants first open *App Permission Management* from the main menu and slowly review each required switch one by one, checking that every entry in the list displays the enabled state before they move on to the next item. If any toggle is still off or appears disabled, participants turn it on immediately, wait briefly for the interface to refresh, and then return to the home screen to confirm that the corresponding status indicators have updated and now correctly reflect the authorized state. On the home screen after login, the interface prominently shows the participant profile and the

current project (ID *jiu*) at the top of the page, allowing verification that they are operating under the correct account and project context. Beneath this header, the main menu presents actions for questionnaires, leaving the project, diagnostics, permission settings, and real-time activity recognition, thereby offering direct and convenient access to each task without additional navigation. A dedicated status indicator on this screen confirms that the notification-listener service is granted and active, making clear at a glance that the listener is properly connected, authorized, and running for the participant's current session. The interface of permission and service status check is shown in Fgiure 9.



Figure 9: Permission and service status check.

Step 6: Finish System Settings (Battery Optimization Whitelist)

Set *iSense* to "Not optimized" in the Android Battery optimization settings so that it is explicitly excluded from the system's Doze and App Standby mechanisms. On a typical device, this involves opening Settings, navigating to Battery, and then selecting Battery optimization to reach the page where per-app power management is configured. Once there, change the view or filter from the default (which often shows only "Not optimized" or "Not allowed" apps) to All apps so that every installed application, including *iSense*, becomes visible in the list. After locating *iSense* in this list, open its entry and switch its status from Optimized (meaning it is subject to battery saving) to Not optimized, thereby instructing the system not to apply power-saving restrictions to this app; if the device displays a confirmation dialog when changing this setting, review the prompt and confirm the change so that the new status is applied. This exemption ensures that *iSense* is no longer automatically paused or constrained by Doze or App Standby, and therefore bypasses the usual background limits that Android may impose to conserve energy. As a result,

the app's background operation can proceed as intended, maintaining stable sensing and logging over time rather than being interrupted or delayed by aggressive power-management behavior. The process of adding *iSense* to the battery optimization whitelist is shown in Figure 10, which presents this procedure in a clear visual form for easy reference.



Figure 10: Add *iSense* to battery optimization whitelist.

Left Interface of Figure 10 - Open Settings: On the device's home screen, locate the Settings app icon and ensure it is visible. Tap Settings to enter the system configuration area, where you can review and modify power-management options that influence background sensing. From here, navigate to the relevant power controls, make the necessary adjustments so background sensing can run as intended, and confirm that your changes are applied. Once finished, return to the home screen to proceed with the remaining setup steps.

Right Interface of Figure 10 - Search "Battery Optimization": Inside *Settings*, tap the search bar at the top and type *Battery optimization* exactly. As you enter the phrase, the device displays matching results below the search field. In the results list, locate the item labeled *Battery optimization* and confirm that it refers to the system's power-management feature. Tap this *Battery optimization* entry to open the corresponding page. This page provides the pathway to the per-app optimization controls. From here, you can proceed to the screen that manages optimization status on an app-by-app basis.

Action Flow in Figure 10:

- *Launch Settings from the home screen by locating the system Settings icon and tapping it to open.* This takes you into the device's configuration area, which is required for changing power-management options. Make sure you are on the main Settings page before proceeding to the next steps.
- *In the search field within Settings, enter Battery optimization exactly and wait for the results to appear beneath the search bar.* When the *Battery Optimization* entry shows up in the results list, tap that result to open the corresponding system page. This is the shortcut that takes you directly

to the optimization controls.

- *On the battery optimization page, change the filter from the default view to* All apps *so that every installed application is visible.* Scroll through the full list to locate *iSense*, then open its entry (if needed) and set it to *Don't optimize* or *Allow background activity*. The exact label can vary by vendor or Android version, but choose the option that exempts *iSense* from optimization and permits continuous background operation.
- *If a confirmation dialog appears, approve the change so that iSense is whitelisted.* Verify that the status now indicates *Not optimized* or *Allowed*, confirming the exemption. This ensures *iSense* can run continuously in the background without being paused by Doze/App Standby or vendor-specific power-saving rules.

Notes for Figure 10: Whitelisting prevents aggressive power management from stopping background logging. This setting does *not* grant additional data access; it only preserves *iSense*'s background operation for reliable sensing.

When participants open the Battery optimization settings menu, they locate the *iSense* application and manually add this app to the battery optimization whitelist, thereby turning off battery optimization for *iSense* by setting it to not allowed. The corresponding system interfaces for this procedure are shown in Figure 11.

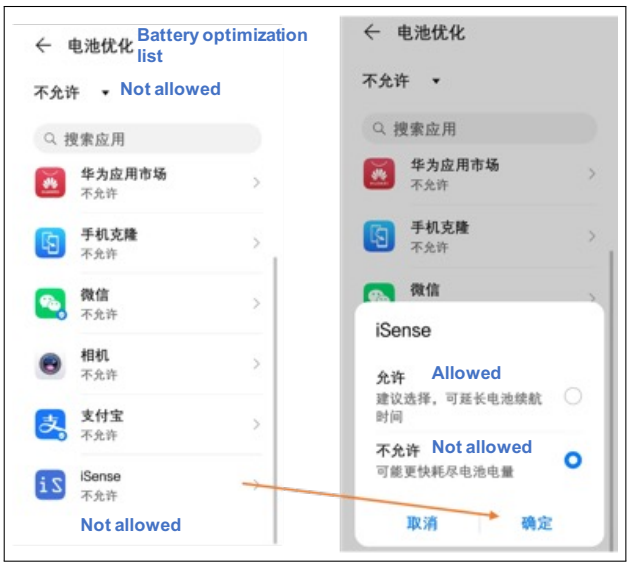


Figure 11: Locate *iSense* in the "Not allowed" list and set to "Not optimized".

Left Interface of Figure 11—Battery optimization list: In *Settings* → *Battery optimization*, confirm that the filter control at the top of the screen is set to *Not allowed*. This view lists only those apps that are currently *exempt* from battery optimization and therefore permitted to run with fewer power-management restrictions. Scroll through this list and locate *iSense*; when the exemption is active, its status label also reads "not optimized," indicating that Doze/App Standby will not limit its background tasks. Seeing *iSense*

in this filtered view, together with the "not optimized" label, confirms that the app is correctly whitelisted on the device.

Right Interface of Figure 11 - Change optimization for *iSense*: Tap *iSense* to open the system's choice dialog for per-app optimization. The dialog presents two mutually exclusive options that control power management for this app:

- *Allow* (optimize battery; may limit background activity and shorten sensing time). Choosing this setting enables battery optimization for *iSense*, allowing the system to pause or restrict background work to save power. As a result, background sensing periods can be shortened, and scheduled or event-driven tasks may be delayed or skipped.
- *Not allowed* (do *not* optimize; allows persistent background operation). Selecting this setting keeps *iSense* out of battery optimization, preserving continuous background behavior. With optimization disabled for this app, sensing services can remain active for longer and are less likely to be interrupted by power-saving routines.

Select *Not allowed* and press *Confirm*. After confirming, the dialog closes and the list updates to show that *iSense* is excluded from optimization, typically reflected by the "not optimized" label in the battery optimization list.

Action Flow in Figure 11:

- *Open Settings and navigate to Battery optimization using the list or the search function.* This takes you to the system page that determines whether each installed app is subject to Doze/App Standby controls and other power-saving behavior, where optimization can be managed on a per-app basis.
- *Ensure the filter at the top is set to Not allowed, then scroll the list to locate iSense.* This view displays only apps already exempted from optimization, so the presence of *iSense* here indicates it is currently whitelisted. Once found, check that its status label reads "not optimized," confirming the exemption is active in the current configuration.
- *Tap iSense to open the choice dialog, select Not allowed, and then tap Confirm to save.* The system applies the change and returns you to the list view; *iSense* should now display the status "not optimized," which verifies that battery optimization is disabled for this app.

Notes for Figure 11: Wording may vary by device vendor (e.g., *Don't optimize*, *Allow background activity*, or *Unrestricted*), but all refer to disabling battery optimization for the selected app. This selection only disables battery optimization for *iSense*; it does not change any data-access permissions such as notification access, usage access, location, or activity recognition. In other words, power-management behavior is adjusted for *iSense* without altering its privacy or permission settings.

3.1.3 Backend and Version Compatibility (from *isense-console* source)

The console stack is Vue 2 + Vite + MDUI; build artifacts are uploaded with `ossutil` to object storage. The File/OSS gateway supports actions such as `makeDir`, `listFiles`, `getTextFile`, `putTextFile`, and `putBinFile`, and organizes data by project/user paths. The MongoDB gateway supports `find`, `findOne`, `insert`, `update`, and `aggregate`. A common Database (DB) `isense` and per-project DBs `{projectId}` are used; the per-project DB contains `isense-user-realtime-status` with fields `{userId, timestamp, realtimeLocation{lat,lng,name}}`. Questionnaire push logs can be exported as NDJSON (newline-delimited JSON). A script fetches the latest client SDK from the server to keep protocols in sync. Data are stored in a private Aliyun (Alibaba Cloud) OSS purchased for this research, with isolation by project/user paths and optional lifecycle rules and versioning.

3.1.4 Power, Privacy, and Security by Design

We prefer event-driven capture, use batch/periodic uploads with offline retry, and use coarse location when possible. We require the battery-optimization exemption for stable background logging. We apply data minimization and de-identification: collect metadata only; hash or map Wi-Fi/Bluetooth Low Energy (BLE) identifiers to functional zones; grid or blur location; use a de-identified research ID mapping. For transport and storage, all client-to-gateway calls use Hypertext Transfer Protocol Secure (HTTPS)/Transport Layer Security (TLS); Object Storage Service (OSS) paths are isolated by project/user; only de-identified derivatives are used in analysis; raw identifiable data are accessible only in authorized environments. Access control and audit follow least privilege and role-based access with confirmation for sensitive actions and audit logs at the gateway and console. Retention and deletion: raw identifiable data are deleted after the minimum useful period; withdrawals trigger deletion; only de-identified/aggregated results are shared.

3.2 Data Collection

This section summarizes the end-to-end design of our co-period data collection framework. It first introduces the weekly intrinsic motivation scale and its release cadence, which are jointly collected with passive smartphone sensing over the same time windows. It then details the smartphone sensing modalities and sampling configurations that define the default logging setup for later quality control, temporal alignment, and feature construction. Finally, it formalizes the standard operating procedure from recruitment and eligibility checks through deployment, monitoring, data ingestion, and withdrawal and deletion, emphasizing metadata-only collection, privacy protection, and reproducible handling of logs.

3.2.1 Scales and Cadence (Co-period Collection)

Informed consent (ICF): We obtained written or electronic informed consent from all participants before any data collection began, prior to the start of any study session or procedure. In practice, each participant first received and read the consent information, then provided their written or electronic agreement, and only after this explicit consent had been recorded we initiate any study-related activities or start logging

any data for that participant.

Self-report Scales: Intrinsic learning motivation was measured with items adapted from the Self-Regulation Questionnaire—Academic (SRQ-A; 1–5 scale) for Chinese university students. In collaboration with psychology experts, we further adapted and translated a higher-education–appropriate Chinese version specifically tailored to university students, translated and refined via a small pilot. The questionnaire we refined and adapted is shown in Table 2. Psychometric properties met accepted standards for reliability and validity, and the instrument content was reviewed and endorsed by our collaborating psychology experts. Originals, translations, item mappings, and scoring rules are provided in this chapter's appendix We release the online questionnaire once per week on a survey platform and allow one day for responses; for example, Friday 00:00 to Saturday 24:00 is one full collection window. We adopt a weekly fixed cycle to accommodate varied course schedules. Each submission time t is the anchor. We construct symmetric or causal windows W around t. By default, the window is 1 day (day-aligned). Passive smartphone sensing (co-period). During the same period as the questionnaire, the smartphone continuously logs daily behaviors and context. These above procedures constitute a common three-step framework applied across the subsequent three staged studies; step-specific implementation details are elaborated in later sections.

Table 2: Intrinsic Learning Motivation Items

Item
1. Why do I do my homework? Because it is fun.
2. Why do I do my homework? Because I enjoy doing my homework.
3. Why do I work on my classwork? Because it is fun.
4. Why do I work on my classwork? Because I enjoy doing my classwork.
5. Why do I try to answer hard questions in class? Because I enjoy answering hard questions.
6. Why do I try to answer hard questions in class? Because it is fun to answer hard questions.
7. Why do I try to do well in school? Because I enjoy doing my schoolwork well.

Response format: 5-point Likert scale — 1 = Not at all true; 2 = Not very true; 3 = Somewhat true; 4 = Mostly true; 5 = Completely true.

3.2.2 Sensor Lists

This section introduces the smartphone sensing modalities and sampling configurations used in our studies. In the three staged studies, we selectively enabled different modality sets based on practical field-deployment constraints and observed data quality. Table 3 below lists event-driven (passive-triggered) modalities (e.g., foreground app, notifications, unlock/lock, connectivity) with their triggers, Android APIs, key metadata fields, sampling granularity, and typical uses. Table 4 below lists active-triggered (periodic/streaming) sensors (e.g., accelerometer, gyroscope, light, location) with their APIs, example fields, physical units, and brief preprocessing notes. We collect metadata only (no content) to reduce privacy risk and energy cost, and these tables define the default sensing setup for later quality control, temporal

alignment, and feature construction.

Table 3: Event-driven modalities (API/fields/privacy)

Modality	**Trigger**	**Source/API**	**Key fields (examples)**	**Gran.**	**Notes / Typical use**
Foreground app	event and periodic	`Usage_Stats_Manager`	pkg; fg start/end; total fg time	s/min	Category aggregation; self-control proxy
SMS metadata	event	`Telephony/\gls{SMS}`	sent/recv; ts; peer hash	count	No content; social intensity proxy
Notifications	event	`Notification_ Listener`	app category; ts	count	No content; interruptions
Unlock/lock	event	`USER_PRESENT/ SCREEN_OFF`	unlock/lock ts	count	Merge to sessions; rhythm
Airplane mode	event	`system_broadcast`	on/off	state	Debounce
Battery	event	`BATTERY_CHANGED`	level%; charging state	%/state	Sessionize; routine
Nearby BLE	periodic	`Bluetooth/ \gls{BLE}scan`	device count/unique	count	Filter static devices
Device sleep	event	`screensleep`	start/end	count	Align with sessions
Headphones	event	`AudioManager`	on/off; type	state	Debounce; day/night split
Media playback	event and periodic	`MediaSession/ Playback`	play/pause; duration	count/s	Dedup; night split
Connectivity	event	`ConnectivityManager`	Wi-Fi/cellular/offline	state	Debounce
Cells nearby	periodic	`Telephony. getAllCellInfo`	cell ID; RSSI	rows	Coarse mobility
Wi-Fi nearby	periodic	`WifiManager. scanResults`	SSID/BSSID/RSSI	rows	Hash BSSID; place stay
Call metadata	event	`Phone_State_ Listener`	in/out; duration	count/s	No content; social intensity

Continued on next page

Modality	**Trigger**	**Source/API**	**Key fields (examples)**	**Gran.**	**Notes / Typical use**
Boot/shutdown	event	`BOOT_COMPLETED/ SHUTDOWN`	type; ts	count	Device stability
Ringer mode	event	`AudioManager`	ring/vibrate/silent	state	Debounce
User present	event	`screen/user-present`	active/inactive	state	Merge with sessions

Above Table 3 summarizes the event-driven (passive-triggered) smartphone sensing modalities used in our studies. They include foreground app usage, SMS metadata, notifications, screen unlock/lock, airplane-mode state, battery level/charging, nearby Bluetooth Low Energy (BLE) scans, device sleep, headphone connection, media playback, connectivity transitions, nearby cellular towers, nearby Wi-Fi access points, call metadata, boot/shutdown, ringer mode, and user-present events. For each modality we record metadata only (no content), apply debouncing or sessionization where appropriate, and derive features for quality control, temporal alignment, and behavioral interpretation as detailed below.

Foreground App: We log package name, foreground start/end, and accumulated foreground time at second/minute granularity. Events are sessionized (merge short gaps) and aggregated by app and by category (e.g., social, education, games). Typical uses: daily/weekly focus vs. distraction, study-tool time, and self-control proxies (sustained use vs. rapid app switching).

SMS Metadata: We record sent/received flags, timestamps, and a hashed peer identifier (no content). Features include counts per window, sent/received balance, unique peers, and diurnal rhythms. Use: social-interaction intensity and regularity without inspecting message contents.

Notifications: We capture app category and timestamps for each posted notification as it appears on the device. Based on these records, we compute notification rate, burstiness, and time-of-day distribution to quantify interruption load and attention fragmentation over the observation period and to describe how notifications are spread across different times of day. No titles or bodies are stored at any point; only minimal metadata are retained, so the semantic content of notifications remains private while still allowing us to characterize their frequency, timing, and basic source categories.

Unlock/Lock: Each unlock/lock marks a screen session and therefore defines a period of active device use. We compute unlock counts, mean session length, and inter-unlock intervals to characterize usage rhythm and possible compulsivity, and to align other events to active screen time so that behavioral signals can be interpreted in relation to when the screen is actually in use. In effect, these metrics summarize how often participants return to the device, how long they tend to stay on each time, and how their interaction windows structure the rest of the sensed events.

Airplane Mode: On/off transitions are debounced to avoid flapping when the mode is toggled rapidly. These events explain connectivity gaps and prevent misclassifying “no data” periods as sensing failure,

since they indicate times when the device was deliberately taken offline. In this way, airplane-mode status acts as a safeguard in the quality-control pipeline and helps distinguish intentional disconnection from technical problems or logging errors.

Battery: We track percentage level and charging state whenever the system reports a battery update. Features include charge cycles, time-on-charge, and discharge rate during active use versus idle, which are useful for routine characterization and for flagging anomalous device states that may compromise sensing stability. Together, these indicators provide a coarse picture of how the phone is used and charged across the day and help interpret potential gaps that arise from shutdowns or very low battery levels.

Nearby BLE: We log counts and uniques of discoverable devices (no MAC content is stored), capturing only anonymized presence information. After filtering static beacons, co-presence density and volatility approximate local crowding or social proximity in a privacy-preserving way without identifying individual people or devices. This allows us to infer whether participants are in more socially dense or more solitary environments while keeping the underlying identities fully obscured.

Device Sleep: We record device-sleep start and end times when the screen enters and exits long idle states. Combined with unlock and usage sessions, this delineates night versus day and aligns "quiet time" windows for other signals that may also drop during sleep. As a result, we can distinguish expected low-activity periods (such as night-time sleep) from unexpected gaps and interpret other behavioral streams within a coherent daily rest–activity pattern.

Headphones: We record on/off status and type (wired/Bluetooth) each time the headphone connection changes. Debounced events, split by day and night, indicate likely audio-listening contexts (music, podcasts, or lectures) without logging any audio content or media titles. These patterns help us understand when participants are engaged in audio-based activities in general terms, while still preserving full privacy regarding what they are actually listening to.

Media Playback: We capture play/pause events and playback durations, and duplicates from multiple players are deduplicated so that overlapping sessions are not double-counted. Day and night splits separate leisure from study playback in a coarse way and support analyses of audio/video engagement time and its coupling with other activities or contexts. In this manner, media-playback logs provide a high-level view of when participants use media without exposing the specific content consumed.

Connectivity: Transitions among Wi-Fi, cellular, and offline states are recorded as discrete states with debouncing to avoid rapid toggling noise. Connectivity state supports upload scheduling, location inference (for example, recognizing common campus Wi-Fi networks), and interpreting missing data that occur during offline periods. It therefore functions both as an explanatory signal for coverage patterns and as a contextual cue for where and how the device is connected.

Cells Nearby: We store cell IDs and RSSI per scan cycle for all observed towers. Tower changes and

signal variability offer coarse mobility and dwell estimation when GPS is unavailable or disabled, providing a lower-resolution alternative for tracking movement. These cellular indicators help us differentiate between stationary periods and broader movement trajectories while maintaining a relatively low privacy risk compared with precise location tracking.

Wi-Fi Nearby: We log SSID, BSSID, and RSSI for nearby networks at regular intervals so that the device can characterize its surrounding wireless environment over time. BSSIDs are immediately hashed so that raw hardware identifiers are not retained in the dataset, which reduces re-identification risk while still allowing us to track whether the same access points are present across scans. Stable BSSID sets indicate place-stays, because the phone continues to see the same collection of nearby routers, while transitions in the observed sets indicate movement between semantic locations (for example, dorm → library or dorm → classroom) without revealing exact street addresses or room numbers. This representation allows us to infer meaningful location routines, such as repeated visits to study spaces or cafeterias, and to detect context changes across the day in a privacy-preserving and device-agnostic way that works across different handset models and operating-system versions.

Call Metadata: We record incoming and outgoing flags and call durations as event-level metadata, but no audio or conversational content is ever captured or stored. From these logs, features include total call time per window, mean call length, and diurnal calling patterns, which act as proxies for synchronous social interaction and availability without exposing the substance of conversations. These summary indicators help characterize how often and when participants engage in voice communication, for example whether they regularly call others in the evening or during breaks, while preserving the confidentiality of whom they call and what they say. In this way, the phone call features provide a coarse but informative view of social-contact patterns that respects privacy and remains suitable for large-scale, ethically governed sensing studies.

Boot/Shutdown: We mark device restarts and shutdowns to gauge device stability, operating-system updates, and potential sensing gaps that occur when the device is off. These flags feed the quality-control pipeline and help distinguish gaps caused by the phone being powered down from gaps due to software errors or permission issues. In turn, this supports more accurate interpretation of coverage and reduces the risk of misattributing missing data to participant behavior.

Ringer Mode: Changes among ring, vibrate, and silent modes (debounced to avoid rapid toggling noise) signal contextual norms, such as setting the phone to silent in class or in the library. These mode transitions are often used to modulate expectations about notifications and foreground-app behavior, since they reflect whether participants are likely trying to reduce audible interruptions. Thus, ringer-mode patterns provide an additional contextual layer when interpreting usage and notification streams.

User Present: This “active user” gate indicates that the device is attended and that a user is physically interacting with it. We merge it with screen-session logic to align other events to user-active periods and to

refine session boundaries so that analyses focus on times when the participant is plausibly engaged. In combination with unlock and usage signals, this event helps distinguish between background processes and genuine, user-driven interaction with the device.

Table 4 lists the active-triggered (periodic/streaming) sensors used in our studies. For each sensor, we collect numeric signals only (no audio, image, or text content), apply light preprocessing (e.g., smoothing, debouncing, gravity removal, coordinate fusion), and align samples to fixed-length behavioral windows for feature construction (mobility, activity, environment). Units and typical uses match the table entries.

Table 4: Active-triggered sensors (API/fields/privacy)

Modality	**Trigger**	**Source/API**	**Key fields (examples)**	**Unit**	**Notes / Typical use**
Accelerometer	periodic	`TYPE_ACCELEROMETER`	x,y,z accel (incl. gravity)	m/s^2	Remove gravity; micro-activity
Ambient temperature	periodic	`TYPE_AMBIENT_TEMPERATURE`	temperature	°C	Smoothing; environment
Location	period and event	`Fused_Location_Provider`	lat, lon, accuracy	m	Grid/zone mapping; mobility
Gravity	periodic	`TYPE_GRAVITY`	x,y,z gravity	m/s^2	Fuse with accel for posture
Gyroscope	periodic	`TYPE_GYROSCOPE`	x,y,z angular vel.	rad/s	Denoise; micro-motion
Light	periodic	`TYPE_LIGHT`	illuminance	lx	Clip extremes; day–night
Linear accel.	periodic	`TYPE_LINEAR_ACCELERATION`	x,y,z (no gravity)	m/s^2	Window energy/-variance
Magnetic field	periodic	`TYPE_MAGNETIC_FIELD`	x,y,z geomagnetic	μT	Calibration (opt.), heading
Orientation (derived)	periodic	`from_rotation_vector`	yaw/pitch/roll	deg	From fusion; posture
Pressure	periodic	`TYPE_PRESSURE`	pressure	hPa/mbar	Smoothing; altitude approx.
Proximity	event and periodic	`TYPE_PROXIMITY`	distance	cm	Ear-on/holding with calls
Humidity	periodic	`TYPE_RELATIVE_HUMIDITY`	relative humidity	%	Smoothing; environment

Continued on next page

Modality	**Trigger**	**Source/API**	**Key fields (examples)**	**Unit**	**Notes / Typical use**
Rotation vector	periodic	`TYPE_ROTATION_VECTOR`	quaternion / DCM	—	Orientation solving/fusion

Accelerometer: We record tri-axial acceleration (including gravity), that is, acceleration along the x, y, and z axes as reported by the device's built-in motion sensor at regular sampling intervals. Preprocessing removes gravity via high-pass filtering or by pairing with the gravity/linear-acceleration streams, so that the remaining signal more closely reflects actual movement rather than static orientation of the phone. Features include magnitude and variance of the acceleration vector, overall energy in each window, jerk-based change measures, and spectral peaks for step frequency, all computed over fixed or sliding windows. Uses: micro-activity detection, short posture-change bursts, and general physical activity level, providing a compact summary of how intensely and how often the device (and thus the participant) is in motion within a given time span.

Ambient Temperature: Single-channel temperature is smoothed (median or low-pass filtering) to reduce sensor jitter and short transients that arise from brief contact changes or local fluctuations near the device. This produces a more stable estimate of the surrounding thermal environment that can be interpreted at the level of minutes or hours rather than second-to-second spikes. Uses: coarse environmental context (hot/cold episodes), often combined with light and humidity to infer indoor versus outdoor conditions and general comfort patterns over the day, without attempting to identify the exact room or location.

Location: We store latitude, longitude, and accuracy for each obtained location fix, keeping track of how precise each sample is according to the fused provider. We apply duty-cycling and accuracy screening to avoid excessive battery use and to drop very noisy points, and raw fixes are snapped into grids or functional zones to protect privacy while preserving coarse mobility information. Features include total distance traveled, radius of gyration, place-stay duration, and transitions among semantic zones (e.g., dorm, library, classroom), which together describe broad movement patterns and time spent in key study-relevant locations.

Gravity: The gravity vector in the device frame is fused with accelerometer readings to separate posture or tilt from translational movement, allowing us to distinguish static orientation from actual motion. By combining these sources, we obtain a clearer description of how the phone is held or positioned over time. Uses: phone orientation (lying, standing, tilted) and posture-aware activity cues, which refine the interpretation of accelerometer-based features in terms of whether the device is stationary, being carried, or being actively manipulated.

Gyroscope: Tri-axial angular velocity is denoised and windowed so that short spikes or sensor noise do not dominate the signal, and the resulting segments can be more reliably summarized. Features include

rotation energy, variance, and the frequency of short rotational bursts, which complement accelerometer-based measures to detect micro-motions, phone handling patterns, and gait-related oscillations. These derived indicators are used as additional cues about how the device is being rotated or moved in three-dimensional space.

Light: We clip outliers and smooth the raw illuminance readings to stabilize fast flicker or abrupt spikes that are not meaningful at the behavioral level. This generates a more interpretable light profile that can be related to daily cycles and typical study settings. Uses: day–night segmentation, circadian regularity, and contextual cues for indoor study locations (e.g., library vs. dorm), supporting coarse distinctions between bright, well-lit study environments and dimmer rest or off-task periods.

Linear Acceleration: Gravity-free acceleration is used to compute window energy, variance, peak counts, and step-related features with less tilt contamination than raw accelerometer signals, because the gravitational component has already been removed. This yields a cleaner representation of true linear movement of the device within each time window. Uses: locomotion intensity and short movement bursts, helping to characterize when the participant is relatively still, moderately active, or engaged in more vigorous physical movement.

Magnetic Field: We record tri-axial geomagnetic strength as measured by the phone's magnetometer, capturing the magnetic field along three axes at each sample. Optional calibration (soft/hard-iron correction) improves heading stability and reduces systematic distortions caused by the device or nearby metallic objects. Uses: heading estimation with gyro fusion and coarse indoor magnetic "fingerprints" to support place inference at a level that distinguishes broad zones without identifying exact coordinates.

Orientation (derived): Yaw, pitch, and roll are derived via sensor fusion (e.g., combining the rotation vector with magnetometer and gyro data) to obtain a stable orientation estimate in degrees. This fused orientation describes how the device is rotated relative to a fixed reference frame at each sampling time. Uses: posture classification, phone-holding angle, and orientation normalization for other motion features, so that activity measures can be made more comparable across different ways of holding or placing the phone.

Pressure: Smoothed barometric pressure provides an altitude proxy, because small changes in pressure correspond to changes in height above sea level or floor within a building. This processed signal filters out rapid spikes and focuses on slower shifts that indicate meaningful vertical movement. Uses: floor-change and elevator events, indoor/outdoor transitions, and refinement of mobility trajectories, especially in multi-storey environments where vertical movement is behaviorally relevant.

Proximity: Near/far (or distance) samples indicate ear-on, handheld, or pocket states depending on how close an object is to the proximity sensor at a given moment. By interpreting these distance readings in context, we can infer how the phone is being carried or used. Uses: call-related handling patterns, simple screen-attention heuristics, and filtering of spurious motion signals when the device is pocketed or held

close to the body rather than actively viewed.

Humidity: Smoothed relative humidity complements temperature and light for environmental characterization, yielding a more complete picture of the ambient conditions in which the device is used. The smoothing step reduces rapid noise and focuses on broader humidity changes across the day. Uses: indoor climate patterns (e.g., dorm vs. classroom) and comfort routines, contributing to coarse distinctions between different everyday settings without revealing specific locations.

Rotation Vector: We store the fused orientation representation (quaternions or direction-cosine matrix), which encodes the device's rotation in a compact mathematical form suitable for further processing. This representation supports robust heading estimation and frame transformations across time. Uses: stable orientation solving, drift reduction when combined with gyro signals, and orientation-invariant motion features that remain comparable even as the phone is rotated in different ways during normal use.

3.2.3 Standard Operating Procedure

Recruitment & Eligibility: Issue and collect informed consent (ICF); check inclusion/exclusion; perform device self-check (Android version/ROM/storage/network); record the institutional review board (IRB) number; complete a baseline questionnaire with course/exam flags. In the app, participants log in with Chinese name + student ID (used as `userId`). In practice, this means that each student is first formally informed about the study and signs the ICF, then is screened against predefined eligibility criteria and has their phone briefly checked to confirm that the operating system and resources meet the sensing requirements. The IRB approval number is documented alongside these steps, and a short baseline form is completed to register course enrollment and upcoming assessment periods. Finally, the login using Chinese name and student ID creates a consistent `userId` that links app data to the baseline record under controlled conditions.

Deployment & Onboarding: Follow the 6-step guide in §3.1.2: install → join project (name + ID) → log in and join `"jlu"` → grant notification/usage/location & activity permissions → pass setup check → set *iSense* to *not optimized* under battery optimization. This stabilizes event/periodic logging and reduces Doze-related gaps. Concretely, participants are guided through installing the app, registering themselves with the correct project label and personal identifier, and joining the `"jlu"` project space so that their data are routed to the proper study. They then grant the required permissions for notifications, usage statistics, location, and activity recognition, run a built-in setup check to ensure all modules are active, and adjust system battery settings so that *iSense* is exempt from aggressive power saving. Together, these steps ensure that passive sensing runs reliably in the background throughout the observation period.

Operation & Monitoring: Run passive sensing co-period with the weekly questionnaire; enable local caching and retry on-device; on server, maintain statistics maps and real-time status; the research team reviews event counts, last-seen time, and recent location in the console. In other words, the sensing app and the weekly self-report schedule are configured to overlap in time, while the app itself temporarily stores

logs locally and retries upload when connectivity is unstable. On the server side, dashboards summarize coverage and device status in near real time, so that the research team can periodically check whether participants are still active, whether recent data have been received, and whether sensing streams remain stable over the study.

Data Ingestion: Receive logs via Aliyun OSS and data gateways; organize by project/user path; pull and verify daily increments; all I/O is audited. Incoming files are thus transferred through the configured gateways into Aliyun OSS, stored under a hierarchical directory structure keyed by project and user, and then retrieved in daily batches for aggregation. Each import is checked for completeness and integrity before being added to the analysis dataset, and all read/write operations are recorded in an access log so that data movement remains traceable throughout the pipeline.

Withdrawal & Deletion: Trigger withdrawal in the console; immediately revoke credentials and start raw-data deletion; update audit records; complete OSS lifecycle cleanup. When a participant withdraws, the research team uses the management console to disable their access token, initiate automated removal of their raw logs from storage, and record the withdrawal event in the audit trail. The OSS lifecycle rules then finalize cleanup of any remaining artifacts, ensuring that the participant's data are no longer retained beyond the point of withdrawal in accordance with the consent terms.

3.3 Preprocessing / Reprocessing

In order to build an analyzable *subject* $\times$ *window* dataset from raw logs and generate aligned artifacts (timestamps, normalization parameters, coverage reports). This section follows the preprocessing layer defined in Chapter 1 and specifies the concrete steps and conventions that make later analyses comparable across participants, time windows, and studies. All intermediate products (aligned timestamps, fitted scalers, coverage/quality summaries) are versioned and retained to enable reproducibility and auditability throughout the dissertation.

Data Intake & Normalization: Unify first, then align and aggregate. Ingest phone logs and scale scores; unify subject identifiers (use student ID as `userId`) and display names (Chinese name); convert all timestamps to a single time zone (for example, Asia/Shanghai) to ensure mergeability. This unification step establishes a single source of truth for identifiers and time, so that subsequent alignment and aggregation operate on consistent keys. Raw records retain their original fields, but downstream processing always references the normalized `userId`, canonical display name, and normalized timestamps to avoid ambiguities or duplicate joins.

De-identification: Anonymize before use; restrict raw access. Map student ID/name to a research ID; grid or zone-map locations; keep only hashes/counts for Wi-Fi/BLE; protect raw logs with access control and audits. De-identified keys are used for all analysis tables, while the mapping between personal identifiers and research IDs is held separately under restricted permissions. Location generalization (grid/zone) and identifier hashing are applied prior to feature computation, so downstream artifacts never

expose fine-grained or linkable identifiers. Access events and audit trails document when protected data are touched.

Temporal Alignment: Align streams to a common window for comparison with the scale. Use the questionnaire submission time t as the anchor; define symmetric windows $[t - W, t + W]$ or causal windows $[t - W, t]$; resample/align streams with different rates and persist aligned timestamps. The cadence matches the weekly schedule, with default $W = 1$ day (day-aligned). All sensor/event streams are snapped to the chosen window relative to t so that behavioral summaries describe exactly the same interval used by the psychometric measure. We persist the exact aligned time grids to make re-derivation of features and cross-checks straightforward.

Quality & Coverage: Check quantity and stability first; weight or exclude if needed. Compute window-level coverage (for example, valid minutes / window minutes) and event distributions; set a minimum coverage rule; label gaps due to missing permissions or Doze kills and apply weighting/exclusion rules. Default: at least one valid reading per 5-minute bin. Coverage summaries and reason codes (e.g., permission revoked, device idle) accompany each window and are propagated to modeling. Windows that fail the minimum rule are either down-weighted or removed, and these decisions are recorded alongside model inputs to support sensitivity analyses.

Cleaning & Harmonization: Deduplicate, fix time, unify schema. Remove duplicates; fix timestamp order and clock drift; standardize event schema and units (seconds/minutes/state); sessionize short blips (for example, unlock $\longrightarrow$ usage session). Harmonization produces a consistent set of columns and units across sources so that aggregation logic is uniform. Sessionization consolidates rapid, fragmented events into interpretable episodes while preserving start/end times and durations for later feature extraction.

Missingness Handling: Diagnose first, then impute cautiously. Determine MCAR/MAR/MNAR; prefer recovery via on-device offline cache and retry; if needed, apply conservative imputation and record rationale and scope. Missingness diagnostics are stored at the window level and inform whether a window is eligible for analysis, requires down-weighting, or should be flagged for robustness checks. Any imputation method used is documented with its parameters and the exact subset of features/windows affected.

Scaling (Normalization Only): Fit normalization on training folds to avoid leakage. Fit parameters (mean/SD or quantiles) only on training folds; reuse them for validation/test; persist per-fold parameters with versioning. Transformations for held-out data apply the frozen parameters from the corresponding training fold, ensuring no information leaks across splits. Normalization metadata (parameter values, fold identifiers, timestamp of fitting) are saved to allow exact re-creation of model inputs.

Aggregation & Join: Summarize events into features, then merge with the scale. Aggregate events into subject $\times$ window features (sessions, durations, ratios, rhythms). At anchor t, merge sensing features with scale scores into a versioned dataset, keeping generation time, parameters, and an audit trail. The join uses the normalized `userId` and the aligned window index, guaranteeing that behavioral summaries

and psychometric measures refer to the same temporal slice. Feature dictionaries and generation settings are stored with the dataset release.

Interfaces to Chapters 4–6 Chapter 4 (RQ1) uses the §3.2–§3.3 outputs for behavior–motivation associations, for example GEE (generalized estimating equations) [67, 68], controlling FDR with TSBH (Two-Stage Benjamini–Hochberg) [69]. The aligned windows, coverage rules, and normalized features provide the comparable inputs that GEE requires, while TSBH is applied to control multiplicity. For Chapter 5 (RQ2), within the §3.3 feature families, run clustering/factor stability and between-group/within-person comparisons; assess interpretability and robustness. The standardized feature tables and window definitions enable consistent comparison across strata and time. For Chapter 6 (RQ3), from the windows and features built in §3.3, textualize and prompt LLMs, use CoT (chain-of-thought) constraints to improve explanation consistency, and follow grouped splits and anti-leakage rules, relying on the same normalization and alignment artifacts to keep inputs synchronized.

Defaults and Deviations:

- *Defaults:* $W = 1$ day (day-aligned); observation window $L = 1$ day; minimum coverage rule: one valid reading per 5-minute bin; normalization parameters are fit only on training folds and persisted. Unless otherwise stated, all analyses in later chapters inherit these defaults for windowing, coverage, and scaling, together with the identifier/time unification and de-identification policies described above.
- *Deviations:* The data-collection and preprocessing procedures here are the shared defaults for all subsequent studies. Any deviations will be stated and explained in a dedicated "Differences from Chapter 3" subsection of the corresponding chapter. Such notes identify exactly which steps differ (e.g., window shape, inclusion criteria, or feature sets) and provide the justification, while keeping all other steps consistent with the defaults.

4

Association Analysis of Daily Behaviors and Intrinsic Learning Motivation

This chapter describes the *study at association stage* in the dissertation's three-stage program (from association to optimization to estimation), operationalizing the hybrid psychometric approach introduced earlier: aligning SDT-grounded self-reports with passive smartphone sensing to assess students' *intrinsic learning motivation* in authentic contexts. Focusing on the *social* facet of behavior, we examine whether and how smartphone-derived social behavioral patterns relate to intrinsic learning motivation.

Within the five-layer architectural workflow (from Perception to Preprocessing to Representation to Analysis to Reporting), this study realizes concurrent collection of weekly SDT measures and passive logs (Perception), standardized preprocessing with temporal alignment and coverage checks (Preprocessing), construction of app-usage features and low-dimensional behavioral patterns (Representation), and longitudinal association analyses with multiple-testing control (Analysis). The results provide candidate behavioral indicators that Chapter 5 will refine and Chapter 6 will leverage for individual-level, concurrent estimated intrinsic learning motivation.

4.1 Alignment with Research Questions (RQ1)

Consistent with the framing and research agenda introduced in Chapter 1, this study specifically addresses the first line of inquiry on behavior–motivation links by examining how observable behavioral traces relate to underlying learning motivation states and by systematically organizing this initial direction of investigation within the overall dissertation framework:

RQ1a: What social behavioral patterns emerge from daily smartphone app-usage signals of university students?

RQ1b: How are these patterns associated with intrinsic learning motivation (intrinsic learning motivation; SRQ-A intrinsic subscale)?

4.2 Method Design Overview (Pointers to Shared Procedures)

Unless otherwise stated, data handling follows Chapter 3 (de-identification, fixed-length analysis windows, coverage metrics, missingness handling, and scaling). Study-specific choices are: a 12-week observation of $N = 58$ undergraduates; weekly intrinsic learning motivation self-reports; 27 app-usage features grouped by social interactivity; principal-component analysis (PCA) for pattern extraction [298]; generalized estintrinsic learning motivationating equations (GEE) for longitudinal associations; and two-stage Benjamini–Hochberg (TSBH) control of FDR. Ethics and data governance follow Chapter 1.

4.3 Contributions of This Chapter

This chapter makes three closely connected contributions to the overall dissertation. First, it demonstrates the *in-the-wild* feasibility of linking passively sensed *social* behaviors, which are captured via smartphone-based traces in everyday life, with SDT-anchored intrinsic learning motivation in authentic higher-education contexts. Furthermore, it shows that such links can be observed naturally and readily without disrupting routine academic activities. Second, it extracts a set of interpretable behavioral patterns and presents them as candidate indicators of learning motivation, thereby seeding a reusable, theory-aligned feature space that can support replication, extension, and cross-study comparison in future work. Third, by establishing these feasibility results and by proposing an initial catalog of candidate indicators, the study lays the empirical groundwork for the subsequent Chapter 5, where these behavioral indicators are further refined, systematically evaluated, and integrated into a more optimized representation of behavior–motivation relations.

4.4 Methods

This section describes the study design and implementation of association analysis study, a three-month mobile sensing study with $N = 58$ second-year undergraduates at a Chinese university. We outline the longitudinal protocol that combined weekly SRQ-A intrinsic motivation surveys with continuous smartphone-based sensing, and we summarize the final analytic sample and its demographic composition. We then detail the *iSense* logging configuration, the construction and social-interactivity grouping of 27 app-usage features, and the association analysis pipeline that used PCA and GEE to link behavioral patterns to intrinsic learning motivation. Finally, we note the embedded ethical safeguards, including prior IRB approval, de-identification, and restricted-access storage on a private server.

4.4.1 Study Design

We collected and analyzed data from a continuing mobile sensing study that followed $N = 58$ university undergraduates from one university in China, combining continuous smartphone sensing with repeated self-report surveys administered via structured questionnaires to capture their learning-related states over

time. The overall study period lasted for three months (12 weeks of active data collection) from March 2022 to June 2022, spanning from the beginning of the second week in March to the end of the first week in June. The observation window covered an entire segment of the spring semester under relatively stable academic conditions. Students were recruited and provided informed consent at the beginning of their spring academic semester (initially $N = 65$ on March 6, 2022), one week before the formal data collection began, ensuring that enrollment and consent procedures were completed before any sensing or survey data were recorded. After recruitment, in the pre-testing phase (the first week in March 2022), we provided a tutorial on installing the mobile sensing app (iSense) on their phones, walked participants through the basic steps so that they could successfully install and open the app, and asked them to keep the app installed and running without interruption until June 5, 2022 (*end of the spring semester*; includes semester intrinsic learning motivation and short breaks). We began collecting and later analyzing the data on March 13, 2022, coinciding with the start of the 12-week sensing window. During the study, we monitored daily data collection, checked whether the expected sensing streams were being uploaded for each participant, and reminded noncompliant participants (e.g., those who accidentally shut down the app) to recheck the app status, reopen or restart the app if necessary, and resume participation so that the continuity of the time series was maintained as much as possible.

Our study was approved by the hosting university's Institutional Review Board on March 4, 2022, and all study procedures were carried out in accordance with this ethical clearance. Weekly self-report surveys (SRQ-A intrinsic motivation subscale) were delivered to participants via the app as short questionnaires, and participants were compensated 20 RMB per week for completing these surveys. The goal of the study is to understand the relationships between students' social behavioral patterns (as captured through passive smartphone sensing) and their intrinsic learning motivation (i.e., their self-reported intrinsic motivation for learning). Unless otherwise noted, all procedures for de-identification, temporal alignment, coverage metrics, missingness handling, scaling of variables, and aggregation of features follow the standardized workflow described in Chapter 3.

4.4.2 Demographics

We initially recruited 65 participants, which formed the original pool of students invited to take part in the study. Of these, 58 participants who complied with the study protocol and completed the weekly intrinsic learning motivation assessments were ultimately included in the analysis, so the final analytic sample is based on those students who provided both sensing data and regular self-reports over the observation period. This screening ensured that only participants with sufficient and usable data across the study weeks contributed to the reported results.

Two participants withdrew from the study in the second week, which meant that their partial data were not considered in the main analyses. In addition, five participants were excluded because they had substantial missing sensing data, for example long stretches without recorded smartphone signals, which made it impossible to derive reliable behavioral features for them. Together, these exclusions reduced the sample

from the initially recruited 65 to the 58 participants whose data met the predefined completeness criteria.

All participants were Chinese undergraduates in their second year of study, so they were at a similar academic stage within their degree programs and experienced comparable institutional and curricular demands during the data collection period. This shared background provides a relatively coherent cohort in terms of age, educational level, and university context.

The majority of participants were male (67.24%, N=39), which means that just under one third of the analyzed sample consisted of female students. Thus, while both genders were represented, the sample is somewhat skewed toward male undergraduates in this setting.

Most students were majors in Sports Training (87.93%, N=51), and only a small number came from other physical education related majors, specifically Rhythmic Gymnastics (5.17%, N=3), Tennis (3.45%, N=2), and Swimming (3.45%, N=2). This distribution indicates that the sample is heavily concentrated in Sports Training, with minor representation from these three specialized disciplines. Table 5 summarizes the participant demographics and provides a concise overview of this composition.

Table 5: Demographics of participants used in the analysis ($N = 58$).

Category	**Count**	**Percentage**
Sex		
Female	19	32.67%
Male	39	67.24%
Major		
Sports Training	51	87.93%
Rhythmic Gymnastics	3	5.17%
Tennis	2	3.45%
Swimming	2	3.45%

4.4.3 Mobile Sensing Data

The *iSense* app was designed and developed to collect mobile sensing data from multiple smartphone sensors in a systematic, structured, and unobtrusive way that runs quietly in the background without interrupting students' daily activities. The app captures phone activities including the number of incoming and outgoing calls, application usage (recording only the name of the running application, with no message, audio, or screen content stored), and the number of text messages sent and received, so that only behavioral metadata are collected. Phone activities are captured every 5 minutes according to a fixed sampling schedule; if the device is offline or temporarily lacks network connectivity at the scheduled sampling time, the data are buffered and stored locally on the device and then uploaded to a private server once the device reconnects to Wi-Fi under normal operating conditions. We compute behavioral features from the sensed data on a daily basis (for example, the duration of using a certain app per day or the aggregated counts of specific activities) to capture participants' behaviors across time in a consistent, comparable

format that can be aligned with other study measures. The overall sensing design and preprocessing pipeline follows the same procedures described in Chapter 3, ensuring that data collection, cleaning, and feature construction remain aligned with the broader study framework and that all studies share a common technical foundation.

4.4.4 Intrinsic Motivation Assessment

Intrinsic motivation (intrinsic learning motivation) was measured using the intrinsic subscale of the Academic Self-Regulation Questionnaire (SRQ-A). Weekly SRQ-A administrations yielded tintrinsic learning motivatione-aligned intrinsic learning motivation scores for association analyses with sensing-derived behavioral patterns. Table 6 summarizes the response range and overall distribution.

Table 6: Statistics of Intrinsic Motivation (SRQ-A intrinsic subscale).

Measure	Response range	Overall mean (std)
Intrinsic Motivation	1–5	3.00 (1.48)

4.4.5 Behavioral Features

We explored daily behavioral patterns with a primary focus on interactive behaviors derived from smartphone usage logs. We labeled and computed 27 features from the sensed data, encompassing usage details (e.g., app name, duration) across 450 applications. Features were grouped by social interactivity into three categories: non-social, indirect-social, and direct-social (Table 7). Categories were determined by the degree of social interactivity intrinsic learning motivationplied by app use (e.g., instant messaging for direct interaction; browsing/news for indirect interaction; camera use as non-social). Several features explicitly capture socially oriented study or communication contexts (e.g., `social_study`, `social_business`, `social_date`, `social_blog`, `social_bbs`).

Table 7: Features extracted and grouped by social interactivity.

Category	Details
Non-Social	`camera`, `health`, `game_single`, `navigation`, `system`, `weather`, `tool_other`, `tool_study`, `finance`, `others`
Indirect-Social	`info_entertainment`, `web_browsering`, `shopping`, `music`, `video`, `social_bbs`, `radio/podcast`, `livestream`, `info_study`
Direct-Social	`social_business`, `intrinsic learning motivation`, `call`, `SMS`, `social_study`, `game_multiple`, `social_blog`, `social_date`

4.4.6 Association Analysis Pipeline

We applied PCA to extract interpretable behavioral patterns from the 27 app-usage features, treating these features as a high-dimensional representation of weekly smartphone behavior. For each participant and each week, we first computed daily means within a 7-day window so that short-term fluctuations were smoothed while the overall weekly profile was preserved. This procedure yielded a matrix X in which rows correspond to weekly SRQ-A response occasions and columns correspond to the 27 app-usage features, providing a consistent structure for multivariate analysis. Running PCA on X produced a set of principal components (PCs) that represent orthogonal directions of maximal variance in the app-usage space. We retained 11 PCs that together explained 90% of the total variance, as shown by the cumulative variance plot in Figure 12, which indicates that these components capture most of the meaningful variability in the original feature set while reducing dimensionality.

We then used GEE to model longitudinal associations between the PC scores and the SRQ-A intrinsic motivation (intrinsic learning motivation) scores over time, allowing us to account for association across repeated weekly measurements. In this framework, weekly PC scores served as predictors and intrinsic learning motivation scores served as the outcome for each observation, enabling us to test which behavioral patterns were positively or negatively related to intrinsic learning motivation across the study period. To control the false discovery rate (FDR) when testing multiple PCs, we applied a two-stage Benjamini–Hochberg (TSBH) procedure, which provides a more powerful and structured way to limit the expected proportion of false positives than simple per-test thresholds while retaining interpretability of the resulting associations.

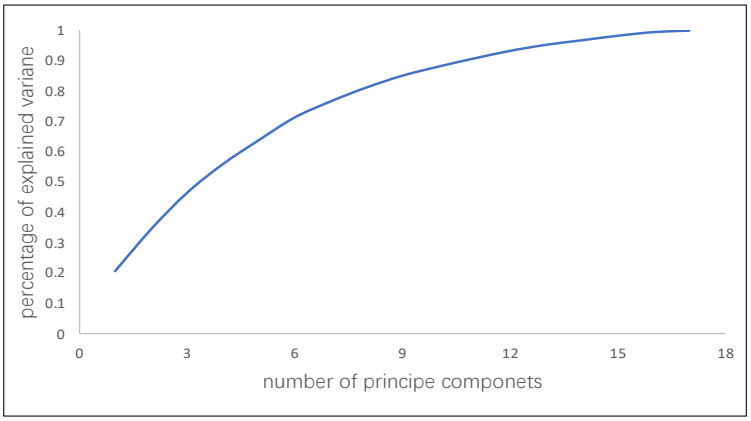


Figure 12: Cumulative variability in the sensed data explained by the top-n principal components.

4.4.7 Ethics and Privacy

Ethics and data governance fully conform to the human-subjects protections described in Chapter 1, and all procedures for the collection, handling, storage, and use of study data are conducted in alignment with those requirements throughout the project. The data collected by the mobile app were de-identified prior to retention, with personally identifying details removed, and then stored on a private, non-public server

that is not publicly reachable, with access limited to authorized personnel under strict access controls.

4.5 Results

In this section, we characterize students' smartphone-based behavioral patterns and examine how these patterns relate to weekly intrinsic learning motivation. Using PCA on app-category usage features, we reduce the high-dimensional sensing data to 11 principal components that capture distinct behavioral patterns, such as using study tools and social-study apps, engaging in online gaming and livestream/video consumption, or relying heavily on social communication and entertainment apps. We then relate these component scores to intrinsic learning motivation using GEE, identifying which behavioral profiles tend to co-occur with higher versus lower intrinsic learning motivation over time. Table 8 summarizes the extracted patterns and their high-weighted features, and Table 9 reports the patterns that show significant associations with intrinsic learning motivation.

4.5.1 Behavioral Patterns (from PCA)

We interpreted the selected 11 PCs as behavioral patterns by inspecting high-weighted features. Table 8 lists the patterns and their matching features. Pattern 1 reflects using social apps for dating together with frequent listening to podcasts or radio and other system/tool use; Pattern 2 reflects frequent use of study tool apps and social-study apps, along with browsing BBS; Pattern 3 reflects playing online multiplayer games and watching game-related livestreams or videos; Pattern 4 reflects listening to music on the smartphone while reading or searching for learning-related information; Pattern 5 reflects intensive instant-messaging communication with others; Pattern 6 reflects frequent use of social networking platforms and actively posting or sharing content (e.g., taking photos); Pattern 7 reflects frequent engagement with shopping apps; Pattern 8 reflects reading or searching entertainment content such as comics or sports news; Pattern 9 reflects heavy use of single-player games; Pattern 10 reflects extensive web browsing; and Pattern 11 reflects frequent use of second-hand trading and finance-related platforms.

Table 8: Behavioral patterns and their matching features (indexed by the n-th PCA component).

Pattern	**Features**
1. Use social apps to date and like listening to podcast/radio	`social_date`, `system`, `radio/podcast`, `tool_others`
2. Use study tool apps frequently and more social-study apps; browse BBS	`tool_study`, `social_study`, `social_bbs`
3. Play online games with others and watch livestream/videos (games related)	`video`, `game_multiple`, `livestream`

Pattern	Features
4. Listen to music on smartphones and read/search learning-related resources	`music`, `info_study`
5. Chat with others a lot	`Instant Message (IM)`
6. Use social networking platforms frequently, like to post or tweet	`social_blog`, `camera`
7. Use shopping apps frequently	`shopping`
8. Read/search content for entertainment (e.g., comics or sports news)	`info_entertainment`
9. Play single-player games a lot	`game_single`
10. Browse the web a lot	`web_browsering`
11. Use second-hand trading platforms a lot	`social_business`, `finance`

4.5.2 Associations with Intrinsic Motivation (intrinsic learning motivation)

We applied GEE to analyse the relationship between PC scores and intrinsic learning motivation scores, with the aim of identifying which behavioral patterns were statistically associated with intrinsic motivation over time. We chose GEE because it is specifically designed for longitudinal data with repeated measurements per participant and can account for the within-person correlation of observations across weeks, while still estimating population-level effects in a robust way. In intuitive terms, GEE can be viewed as an extension of regression methods for panel data that allows us to examine how changes in PC scores relate to changes in intrinsic learning motivation across multiple time points for the same student, instead of treating each week as an independent case.

The resulting associations are summarized in Table 9, which reports, for each identified behavioral pattern, the direction and relative strength of the effect estimate in a compact form. In this framework, positive associations indicate that students who more strongly express the corresponding pattern tend, on average, to have higher intrinsic learning motivation levels, whereas negative associations indicate comparatively lower intrinsic learning motivation for students whose behavior is more heavily characterized by that pattern. This sign-based interpretation provides a direct link between latent behavioral patterns and motivational tendencies and helps clarify which types of smartphone use are aligned with, or potentially in tension with, intrinsic learning motivation. Consistent with this interpretation and within the significance thresholds adopted in this study (all $p < .05$; FDR $< .10$ where noted), we found that Pattern 2 was positively associated with intrinsic learning motivation, whereas Pattern 3 and Pattern 9 showed negative associations, indicating that these latter patterns are characteristic of students with relatively lower

intrinsic learning motivation.

Table 9: Behavioral patterns associated with intrinsic learning motivation (intrinsic learning motivation).

Pattern	Association with intrinsic learning motivation
2	(+)
3	(–)
9	(–)

Note: Signs indicate direction of association; all $p < .05$. Associations passing FDR $< .10$ are bolded in text.

4.6 Discussion

This study demonstrates that passively sensed *social* behavioral patterns are systematically associated with intrinsic learning motivation (intrinsic learning motivation) in authentic, real-world learning settings, rather than only under controlled laboratory tasks that may not reflect everyday student life. In other words, the way students routinely use their smartphones to communicate, study, and entertain themselves throughout an academic term carries structured information about how strongly they are intrinsically motivated to learn and how this motivation fluctuates over time in context. A plausible interpretation is that students who are more active in *socially supported studying* and *study tool* usage (Pattern 2), for example students who frequently rely on study tools, interact with social-study apps, and browse BBS for learning-related content, benefit from behaviors that are aligned with self-regulated learning processes, such as planning their study sessions, monitoring their progress, seeking help and feedback from peers, and staying engaged with course materials in a deliberate and sustained way. These behaviors are consistent with autonomous, self-endorsed forms of engagement that reflect interest, perceived value, and a sense of ownership over learning, and thus correspond to higher intrinsic learning motivation as conceptualized within SDT. In contrast, heavy engagement in *online gaming and livestreaming* (Pattern 3) or *single-player gaming* (Pattern 9) may reflect competing time use, distraction from academic activities, or motivational displacement, in which sustained interest and persistence are redirected from learning tasks toward gaming-related pursuits that are more immediately rewarding but less educationally beneficial. This pattern is associated with lower intrinsic learning motivation in the observed data.

Overall, the findings support using smartphone-derived social behavioral indicators as interpretable, SDT-consistent signals for assessing intrinsic learning motivation in a continuous and ecologically valid manner that aligns with students' everyday routines and naturally occurring study environments. These indicators capture patterns in how students communicate, study, and engage with their devices over time, and they can be monitored with relatively low burden, which makes them suitable for repeated or long-term use in authentic educational contexts where frequent surveys would be impractical. These patterns do not replace traditional self-report scales but rather complement them by providing objective, behavior-based proxies that can be tracked over time and linked to specific contexts and routines, thereby strengthening assessment through triangulation between subjective reports and observable behavior. In practice, such

indicators can help bridge the gap between static questionnaire snapshots and the dynamic flow of daily behavior, since they offer a way to follow short-term fluctuations and gradual trends in motivation while still remaining grounded in theory and anchored to validated self-report measures. This motivates the refinement and extension of such indicators in the next chapter, where we further sharpen how passively sensed behaviors can be mapped onto theoretically grounded constructs of intrinsic learning motivation, examine their robustness and interpretability across different analytic choices, and explore how these mappings can be used to support more nuanced and interpretable assessment at the individual level in both research and potential applied settings.

4.7 Limitations

This study has several limitations that should be considered when interpreting the findings. First, the sample size is modest, and the participants are drawn from a specific and relatively narrow demographic group, because they are mostly physical education majors within a single Chinese university. This restricted sampling frame limits the generalizability of the results to other institutions, disciplines, and cultural or educational contexts. Therefore, replication in larger and more diverse cohorts is needed before broad conclusions can be drawn. Second, we focused on app-usage features rather than exploiting the full available sensor suite, which constrains the range of behaviors that can be captured and may overlook aspects of mobility, social context, or physiological states that could also be relevant for intrinsic learning motivation. Third, heterogeneity in compliance across participants, together with device and platform-specific policies that govern background sensing, introduce nonrandom missingness patterns in the data, and these patterns may bias estimates in ways that are difficult to fully correct. Fourth, non-stationarity around academic calendar events, for example examination periods or holidays, may affect the stability of both behavioral patterns and intrinsic learning motivation. Residual calendar effects may remain even after temporal alignment. Finally, causal interpretations are not warranted given the observational design.

4.8 Summary

This study identified multiple smartphone-derived *social* behavioral patterns and documented their differential associations with *intrinsic learning motivation*. At this stage, the association-analysis study demonstrates the feasibility of assessing intrinsic learning motivation from smartphone-sensed behavioral data by establishing behavior–motivation associations. It also suggests actionable behavioral indicators that will be *refined and stress-tested* in Chapter 5.

5

Optimizing Behavior Identification by Refining Indicators from Behavioral Changes

This chapter constitutes optimization-stage study in the three-stage trajectory (from association to optimization to estimation). Building on the association-stage study, it refines the mapping from passively sensed smartphone behaviors to *intrinsic learning motivation* by (i) expanding and structuring behavioral features, (ii) formalizing behavior identification via clustering and expert-informed labeling, and (iii) validating candidate indicators through *between-group* comparisons (high vs. low intrinsic learning motivation) alongside longitudinal associations with intrinsic learning motivation. All core pipelines (de-identification, fixed analysis windows, coverage metrics, missingness handling, scaling) follow Chapter 3.

5.1 Alignment with Research Questions (RQ2)

RQ2: What passively sensed behavioral indicators can effectively characterize the temporal dynamics of students' *intrinsic learning motivation*?

Objectives: (1) Optimize behavior identification and feature construction; (2) derive interpretable clusters/patterns as candidate indicators; (3) test robustness using association analysis with intrinsic learning motivation and *between-group* comparisons; (4) prepare the refined indicator set for individualized, concurrent estimation as the next stage study.

5.2 Study-Specific Method Design Beyond Chapter 3

Beyond the standardized pipeline in Chapter 3 (de-identification, fixed windowing, coverage/missingness handling, and scaling), this study introduces four methodological advances geared to behavior identification optimization: a broader evidence base (a larger cohort observed for 14 consecutive weeks), an enriched sensing configuration that expands the representation to a 59-feature, multi-family space (mobility/activity, semantic locations, sleep and ambient context, and fine-grained app/phone interaction), a clustering-based behavior identification pipeline that replaces PCA with k-means (silhouette-selected k=9) coupled with permutation-based feature importance and expert labeling for interpretability, and a

strengthened validation strategy that combines longitudinal GEE associations with intrinsic motivation and high–low intrinsic learning motivation feature contrasts. Together these outputs yield a refined, reusable indicator set that underpins the individualized, concurrent estimation developed in Chapter 6.

5.3 Methods

This section describes the participants, sensing protocol, behavioral feature set, behavior identification, and statistical analyses used to address RQ2. We first report recruitment flow and demographics for 118 Chinese undergraduates across multiple majors, then summarize how passive smartphone sensing and weekly SRQ-A intrinsic learning motivation self-reports were collected and aligned. Next, we describe the daily and windowed features that capture mobility, physical activity, semantic locations, sleep and ambient context, and detailed smartphone usage over 42 app genres, yielding a 59 dimensional standardized feature set. We then outline behavior identification via *k*-means clustering on z scored features, supported by permutation feature importance and expert informed labels. Finally, we introduce the association analysis and between group comparisons that merge cluster assignments with longitudinal intrinsic learning motivation scores using GEE, Wilcoxon tests, and FDR correction to evaluate and validate candidate behavioral indicators.

5.3.1 Demographics

Of 130 recruits, 118 completed the study and were included in the final analysis, so the analyzed sample represents the large majority of the initially enrolled participants and provides a reasonably solid basis for the planned longitudinal analyses. In total, 7 participants were excluded because of substantial sensing loss that prevented reliable feature extraction from their smartphone logs, meaning that the available data for these individuals were too sparse or fragmented to support consistent computation of behavioral features over time. A further 3 participants were excluded because of missing self-reports on intrinsic learning motivation, which are essential for providing the outcome labels against which the sensing based features are compared, and 2 participants chose to withdraw in week 1, which reduced the effective sample size but ensured that only those who consented to continued participation and provided sufficient data contributed to the final results. These exclusion steps necessarily reduce the number of participants, but they also improve the internal consistency of the dataset by focusing the analysis on individuals whose sensing coverage and self-report data meet the predefined quality criteria.

All analyzed participants were Chinese undergraduates from the same institution, so the sample reflects a relatively homogeneous cultural and institutional context and should be interpreted as such when considering generalizability to other universities or countries. As shown in Table 10, males constituted 60.2% of the sample (N=71), with the remaining 39.8% of participants being female, which reflects a moderately male dominated cohort and should be kept in mind when interpreting any gender related patterns. Academic majors spanned a range of disciplines, including Clinical Medicine (25.4%), Medical Imaging (19.5%),

Pharmacy (16.1%), Biochemistry (10.2%), Vehicle Engineering (9.3%), Electronic Automation (6.8%), Software Engineering (5.1%), Economics (3.4%), English Literature (2.5%), and Philosophy (1.7%), indicating that the sample covers both health related and engineering disciplines as well as smaller groups from economics and the humanities. Together, this distribution suggests that the findings primarily reflect the experiences of students in health and engineering related fields, complemented by more limited representation from social science and humanities majors, while still preserving sufficient diversity to explore behavioral and motivational variation within this institutional context.

Table 10: Participant demographics and sample flow

Category	Item	N (%)
Recruitment / Flow	Recruited	130
	Completed & analyzed	118
	Excluded: sensing loss	7
	missing self-reports	3
	withdrew in week 1	2
Gender	Male	71 (60.2%)
	Female	47 (39.8%)
Majors	Clinical Medicine	30 (25.4%)
	Medical Imaging	23 (19.5%)
	Pharmacy	19 (16.1%)
	Biochemistry	12 (10.2%)
	Vehicle Engineering	11 (9.3%)
	Electronic Automation	8 (6.8%)
	Software Engineering	6 (5.1%)
	Economics	4 (3.4%)
	English Literature	3 (2.5%)
	Philosophy	2 (1.7%)

5.3.2 Data Collection

Passive smartphone sensing: The app passively logged data on participants' smartphones at regular intervals, collecting records every 5 minutes where this sampling rate was applicable. The logs covered physical activities (e.g., walking, running, cycling), counts of calls and SMS, ambient sound and light levels, battery and screen events (unlock/lock, charge cycles), coarse location, and the name of the foreground app that was in use at that moment. These data were first stored locally on the device and were then uploaded to private servers when connectivity was available, so that the complete time series could be centralized for subsequent analysis.

Intrinsic motivation self-reports: Intrinsic learning motivation was measured with the SRQ-A intrinsic subscale on a 1–5 scale, and a score was obtained for each participant in each week of the study. These weekly intrinsic learning motivation scores were aligned to fixed-length behavioral windows for analysis,

so that self-report and sensing features referred to the same time periods. Across all participants and all weeks, the overall intrinsic learning motivation mean was 2.95 (SD 1.48).

5.3.3 Behavioral Feature Set

Features were computed daily, then aggregated into analysis windows aligned with the intrinsic learning motivation reports. The set spans mobility, physical activity, semantic locations, sleep and ambient context, and smartphone usage (screen/battery/calls/SMS and app usage over 42 genres), yielding 59 standardized windowed features.

Mobility: Taking into account both phone energy conservation and data quality, the mobile application retrieves GPS data at a fixed interval of every 5 minutes, rather than attempting continuous high-frequency sampling. This schedule provides a reasonable compromise between preserving battery life and maintaining sufficient temporal resolution to capture daily movement patterns. We use the collected GPS points to derive the locations that participants visited, such as different campus areas or off campus locations, and to compute the distance they traveled within each day and analysis window.

Physical activity: We discern the participant's ongoing activities through the utilization of the activity recognition API, which classifies movement into categories such as walking, running, or cycling based on sensor input. We then use this information to determine the participants' physical status over time and to calculate the duration of different physical activities, for example the total amount of time spent walking during a given day or within a specific analysis window.

Semantic locations: Since all the participants live in the student residence hall on campus and the vast majority of their campus life takes place within the campus area, we focus on identifying meaningful places in this environment. We first collect raw GPS coordinates with our mobile application and then assign semantic labels to these locations using a third party API together with the university campus map, which provides detailed labels for various zones and buildings within the campus, such as teaching buildings, dormitories, and sports facilities.

Sleep patterns and environment context: We extract sleep related data from the smartphones and use established heuristics to derive sleep duration as well as sleep start and end times, with a typical margin of error of around 40 minutes. In addition, we derive environmental data from the smartphones, including the sound level and ambient light level, which together provide a coarse description of the participant's surrounding context during different periods of the day and night.

Smartphone usage: We use the mobile application to capture a variety of phone usage signals, including screen events such as the duration of the unlock state, battery events such as the number of charging cycles, and communication events such as the number of calls and SMS messages. We also record the detailed daily use of different mobile apps. Data on specific app usage are collected once every 5 minutes, and we then categorize the 779 mobile apps observed in the data into 42 genres, for example education, entertainment, or music/audio, and compute the duration of use for each app genre within each day and

analysis window.

5.3.4 Behavior Identification via Clustering

We applied k-means clustering with Euclidean distance across a candidate range of k values and then selected k=9 using the silhouette criterion, because this choice produced the highest average silhouette value of 0.28 [299], as shown in Figure 13. For each resulting cluster, we computed permutation feature importance in order to identify which features contributed most strongly to the cluster separation, and on this basis we assigned expert-informed labels that summarize and interpret the characteristic behavioral profile represented by each cluster.

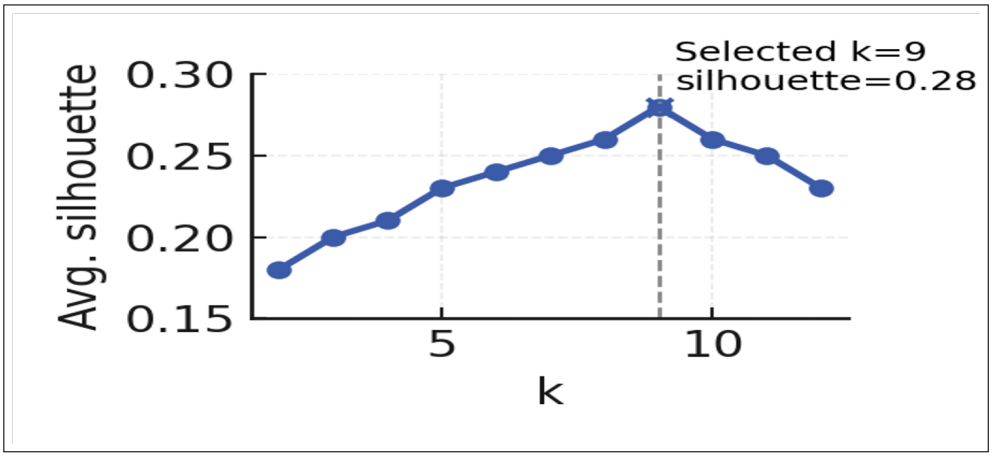


Figure 13: Average silhouette vs k fork-means (Euclidean) on z-scored features. The maximum occurs at k=9 (0.28), which we adopt for subsequent clustering and per-cluster interpretation

5.3.5 Association Analysis and Between-Group Comparison

We merged sensing windows, weekly intrinsic learning motivation scores, and cluster labels, then used GEE to model longitudinal associations between behavior patterns and intrinsic learning motivation with FDR control. To validate indicators at the feature level, participants were stratified into high/low intrinsic learning motivation groups and daily-average features were compared via Wilcoxon tests with FDR correction; results are reported as percentage changes.

5.4 Results

In this section, we first describe the nine clustered behavior patterns derived from 59 passive-sensing features and provide expert-informed labels that summarize their dominant signals. We then examine how these patterns relate to intrinsic learning motivation over time, reporting which clusters show positive or negative longitudinal associations and comparing high- versus low-motivation groups at the feature level using nonparametric tests with FDR control. Next, we visualize the sensing features that differ significantly between groups and interpret how study-supportive behaviors, entertainment-focused phone use, sleep patterns, and mobility jointly characterize higher or lower intrinsic learning motivation. Finally, we distill

a refined set of behavioral indicators and map them to their expected direction with respect to intrinsic learning motivation, yielding an interpretable schema that will be reused for individualized, concurrent estimation.

5.4.1 Nine Behavior Patterns (Clusters) and Interpretations

Table 11 summarizes the nine behavior patterns derived from clustering and the dominant signals that define each pattern; these labels are used throughout the remainder of the chapter.

Table 11: Behavior patterns (clusters) identified from 59 sensing features and their top feature signals.

Top features (signals)	**Behavior label and interpretation**
Social/intrinsic learning motivation /communication duration; long sedentary time	**C1: Socializing and chatting on smartphones.** Spends substantial time on social media and instant messaging; prolonged sedentary periods.
Long stay in library; productivity-app duration; high unlock count	**C4: Productivity tools in the library.** Frequent use of office/collaboration apps while staying in library.
Short-video app duration; unlock duration	**C2: Regular short-video viewing.** Habitual consumption of short videos on smartphones.
Multiplayer-game duration; long dorm stay; late sleep start; voice-call duration	**C5: Online gaming & late sleep in dorm.**
Video play duration; education/utility/social app durations	**C7: Learning-related video/resources & social learning tools.**
Dorm stay; audio-play duration; music-app duration; ambient sound level; sedentary time	**C9: Music listening in dorm & sedentary.**
Shopping/lifestyle/entertainment app durations; sleep start time	**C6: Life-sharing/entertainment & shopping near bedtime.**
Livestream duration; food/drink delivery app usage	**C8: Livestream watching & frequent takeout orders.**
Early sleep end; longer sleep; jogging; higher daylight; outdoor stay (sports field)	**C3: Early riser with daytime jogging/outdoor activity.**

For each behavioral pattern, the left column lists the top contributing signals that most distinctively characterize the pattern, and the right column provides a concise, human-readable label and interpretation. The patterns span social communication and sedentary use (C1), focused productivity activity in library settings (C4), habitual short-video viewing (C2), online gaming with late sleep concentrated in dorms (C5), learning-related video consumption and social learning tools (C7), music listening in dorm environments

with sedentary time (C9), life-sharing/entertainment and shopping near bedtime (C6), livestream watching accompanied by frequent food/drink delivery (C8), and early-riser routines with daytime jogging and outdoor activity (C3).

The behavioral pattern labels are descriptive rather than normative and are intended to condense recurrent co-occurrences of sensing signals into interpretable behavior motifs. Importantly, the labeling and narrative interpretations were developed in collaboration with psychologists through iterative review of feature summaries and representative weeks, ensuring that the behavioral descriptions are theoretically grounded and consistent with established constructs in educational and motivational psychology. As such, the table provides an interpretable bridge from multi-sensor signals to behavior categories that can be reused in downstream analyses, while acknowledging that these patterns are directional characterizations and not causal claims.

5.4.2 Associations Between Behavior Patterns and intrinsic learning motivation

We next assessed longitudinal associations between these patterns and intrinsic learning motivation; Table 12 reports which patterns relate positively or negatively to intrinsic learning motivation (all $p<.05$; patterns passing FDR $< .10$ are highlighted in the text where relevant).

Table 12: Behavior patterns associated with intrinsic learning motivation (intrinsic learning motivation).

Positively associated	**Negatively associated**
C1, C3, C4, C6, C7, C9	C2, C5, C8

To compare and analyze the behavioral changes among students with distinct learning motivation levels, we divide the participants into two groups based on their SRQ-A scores, corresponding to high and low motivation levels, respectively. As a result, the high learning motivation level group achieves the overall mean of 3.6 with the standard deviation of 1.28, while the low group achieves the overall mean of 2.0 with the standard deviation of 1.47. Furthermore, as we have obtained the behavioral features computed and extracted from passive sensing data, we analyze the behavioral changes by exploring features that exhibit variations between the high-level and low-level groups. We conduct the comparison by using Wilcoxon signed-rank test [300]. After applying the Benjamini/Hochberg FDR correction procedure for multiple comparisons, the displayed results reveal statistical significance with a p-value below 0.05.

Features with significant differences after FDR correction are visualized in Figure 14, where each bar represents one sensing-derived behavioral feature that survived multiple-comparison adjustment. In this visualization, positive bars indicate that the corresponding feature has, on average, higher values in the high-intrinsic learning motivation group, whereas negative bars indicate that the feature is instead elevated in the low-intrinsic learning motivation group. This sign-based encoding makes the direction of the between-group difference immediately apparent, allowing the reader to quickly identify which behaviors

tend to be more pronounced in each group. Overall, the plot provides an intuitive, at-a-glance summary of how specific behavioral features differ between students with higher versus lower intrinsic learning motivation.

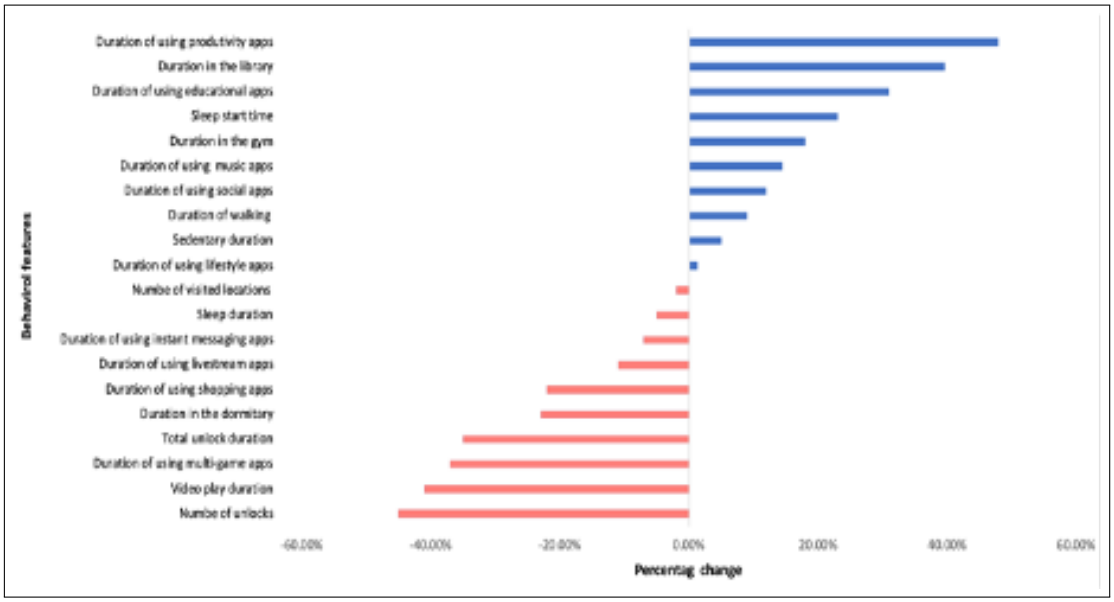


Figure 14: Features with significant differences between high- and low-intrinsic learning motivation groups (positive bars indicate higher values in the high-intrinsic learning motivation group).

Additionally, please note that our statistical comparisons specifically pertain to daily averages. The y-axis represents sensing behavioral features that reveal significant differences between the two groups, while the x-axis illustrates the percentage change. The bars in Figure 15 diverge to indicate whether a feature undergoes positive or negative changes between the two groups (e.g., a positive change indicates that the particular feature is more prominent in the high learning motivation level group). The positive percentage changes are signified by blue bars, while the negative percentage changes are denoted by red bars. More specifically, we observed from the figure that for the positive changes, the features including duration of using productivity apps (+48%), duration of using education apps (+31%), and duration in the library (+39.6%) change with a relatively higher rate. It is reasonable to infer that students with a high level of learning motivation might have the habit of using learning-supportive apps. Correspondingly, the duration in the gym (+18%), duration of using music apps (+14.5%), duration of using social apps (+12%), duration of walking (+9%), sedentary duration (+5%), and duration of using lifestyle apps (+1.3%). An interesting finding is the positive change in sleep start time (+23%) and the negative change in sleep duration (-5%), which might indicate that students with the higher level of learning motivation are likely to go to bed and rise early. As for other features that change negatively, we find that features including number of unlocks (-45%) and video play duration (-41%), duration of using multi-game apps (-37%), total unlock duration (-35%), duration in the dormitory (-23%), duration of using shopping apps (-22%), duration of using livestream apps (-11%), duration of using instant messaging apps (-7%), and number of visited locations (-2%). This suggests that students with lower learning motivation levels overall spend more time using

their phones for entertainment purposes.

5.4.3 Refined Behavioral Indicators

Based on the behavioral differences summarized above, we distilled a refined set of passive-sensing behavioral features that reliably distinguish high versus low intrinsic motivation. Table 13 lists these features, their expected direction relative to intrinsic learning motivation, the observed percentage change (daily averages; Benjamini–Hochberg FDR at $p < 0.05$), and a brief interpretation. The column widths are set as fractions of `\linewidth` so the table never exceeds the text block.

Table 13: Refined behavioral features differentiating high vs. low intrinsic motivation. Values reflect daily averages; a positive sign indicates the feature is higher in the high-intrinsic learning motivation group. All displayed differences remain significant after Benjamini–Hochberg FDR correction at $p < 0.05$ (Figure 15).

Behavioral feature	**Direction**	**Change**	**Interpretation**
Productivity app duration	Positive	+48%	Greater study-supportive app use
Education app duration	Positive	+31%	More direct learning activity on phone
Library duration	Positive	+39.6%	More time in study-supportive locations
Gym duration	Positive	+18%	More health/activity routines
Music app duration	Positive	+14.5%	Mild increase in music use (possible regulation aid)
Social app duration	Positive	+12%	Slightly higher social engagement
Walking duration	Positive	+9%	More light physical activity
Sedentary duration	Positive	+5%	Marginal increase in seated periods
Lifestyle app duration	Positive	+1.3%	Small rise in daily-life management apps
Sleep start time	Positive	+23%	Earlier sleep/wake tendency inferred from pattern
Sleep duration	Negative	−5%	Slightly shorter sleep duration
Number of unlocks	Negative	−45%	Fewer phone checks
Video play duration	Negative	−41%	Less entertainment video consumption
Multiplayer game duration	Negative	−37%	Less multi-game usage
Total unlock duration	Negative	−35%	Less overall screen-on time
Dormitory duration	Negative	−23%	Less time confined to dorm
Shopping app duration	Negative	−22%	Reduced online shopping activity
Livestream app duration	Negative	−11%	Less livestream viewing

Continued on next page

Behavioral feature	Direction	Change	Interpretation
Instant messaging app duration	Negative	−7%	Slightly less intrinsic learning motivation usage
Number of visited locations	Negative	−2%	Slightly fewer distinct places visited

Table 14 consolidates the refined indicator set by mapping observable patterns and example features to their expected direction with respect to intrinsic motivation; this same set is reused in Chapter 6 for individualized, concurrent estimation. Indicators expected to relate *positively* to intrinsic learning motivation include (i) library presence and longer durations in productivity/education apps and social-learning tools, and (ii) daytime physical activity, earlier rising, and reduced dorm confinement. These patterns are consistent with SDT, in which competence- and autonomy-supportive study contexts and self-regulated routines tend to accompany higher intrinsic learning motivation. In contrast, indicators expected to relate *negatively* to intrinsic learning motivation include short-video and livestream consumption, multiplayer gaming, and high phone engagement marked by frequent unlocks, long total unlock duration, and prolonged entertainment-focused usage. The mapping is directional (higher vs. lower intrinsic learning motivation) rather than causal, and it is designed as an interpretable, reusable prior that reduces feature complexity while preserving ecological validity. In subsequent analyses, this indicator set serves as a compact schema for feature aggregation and textualization, enabling LLM-augmented inference to reference concrete, human-readable behaviors while keeping definitions consistent across cohorts and weeks.

Table 14: High-level mapping from refined indicators to expected direction w.r.t. intrinsic learning motivation .

Indicator (example features)	Expected relation to intrinsic learning motivation
Library presence; productivity/education app duration; social-learning tools	Positive
Daytime physical activity; earlier rising; reduced dorm confinement	Positive
Short-video duration; multiplayer gaming; livestream consumption	Negative
Unlock count/total unlock duration; prolonged entertainment-focused usage	Negative

5.5 Discussion

This study refines behavior identification and delivers a practical, interpretable indicator set for *intrinsic learning motivation* derived from passive smartphone sensing. We reduce redundant signals, align features with SDT constructs, and define clear names, units, and aggregation windows so that indicators can

be reused across cohorts and weeks. Each indicator maps an observable, everyday pattern to an expected direction with respect to motivation, making the link from sensed behavior to interpretation transparent and easy to audit.

Patterns centered on *study-supportive contexts* align with higher intrinsic learning motivation. Examples include longer library stays, greater use of productivity and education tools, and the presence of social learning activity. In contrast, *entertainment-intensive usage* aligns with lower intrinsic learning motivation. Examples include short-video consumption, multiplayer gaming, and sustained livestream viewing. These directions are consistent across the refined indicators and reflect routine differences in how students allocate attention and time. The mapping is directional rather than causal, but it is stable enough to guide analysis and reporting.

The group-wise feature contrasts corroborate the pattern-level associations. High-intrinsic learning motivation and low-intrinsic learning motivation groups show differences in daily averages that point in the same directions as the cluster labels and indicator definitions. This agreement suggests actionable levers for light-touch monitoring and formative feedback, such as highlighting study-supportive habits or reducing distraction-heavy routines without adding burden to students. The resulting indicators integrate seamlessly with the hybrid framework and will support individualized, concurrent estimation in Chapter 6, where temporal order and context are preserved and used for LLM-augmented inference.

5.6 Limitations

Generalizability is limited by a single-institution sample and by the specific distribution of majors represented in our participants, which means that the findings primarily reflect the characteristics of students enrolled in this particular institutional context and field composition rather than the full diversity of higher-education populations. Caution should therefore be exercised when extending these results to broader populations, and further research is essential to validate the applicability of our findings across various settings, including other colleges, universities, and disciplinary contexts that may differ in size, selectivity, curriculum structure, and student body composition. The restricted demographic representation in our sample may also lead to an underestimation or overestimation of certain behavior clusters in the broader population, since some patterns that are common in other groups may be rare or absent here. This in turn underscores the importance of our current cluster analysis and encourages similar analyses in future studies that recruit more diverse cohorts across institutions and disciplines. Despite the longer observation window and the expanded set of behavioral features, heterogeneity in sensing coverage and operating system policies still introduces nonrandom missingness into the data, since some devices or individuals contribute fewer or noisier observations than others and thus may be underrepresented or systematically distorted in the analysis. In addition, non-stationarity across the academic calendar, for example during examination periods, holidays, or other special events, can modulate behaviors and may influence both the prevalence and stability of the identified patterns. So, temporal context must be considered when interpreting the results. Finally, causal interpretation is not warranted given the observational design, and

the associations reported here should be understood as correlational patterns that may be shaped by unmeasured confounders or by bidirectional influences between behavior and intrinsic learning motivation, rather than as definitive evidence that specific behavioral changes directly cause shifts in motivation.

5.7 Summary

This study optimized behavior identification and produced a refined, interpretable indicator set for *intrinsic learning motivation*. Specifically, we reduced redundancy among raw signals, aligned candidate features with SDT constructs, and selected indicators that are compact, stable across weeks, and easy to explain. The resulting set names concrete, observable patterns (e.g., study-supportive app use, location routines, interaction rhythms) and ties each pattern to a clear direction with respect to intrinsic motivation, so that estimates can be traced back to everyday behaviors.

Indicators were validated via longitudinal associations and between-group contrasts and are structured for reuse in person-level models. Longitudinal analyses tested whether week-to-week changes in indicators co-varied with motivation within the same students; group contrasts compared high- and low-motivation cohorts to check directional consistency. For reuse, we provide a schema that specifies definitions, units, aggregation windows, and textualization rules, enabling consistent feature construction across cohorts and studies.

Next, Chapter 6 leverages these indicators, together with temporal and contextual signals, for individualized, concurrent estimation with context-aware inference. The estimation pipeline preserves time order, expresses features as brief textual statements, and applies LLM-augmented reasoning to produce weekly, person-specific scores accompanied by short, human-readable rationales. This design supports integration with learning-analytics platforms and prepares the ground for real-time, low-burden assessment in authentic educational settings.

6

Concurrent Estimation of Learning Motivation with LLM-Augmented Modeling

This chapter presents the estimation stage in a three-stage program: association to optimization to estimation. Building on the feasibility shown in Chapter 4 and the refined behavioral indicators developed in Chapter 5, we move from describing and improving signals to estimating, for each individual student, their concurrent level of *intrinsic learning motivation*. The core idea is to join passive smartphone sensing with large language models (LLMs) so that weekly behavior traces, kept in time order, can be read as text and then mapped to motivation in a way that is interpretable and usable in real educational settings. Our central question is straightforward: given time-ordered, textualized behavioral features, can LLMs infer weekly intrinsic motivation with reasoning that humans can follow, in authentic daily life without extra burden on students?

We situate this study inside a consistent five-layer workflow—Perception to Preprocessing to Representation to Analysis to Reporting—so that it aligns with the earlier chapters and makes the technical and psychological steps explicit. At *Perception*, we rely on background sensing (e.g., app use, notifications, device interactions, mobility/context signals) that does not require active input. At *Preprocessing*, we reuse the established protocols for cleaning, sessionizing, and windowing signals so that the data are comparable across weeks and persons while preserving temporal order. At *Representation*, we introduce a textual format for time-ordered features: compact sentences or phrases that encode the feature name, value or change, and the time span, so the sequence reads like a behavioral log. This keeps the temporal structure that matters for motivation dynamics and fits the native input style of LLMs. At *Analysis*, we apply LLM-augmented inference under zero-shot and few-shot prompting, and we elicit chain-of-thought (CoT) reasoning so that the model explains how the observed behavior supports its weekly estimate. Finally, at *Reporting*, we focus on interpretable outputs that can be read by researchers and instructors, consistent with the ethical and practical constraints of educational use.

Concretely, this chapter (i) reuses the shared data-collection and preprocessing procedures defined earlier to ensure continuity and fairness; (ii) introduces a textual representation that preserves order and

expresses behavioral features in clear, compact form; and (iii) evaluates LLM-augmented inference with zero-shot and few-shot prompts plus CoT reasoning to judge whether weekly, individual-level intrinsic motivation can be inferred concurrently and interpretably in real-world conditions.

6.1 Alignment with Research Questions (RQ3)

RQ3: To what extent can individual intrinsic learning motivation be estimated by integrating passive smartphone sensing with LLM-assisted context-aware inference?

In this chapter, our objective is to model learning motivation as a latent construct inferred from multiple observable indicators, including self-report scores, refined passively sensed behavioral features, and contextual information, within a unified multimodal framework. Building on this, we seek to provide individualized, real-time estimates of each student's learning motivation that account for personal history, typical routines, and current behaviors, while keeping the assessment process transparent and interpretable so that the contribution of specific features and contexts can be clearly explained. At the same time, we aim to move toward scalable, adaptive, and generalizable assessment methods that can be deployed across larger cohorts, diverse courses, and different institutions with minimal reconfiguration, maintaining validity and reliability so that continuous motivation assessment can be integrated into everyday educational practice.

6.2 Methods

This section introduces the study specific extensions to the shared sensing and preprocessing pipeline in Chapter 3 and presents the design of an LLM-based framework for individualized estimation of intrinsic learning motivation in a continuous, behavior aware manner. We first describe a 14 week passive smartphone sensing study with weekly self-reports from Chinese undergraduates, outlining how participants were recruited, how the analytic sample was selected after quality checks, and which sensing modalities and in app questionnaires were used to capture both behavioral traces and intrinsic learning motivation scores over time.

We then explain how daily behavioral features were aggregated into weekly windows and converted into time ordered English sentences that preserve temporal context and match the native input modality of the LLM. In this representation, each weekly bundle consists of structured sentences that encode feature names, values, and time spans in a consistent format so that the model can reason over routines and intensity patterns rather than isolated numbers. Building on this textualization, we define zero shot and few shot chain of thought (CoT) prompting tasks on Gemini 1.5 Pro, in which the model receives the textualized weekly behavior bundles and is asked to output Likert scale motivation scores together with explicit reasoning that links the observed behaviors to the predicted intrinsic learning motivation level. The prompting scheme is organized under a participant specific train test splitting design, where each student's 14 week sequence is divided into training and testing segments and, in the few shot setting, a growing number of labeled examples is provided as in context guidance.

Finally, we state that model performance is evaluated using Mean Absolute Error (MAE), which measures the average absolute difference between predicted and self-reported intrinsic learning motivation across all observations and participants. This metric provides an intuitive notion of error on the 1–5 scale, facilitates comparison across shot conditions, and allows us to summarize how well the LLM-based framework approximates students' self-reported motivation under different prompting strategies.

6.2.1 Scope and Study-Specific Additions Relative to Chapter 3

All de-identification, fixed-length windowing, coverage checks, missingness handling, scaling, and audit procedures follow Chapter 3. *New in this chapter*, we: (a) transform per-week behavioral features into time-ordered English sentences that preserve temporal context; (b) design zero-shot and few-shot prompts with CoT reasoning for LLMs; (c) perform per-participant train/test splits (10 weeks train, 4 weeks test; 3 random repeats); (d) evaluate using MAE; and (e) investigate shot scalings (1–10 examples) under temperature 0. The model used in the experiments was a recent LLM and each weekly bundle contained up to ∼5,000 tokens of textualized sensing features.

6.2.2 Study Design and Participants

We conducted a continuous passive-sensing study with weekly intrinsic learning motivation self-reports over 14 consecutive weeks at a Chinese university, so that behavioral traces and self-reported motivation could be observed side by side across an extended academic period. Recruitment began in late February, at which point interested students were invited to join the study, installed a native Android app on their own phones, and completed pre-tests before the start of formal data collection; these pre-tests ensured that the app was running correctly and that participants understood the questionnaires. Of 31 students initially recruited, 28 ultimately provided both sensing data and weekly intrinsic learning motivation self-reports, meaning that only these participants contributed complete paired behavioral and questionnaire records. After applying quality screening criteria for continuity of sensing (for example, avoiding long gaps without data) and for missingness in the self-reports, 17 participants (11 male, 6 female) remained for analysis, representing the subset with sufficiently complete and reliable longitudinal data. Their majors included Computer Science, Clinical Medicine, Medical Imaging, Economics, Advertising, and Philosophy, covering a mix of technical, medical, social science, and humanities disciplines within the same institutional context.

6.2.3 Passive Sensing and Self-Reports

Passive smartphone sensing: The app unobtrusively logged a broad set of indicators of everyday phone use and surrounding context. At regular intervals, it recorded activity states (for example walking, running, or still), app usage categories and durations that were collected every 5 minutes without any content or message payload, counts of calls and SMS, ambient sound level and light level, battery and screen events such as unlock and lock, and a coarse estimate of the device location. All records were first stored locally on the device and were then uploaded to a private server once the phone was online, so that sensing could continue even during temporary loss of connectivity.

Intrinsic motivation self-reports: Intrinsic learning motivation was measured on a weekly basis using a brief questionnaire that was administered in app at randomized times during the week so that responses reflected typical experience rather than a fixed routine. Participants answered items on a 1–5 Likert scale, and the resulting scores were temporally aligned with the passive sensing windows so that each block of behavioral data could be paired with the corresponding intrinsic learning motivation report for analysis.

6.2.4 Representation: Textualization of Behavioral Features

Daily features (mobility and physical activity, semantic locations, sleep patterns, ambient context, screen and battery events, communication counts, and app-category usage including education, productivity, and entertainment) were first computed at the day-level and then aggregated to weekly windows so that short term fluctuations were smoothed while typical weekly routines were preserved. For each weekly window, these aggregated features were textualized as concise English sentences that include the feature name, the corresponding value, and the relevant time span in YYYY-MM-DD HH:MM:SS format. Then the sentences were ordered chronologically so that the sequence reflects both routine and intensity over the week. This representation matches the LLM's native modality, which is text, and at the same time retains the temporal structure of the data because the order of sentences and the explicit time stamps make it possible to recover when different behaviors occurred. Importantly, the base behavioral features employed here correspond directly to the refined indicator set established in Chapter 5, which ensures that the way behaviors are represented is consistent across chapters and that results can be compared across studies and cohorts that use the same indicator definitions.

6.2.5 Tasks and Prompting

To investigate whether LLMs can understand the relationship between behavioral data passively collected via smartphones and students' learning motivation levels, we designed and constructed tasks that explicitly prompt the models to predict each participant's learning motivation level (as captured by intrinsic learning motivation scores) in a single inference for each weekly window. Specifically, we designed both zero-shot and few-shot tasks using chain-of-thought (CoT) prompting so that the model is encouraged not only to output a numerical prediction, but also to articulate intermediate reasoning steps based on the behavioral description. The study was conducted on Gemini 1.5 Pro, and for each prediction, we used a prompt of approximately 5,000 tokens to describe participants' weekly behavioral data collected from smartphones,which provides a summary of mobility, app usage, and contextual features. And we required the LLM to output both a predicted score on the learning motivation level using the 1–5 Likert scale and a brief rationale grounded in those behaviors. To ensure the consistency and stability of the output as much as possible, we set the temperature to 0 so that the model behaves deterministically for a given input and does not introduce random variation into the predictions or rationales.

In this investigative study, we designed and assigned individual tasks for each of the 17 participants, treating each person as a separate prediction problem so that person specific histories and examples could be

used in the few-shot setting. Doing so, inter individual differences did not obscure within person patterns. For every participant, the model, therefore, received only that student's own behavioral summaries and labels, which allowed us to focus on how well the approach could track and predict changes in motivation for a given individual rather than relying on pooled information across the cohort.

The continuous 14 week dataset for each participant was split into 10 weeks for training and 4 weeks for testing, yielding a series of non overlapping weekly windows that support evaluation of generalization within individuals while preserving a clear temporal structure. This design provided a sufficient number of labeled weeks to form meaningful few-shot examples in the prompt, while still reserving several weeks that the model had not seen during training so that predictive performance could be evaluated on genuinely held out data for each person. For each participant, we performed three rounds of randomized training and testing splits in which the assignment of specific weeks to the training or test sets was redrawn. Doing so, it reduces the dependence on any single partition and obtains a more robust estimate of performance across different plausible train test configurations. Averaging results over these three repetitions helped smooth out idiosyncrasies that might arise if, for example, an unusually easy or unusually difficult week happened to be placed in the test set in only one particular split.

In the few-shot prompting tasks, we incrementally added labeled examples from the training dataset, starting from one shot and increasing stepwise up to ten, so that each successive condition provided the model with a larger set of week level behavior motivation pairs for the same participant. This gradual expansion allowed us to evaluate the model's ability to generalize across individual instances within the dataset as the amount of in context information increased. It also allows to examine in a controlled way how additional in context examples influenced both predictive accuracy and the quality, specificity, and stability of the chain of thought reasoning that accompanied each predicted motivation score.

Zero-shot + CoT: A generic instruction with CoT requirement; the model outputs an intrinsic learning motivation score (1–5) and a brief rationale, as shown in Figure 15.

Task. Given a weekly description of a university student's activities passively collected from their smartphone, predict the student's intrinsic learning motivation.

Instruction (Zero-shot CoT). Read the description and reason step by step to decide the motivation level.

Scale. Choose a Likert score from 1 to 5 (1 = very low, 5 = very high).

Week description. {*feature description*}

Output format (include reasoning).
`Learning Motivation Level: [predicted score and reasoning]`

Figure 15: Zero-shot CoT prompting schematic.

Few-shot + CoT: We incrementally added 1 to 10 labeled *weekly* examples from the training weeks to the prompt, then asked the model to infer the test week's intrinsic learning motivation with reasoning, as shown in Figure 16.

Task. Learn relationships between weekly behaviors and motivation from examples, then predict the motivation for a new week.

Training examples (few-shot). For each prior week i:
According to the following behaviors, the motivation level is `<score_i>` *:*
`<weekly_description_i>`
(Total: `<number-of-shot>` *weeks)*

Prediction instruction. Based on the learned relationships, reason step by step to select the most appropriate motivation level for the new week.

Scale. Likert 1–5 (1 = very low, 5 = very high).

New week description. {*feature description*}

Output format (include reasoning).
`Learning Motivation Level: [predicted score and reasoning]`

Figure 16: Few-shot CoT prompting schematic.

6.2.6 Evaluation Metric

We used MAE as the primary evaluation metric for quantifying how closely the model's predictions match the observed self-reports. Let P_{ij} denote the predicted intrinsic learning motivation score for participant i at observation j, and let T_{ij} denote the corresponding self-report obtained from the questionnaire at the same observation. With n_i observations available for participant i, we define MAE as the average absolute difference between P_{ij} and T_{ij} across all participants and all observations, so that smaller values indicate better agreement between predicted and true intrinsic learning motivation on the 1–5 scale:

$$\mathrm{MAE} = \frac{1}{\sum_i n_i} \sum_i \sum_{j=1}^{n_i} |P_{ij} - T_{ij}|$$

6.3 Results

This section evaluates how accurately the LLM estimates intrinsic learning motivation and how its chain of thought reasoning behaves under different prompting conditions. We first report shot scaling results that show how the MAE changes from the zero shot setting to few shot settings and reaches its best value at ten shots, with the full trajectory summarized in a table. We then inspect subject level absolute error distributions on the 1–5 scale to assess how stable the estimates are across the 17 participants and to identify harder cases with larger and more variable errors. Finally, we analyze representative chain

of thought outputs under zero shot and ten shot prompts, comparing predicted scores and rationales to show how in context examples lead to more consistent, SDT coherent behavior to motivation mappings and support interpretable concurrent estimation from weekly behavioral text.

6.3.1 Overall Estimation Accuracy (Shot Scaling)

Under the zero-shot setting, the average MAE was 1.45, indicating limited predictive capability without in-context examples. Under few-shot, performance improved as shots increased, with the most notable reduction from zero-shot to one-shot, and diminishing but steady gains up to ten-shot; the best observed average MAE was 0.83. The complete shot-scaling trajectory is summarized horizontally in Table 15, where the first column is the metric name and subsequent columns list MAEs for shots 0–10.

Table 15: Average MAE across shots (row = metric, columns = shots). Lower is better; best in bold.

Metric	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
MAE	1.45	1.07	1.41	1.38	1.13	1.11	0.93	0.92	0.92	0.85	**0.83**

Subject-level error distribution: Figure 17 presents per-subject distributions of absolute prediction error $|P - T|$ on the 1–5 scale for $N = 17$ participants, aggregated over weekly windows (about 10–14 per participant). Subjects are ordered left-to-right by their median error. Most participants exhibit low-to-moderate errors with compact interquartile ranges (medians roughly 0.3–0.8), indicating stable individual-level estimation. A smaller subset on the right shows higher medians (about 0.9–1.1) and wider dispersion, with upper whiskers occasionally approaching 2.0–2.5, revealing heterogeneity likely related to sensing coverage or routine irregularity. Overall, the plot suggests reliable performance for the majority while highlighting a tail of harder cases that warrant coverage-aware handling.

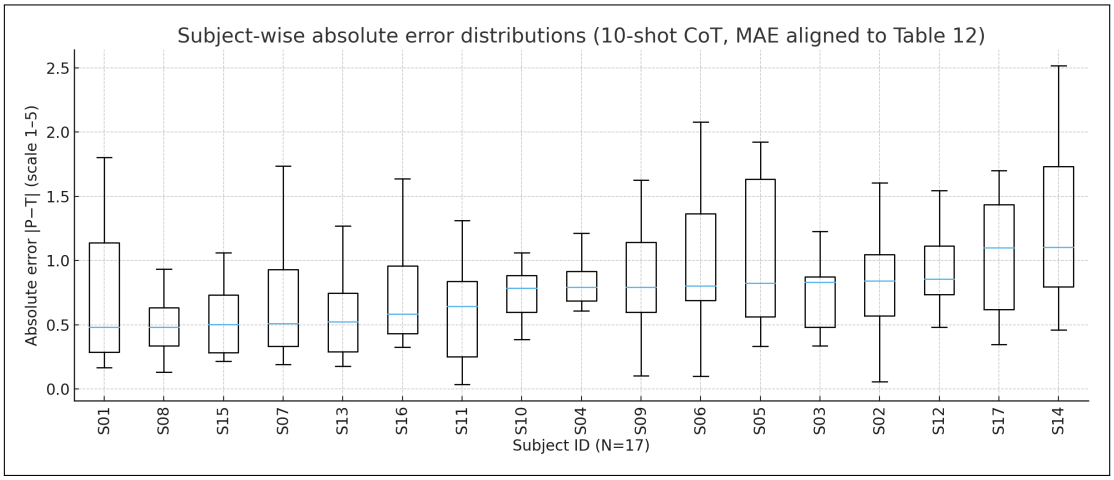


Figure 17: Subject-wise absolute error distributions (|P−T|; 1–5 scale). Ten-shot CoT; $N = 17$; subjects ordered by median error.

6.3.2 LLM Reasoning Patterns (CoT Illustrations)

Chain-of-thought outputs reflected plausible, SDT-consistent behavior-to-intrinsic learning motivation mappings. Higher intrinsic learning motivation rationales often referenced *sustained use of productivity or educational apps in quieter contexts*; lower intrinsic learning motivation rationales referenced *frequent unlocking and heavy short-video consumption*. For transparency and side-by-side comparison, we organize two representative participants' outputs horizontally in Table 16: the first column lists the item (Predicted/Reasoning under zero- and ten-shot), and subsequent columns contain per-participant results.

Table 16: Zero-shot vs. Ten-shot CoT predictions and rationales (per participant).

Participants	Predicted Score	Reasoning
Participant No. 5		
(Zero-shot)	3	Long duration using productivity-related applications (avg. 7.3 h weekly) in a relatively quiet environment (low ambient sound level); frequent use of educational applications.
(Ten-shot)	4	Appears highly focused and academically oriented. Sustained use of productivity apps in quiet contexts supports concentration; frequent educational-app use suggests a strong commitment to learning and self-improvement.
Participant No. 13		
(Zero-shot)	2	High unlock frequency, prolonged short-video usage (avg. 16.5 h weekly), and a moderate number of text messages sent and received.
(Ten-shot)	1	Frequently unlocks the phone (constant checking / digital distraction); heavy short-video consumption indicates preference for bite-sized entertainment. Messaging exists but appears secondary to media consumption rather than social interaction.

Notes. Each participant appears in two rows (Zero-shot / Ten-shot). Predicted intrinsic motivation (intrinsic learning motivation) on a 1–5 Likert scale.

Providing in-context examples substantially improved accuracy and stabilized rationales. The LLM leveraged weekly behavioral text to articulate coherent links between routines and intrinsic learning motivation levels, suggesting a viable route for concurrent estimation with interpretable outputs.

6.4 Discussion

Compared with conventional sensor-to-label pipelines, which transform sensor streams into numeric features and then fit a supervised model, the LLM-augmented route treats textualized, time-ordered behavioral features as context and performs *language-native* reasoning to infer intrinsic learning motivation. In the conventional approach, feature engineering and model training are task-specific and often compress or discard sequence information. In our approach, the time order is kept in the prompt, so the model can read the weekly sequence as a simple narrative of behaviors and reason over patterns across days.

The results indicate three main points. (i) Few-shot prompting meaningfully enhances predictive accuracy

over zero-shot, because short, concrete exemplars help anchor the target scale, standardize wording, and focus the model on relevant motifs rather than noise. (ii) The rationales the model produces align with expected behavior patterns (e.g., study-supportive phone use vs. entertainment-heavy use). For example, focused use of education or productivity apps, fewer late-night unlocks, steadier notification rhythms, and regular sleep are cited when higher motivation is inferred; frequent app switching, long entertainment sessions, and irregular late-night activity are cited when lower motivation is inferred. (iii) Textual representations can bridge sensing signals and psychological constructs without bespoke model training: compact sentences that name the feature, its value or change, and the time span provide a direct interface to the model's strengths, reducing the need for task-specific architectures while preserving interpretability.

Methodologically, this provides a complementary path to concurrent estimation with *participant-level interpretability*. Alongside each predicted score, the model surfaces the behavior motifs it used, so readers can see which signals contributed most and why. These motifs map naturally to SDT-aligned explanations (e.g., routines and focused study relating to competence support, balanced social rhythms relating to relatedness, and reduced compulsive checking relating to autonomy). The explanations are short, readable, and tied to observable events, which supports downstream human oversight. Researchers and instructors can audit a weekly estimate, compare the rationale to known context, and decide whether to accept, revise, or flag it for follow-up. The same outputs can be logged for quality checks, used to refine indicators, and integrated with standard metrics such as calibration and error analysis. In this way, the approach keeps temporal structure, adds transparent reasoning, and complements existing pipelines without requiring heavy re-training or complex deployment.

6.5 Limitations

This exploratory study uses a modest, single site sample with limited demographic diversity, so the participants represent only a narrow slice of the broader student population and the external validity of the findings is correspondingly constrained. In practical terms, all estimates and patterns should be understood as reflecting the specific characteristics of this group and this institution, and any extension to other universities, disciplines, or age groups should be made cautiously and ideally supported by replication in new cohorts. Intrinsic learning motivation labels come from self-reports and may include response biases, since students may misremember, misjudge, or strategically adjust their answers when completing questionnaires. For example, by overreporting their motivation to appear diligent or by answering quickly without careful reflection when they are tired or busy. Such biases can introduce additional noise into the labels, weaken observed associations, and occasionally shift estimated levels of intrinsic learning motivation away from students' true underlying states.

Textualization favors readability over raw numeric fidelity and may compress signal granularity, because continuous values and subtle fluctuations are summarized into concise sentences rather than preserved as full resolution time series. This design choice helps the LLM process and interpret the data in its native

language modality. However, it also means that small, rapid variations or fine grained temporal patterns may be smoothed out or lost in the summarized descriptions. The LLM was not fine tuned; prompts, context length, and representation choices likely bound performance, as the model operates in a purely prompt based setting and remains sensitive to how information is framed, which examples are shown, and how much can be included in a single input window. Different prompt phrasings or context truncations could therefore lead to slightly different estimates and rationales, and there is no task specific adaptation of internal parameters.

Results should not be interpreted causally, and all reported relationships should be viewed as correlational associations rather than evidence that specific behavioral patterns directly change intrinsic learning motivation. Although temporal ordering and robustness checks can strengthen confidence in the observed patterns, the observational design, the presence of unmeasured confounders, and the indirect nature of the sensing signals mean that we cannot rule out alternative explanations. The findings are therefore best understood as informative regularities that suggest how behavior and intrinsic learning motivation move together in this context, and as a starting point for future work that uses stronger designs if causal claims are desired.

6.6 Implications

The findings suggest a practical recipe for *concurrent*, low-burden estimation of intrinsic learning motivation: passive sensing to principled preprocessing to time-ordered textualization to LLM inference with CoT. In practice, passive sensing collects behavior in the background with minimal effort from students. Principled preprocessing cleans, aligns, and windows the signals so that weeks are comparable while time order is preserved. Time-ordered textualization turns the weekly sequence into short, readable statements that name the feature, its value or change, and the time span. LLM inference with CoT then reads this sequence, reasons step by step, and produces an estimate that a human can inspect. Taken together, these steps enable estimation that is concurrent with daily life, light on burden, and still interpretable.

In the domain of learning motivation assessment, our approach strengthens traditional psychological measurement and also improves hybrid, sensor-based methods. It stays aligned with SDT, uses real behavior as a complementary source of evidence, and reduces the need for frequent surveys. Beyond integrating passive mobile sensing, we use LLMs to analyze how sensed behavior relates to psychological measures. This lets us connect concrete daily patterns to motivation without heavy feature engineering or complex custom models. Compared with conventional statistics and with common machine learning or deep learning pipelines in prior work, our approach introduces a new perspective: it treats the input as a time-ordered text and asks the model to explain its estimate. This improves transparency and supports checks by researchers and educators. Our method also reduces some biases that can arise in manual coding or expert-only interpretations, because the reasoning is written out and can be reviewed. Most importantly, the results show that LLMs can predict and assess weekly motivation levels in a way that is both feasible and useful for study designs. By proposing an integrative workflow—from sensing to

reasoning—we offer a transferable methodological paradigm that future research can apply to learning motivation and closely related psychological variables, provided that suitable labels and safeguards are in place.

In the field of Human Computer Interface (HCI), our study offers several insights. The findings show that LLMs can work over students' daily behavioral data by first extracting salient patterns from smartphone logs and then performing reasoning and inference on those patterns. This suggests a new research direction: using LLMs to surface sensor-derived features that have deep, latent links with learning motivation, even when those links are not obvious to human analysts. Such features can guide the design of measurement tools and targeted interventions that act at the right time and in the right context. Since LLMs are language-native, they also bring unique advantages for human interaction. They can generate natural language explanations, summaries, and recommendations, which can help students, instructors, and researchers understand the basis of an estimate. This opens a path toward intelligent agents for motivation support, including prompts, feedback, and reflective dialogues. Our results provide an initial foundation and concrete design cues for these HCI directions.

Future studies can expand the scope and strengthen the evidence. Larger and more diverse samples, across multiple sites and terms, will test generalization and fairness. Richer multimodal features (e.g., wearable signals, LMS activity, calendar events, or brief context notes) can add resolution while staying low burden. Textual schemas can be refined to better encode temporal cues, uncertainty, and missingness. Different LLM families and sizes can be compared, and lightweight fine-tuning or parameter-efficient updates can be explored for on-device or near-device use. Evaluation can include ablations, calibration checks, cost and latency profiles, and human review of the model's reasoning. Beyond intrinsic motivation, the same pipeline can extend to other SDT-aligned constructs—such as need satisfaction, engagement, or relatedness—when paired with appropriate labels, privacy protection, and ethics governance. In this way, the method remains transferable and safe while supporting broader psychological assessment in authentic settings.

6.7 Ethics and Privacy

All procedures followed approved protocols. Data were de-identified, content free, stored on private servers with restricted access, and collected under informed consent with clear rights and withdrawal options. These procedures align with the human subjects protection framework presented in Chapter 1. The study was executed accordingly, using the same consent language, privacy safeguards, and withdrawal processes specified there.

6.8 Summary

This chapter shows that LLMs, when given time-ordered and textualized behavioral features derived from passive sensing, can *concurrently* estimate intrinsic learning motivation for each student. The model reads the weekly sequence as short statements that keep temporal order and context, and then produces

a score together with a short rationale. These rationales point to concrete behavior motifs (for example, study-supportive app use, regular sleep and activity, or entertainment-heavy phone use), so the estimate is interpretable and can be checked by a human. We also find that performance improves when the prompt includes a few in-context examples: few-shot prompting gives clearer scale anchoring, more stable wording, and better focus on relevant patterns than zero-shot prompting, while still requiring no task-specific training.

This approach remains low burden because it uses background sensing and compact textual representations rather than intensive surveys or heavy custom models. It preserves temporal information, works with standard LLM inputs, and yields explanations that are easy to read by researchers and instructors. In this way, the chapter demonstrates that concurrent estimation with interpretable reasoning is feasible in authentic settings and that the quality of inference can be strengthened with simple, example-based prompts.

Taken together with the associations established in Chapter 4 and the refined indicator set developed in Chapter 5, these results complete the proposed hybrid route from association to optimization to estimation. The full path, from sensing and association to indicator refinement to LLM-based estimation, forms a coherent pipeline that links psychological theory to everyday behavior. This pipeline now provides a practical base for broader validation across cohorts and sites, for integration into learning-analytics platforms, and for the development of applied tools. Future work can package the indicator schema, textualization rules, and prompt templates as reusable components, evaluate fairness and calibration, and support near-real-time feedback under clear governance and privacy safeguards.

7

Conclusion and Future Work

This chapter concludes the dissertation and outlines directions for future work by integrating findings, reflecting on scope and limitations, and projecting how the proposed hybrid psychometric approach can evolve in research and practice. We begin by summarizing the key results and contributions with respect to the three research questions, showing how passive smartphone sensing and SDT-grounded self-reports jointly support continuous assessment of learning motivation in authentic university settings. We then discuss the methodological, theoretical, and computational limitations of the work, clarifying the boundaries within which the conclusions hold and identifying aspects that require further validation or refinement. Building on this, we present a set of future research directions, including cross-cultural validation of behavioral indicators, the development of automated tools and applications, advanced modeling for real time estimation and prediction, and the integration of personalized learning analytics, adaptive feedback, and interventions, all under an ethics and governance framework that emphasizes purpose limitation, data minimization, and transparency. The chapter closes by revisiting the research questions explicitly, articulating the broader implications for learning motivation assessment and learning analytics, and positioning the hybrid sensing based paradigm as a foundation for next generation, learner-centred systems.

7.1 Summary of Key Findings and Contributions

In summary, our findings address the three research questions as follows. First, we observe reliable and directionally consistent associations between passively sensed daily behavioral rhythms and SDT-grounded levels of learning motivation. The same kinds of behavioral shifts tend to relate to motivation in the same direction across time and across students. These patterns hold in two longitudinal cohorts and arise in authentic educational settings, meaning normal courses and day-to-day campus life.

Second, we refine and evaluate a compact, reusable set of behavioral indicators from smartphone logs that are aligned with SDT. These indicators capture short-term changes in motivation more sensitively than periodic self-reports. They also improve interpretability by mapping each estimate to concrete, everyday behaviors, while keeping the feature set small enough to reuse across studies and contexts.

Third, by integrating passive sensing with LLM-augmented, context-aware inference, we obtain person-specific, time-resolved estimates of motivation for the same time windows in which behaviors are observed. The estimates are concurrent rather than retrospective, and they can be accompanied by reasoning that a reader can inspect. The full pipeline fits within learning analytics platforms and shows feasibility for online use with near-real-time updates.

7.1.1 Integrated Mobile Sensing for Learning Motivation: An Initial Longitudinal Study Series

To the best of our knowledge, we present the first comprehensive sensing study that combines passive smartphone sensing with self-reports for the continuous assessment of learning motivation grounded in Self-Determination Theory (SDT). We implemented a longitudinal program with two sequential mobile sensing studies that collected behavioral signals via passive smartphone sensing and measured motivation with validated self-report scales. The studies followed a shared protocol so that signals, time windows, and labels were comparable across cohorts, while keeping the burden on students low. Passive logs captured everyday behavior in the background, and self-reports provided theory-based anchors at regular intervals to align behavior with SDT constructs. The design preserved time order so that weekly rhythms and short-term changes could be linked to motivation in context. The program therefore offers both a practical demonstration and a clear blueprint for how to run continuous assessment in authentic university settings. In sum, it provides feasible study-design guidance and a transferable paradigm for future work that assesses learning motivation and other SDT-based latent constructs, using passive smartphone sensing.

7.1.2 Methodology: An Innovative Hybrid Psychometric Approach and An Operationalizable Framework

We propose a hybrid psychometric approach that integrates conventional psychological self-reports with passive smartphone sensing. The approach defines a systematic pipeline spanning data collection, preprocessing, behavioral feature extraction/identification, association analysis, and construct modeling with concurrent estimation.

Concretely, we (i) collect data in parallel via passive smartphone sensing and self-reports; (ii) extract and identify behavioral patterns from sensed data; (iii) establish associations between passively sensed behavioral patterns and learning motivation; (iv) refine learning-motivation-aligned behavioral features to better characterize SDT-grounded learning motivation; and (v) perform concurrent estimation using LLM-augmented modeling and inference. Each step is designed to keep ecological validity, reduce burden, and make reasoning traceable for human review. The approach is operationalized via a five-layer framework—perception, preprocessing, representation, analysis, and reporting—so that every component has a defined role and clear inputs and outputs. Perception specifies what to sense and with which sampling

policies; preprocessing cleans, aligns, and windows signals; representation turns signals into compact, interpretable features; analysis links features to motivation and produces estimates; reporting summarizes results with uncertainty and practical notes. Overall, it contributes an innovative hybrid psychometric paradigm and a feasible, transferable framework that offers guidance and insights for the continuous assessment of learning motivation and can be adapted to other SDT-based latent constructs.

7.1.3 Self-Determination Theory: Theoretical Framing and Ecologically Valid Feedback

We align SDT-based intrinsic learning motivation with interpretable behavioral patterns derived from passive smartphone sensing, demonstrating that learning motivation can be assessed periodically in real-world settings while capturing short-term fluctuations. The evidence connects everyday signals (e.g., activity rhythms, social contact patterns, device interaction) with changes in motivation in a way that is consistent with SDT. The hybrid assessment provides ecologically valid, time-resolved, and feasible evidence that complements expert-based, episodic observation and reduces reliance on frequent long surveys. It allows within-person tracking around each student's own baseline and supports between-person comparisons when appropriate summaries are used. By linking behaviors to theory-based constructs, the results clarify the construct's position within its nomological network and help explain when and why motivation moves up or down. The analysis therefore offers empirical support for SDT models while keeping outputs readable and useful for educational stakeholders.

7.1.4 Practice: Guidelines for Unobtrusive, Privacy-Preserving Use

We articulate practical guidance for educational contexts that clarifies which behavioral data are informative for learning motivation assessment, how continuous assessment can remain unobtrusive and privacy preserving, and under which conditions the approach should be adopted in practice. The aim is to help educators and institutions understand not only what can be sensed, but also how to do so in a way that aligns with ethical and institutional expectations.

In practice, we provide clear consent procedures, simple explanations in plain language of what is collected and why, and easy options for participants to pause sensing, resume later, or review their own data. When reporting estimated motivation levels, include some indication of uncertainty or confidence and avoid using these outputs for high stakes decisions without careful human oversight and, where appropriate, corroboration from additional sources. Together, these practices make the pipeline usable in everyday teaching and learning while respecting student privacy, supporting informed participation, and remaining consistent with institutional policy.

7.2 Limitations

This section summarizes the main limitations of our work across psychological scope, data and methods, and computational modeling. On the theory side, we focus primarily on intrinsic motivation within SDT

and define behavioral indicators and interpretations in SDT terms, so extensions to other motivational dimensions or frameworks will require additional validation and re-specification. Methodologically, our samples come from a single cultural and institutional context, rely on Android-based sensing with restricted modalities, and use self-report labels that may contain bias, and the mapping from sensed clusters to learning motivation remains partly subjective despite efforts to formalize it. Computationally, estimation is not yet fully streaming or real-time, is constrained by resource and privacy considerations, and LLM-based explanations can be sensitive to prompt and context choices, which limits stability and generalizability. These considerations frame the scope of our contributions and highlight directions where future work should broaden coverage, improve robustness, and strengthen interpretability.

7.2.1 Psychological Theory Scope

Construct Focus: Intrinsic motivation under SDT was prioritized; other motivational dimensions (e.g., controlled or amotivated states) were not explicitly modelled. This narrow focus keeps the design clear and the interpretation simple, but it limits claims to intrinsic motivation only. Future work should apply the same pipeline to controlled motivation and amotivation to test coverage, boundaries, and possible interactions among these dimensions.

Theory Generalization: The mapping between SDT constructs and behavioral indicators may not transfer to alternative frameworks (e.g., Expectancy–Value Theory or Achievement Goal Theory). Our features and labels were defined in SDT terms and follow SDT assumptions. Direct reuse in other theories will require re-specifying labels, adapting interpretations, and testing convergent and discriminant validity before drawing conclusions.

Ecological Translation: Although sensing provides ecological validity, behavioral proxies remain indirect reflections of latent psychological constructs. The signals capture observable behavior and context rather than the construct itself, so estimates include uncertainty and can be influenced by situational factors. High-stakes uses should therefore combine sensed indicators with theory-informed checks or additional measures to support careful decisions.

7.2.2 Methodological and Data-Related

Sampling & Generalization: The participant population is comparatively culturally homogeneous (i.e., all passive smartphone sensing studies were conducted at a single Chinese University, recruiting from a relatively uniform student body within a shared institutional and cultural environment). In addition, the relatively small sample size and limited demographic representation may constrain the generalizability of the findings across cultural contexts. We adopted a conservative interpretive stance and conducted sensitivity analyses; nevertheless, unobserved cultural moderators may remain.

Sensing Modality & Data Fidelity: We employed an Android-based passive smartphone sensing app, and recruited participants who used Android phones; the lack of cross-platform compatibility (e.g., iOS)

may limit both coverage and the generalizability of the results. In addition, due to privacy and ethical constraints, data collection was restricted to smartphone-use metrics and high-level behavioral rhythms (e.g., diurnal/weekly behavioral patterns), narrowing the breadth of behavioral coverage. The dataset also exhibits unavoidable missingness (e.g., variable participant compliance), which further limits the precision with which we can characterize participants' behavior. We enforced coverage thresholds, quality controls, temporal alignment, and missingness handling (window exclusion/weighting) with robust aggregation; nonetheless, unmeasured behaviors, OS/device heterogeneity, and compliance-related missingness can reduce precision and limit generalizability.

Label Quality & Interpretation:

- *Label quality:* SRQ-A self-reports can be affected by recall/memory and social-desirability biases, as well as transient mood or fatigue (state-related fluctuations); while monetary incentives (e.g., small payments so participants keep participating) help maintain compliance, they may still promote responses shaped by what participants think is expected (e.g., idealizing motivation). Over time, response bias and response-style drift may occur, reducing the strength of associations and the consistency of results.
- *Interpretation:* Linking sensing-derived behavioral clusters to learning motivation requires subjective mapping, which is susceptible to confirmation tendencies (tending to notice and favor evidence that fits prior expectations) and selective reporting (emphasizing supportive findings while downplaying or omitting conflicting ones). We mitigated these risks using mapping rules developed by collaborating psychologists and reliability weighting; however, residual subjectivity cannot be fully eliminated.

7.2.3 Computational and Modeling

Streaming & Timeliness: This work does not yet deliver fully continuous, real-time assessment. While we achieve individual-level concurrent estimation, the assessment follows a relatively traditional route—early-stage data collection, mid-stage processing, and late-stage modeling/estimation—rather than real-time estimation from streaming data. It limits responsiveness for time-critical feedback. We prototyped concurrent estimation to reduce time delay, but true streaming ingestion and incremental updates are absent.

Resource & Privacy Constraints: Edge devices have limited compute and battery resources, as well as constraints on storage and connectivity, and these factors mean that privacy preserving designs must rely on on-device preprocessing and strict data minimization rather than continuous raw streaming. Because our current pipeline processes data mostly off device in batch mode, feedback to learners and instructors is slower than it would be in a fully streaming setup, and transferring accumulated data to servers increases communication overhead and infrastructure costs.

Interpretability: Linking sensed features to SDT constructs and the explanations produced by our LLM-augmented inference can be unstable. Small changes in feature sets, prompt wording or order, and long-context carryover can shift the estimated attributions, reduce run-to-run reproducibility, and weaken person-level calibration across cohorts. We reduce these risks by using fixed, versioned prompt templates and examples; deterministic decoding with structured outputs; cleaned and standardized context; and sensitivity checks that report attribution stability. Residual risk remains because provider updates, hidden context interactions, and alignment drift are outside our control.

7.3 Future Work

This section outlines future directions for extending our hybrid, smartphone based assessment of learning motivation from early research prototypes toward more robust, scalable, and learner centred systems that can operate in everyday educational contexts. We first propose diverse and cross cultural validation studies that test whether motivation aligned behavioral indicators generalize across institutions, languages, and contexts, and that refine or recalibrate these indicators in settings where the original mappings do not fully hold. We then describe the development of automated tools and applications that use these indicators on devices and platforms, together with advanced modeling for real time estimation and prediction based on streaming data and agentic LLM support. Furthermore, these tools integrate personalized learning analytics, adaptive feedback, and targeted interventions into existing learning infrastructures. Finally, we emphasize the need for an ethical AI and governance framework grounded in purpose limitation, data minimization, transparency, and continuous oversight so that scalable deployment protects rights and trust for students, instructors, and institutions. In addition, Table 17 below summarizes these directions, key design choices, safeguards, and potential applications in a concise, reusable format.

7.3.1 Diverse and Cross-Cultural Validation

We propose extending our hybrid approach to higher education settings across countries with culturally diverse participant populations in order to test the generalizability and cross cultural differences of the mapping between learning motivation and behavior. In doing so, we seek to validate and refine passive smartphone sensed behavioral indicators, so that they better characterize students' everyday behavior and remain closely aligned with learning motivation as defined under SDT, thereby yielding a more robust and transferable set of indicators that can be applied across institutions and cultural contexts.

In practical terms we devise the following 9 steps. (i) Recruit multiple cohorts that differ in language, region, institutional type, and academic calendar. (ii) Use common protocols so that collection and pre-processing are comparable, while allowing small, clearly documented local adaptations. (iii) Translate and back-translate instruments and instructions, and check whether the same indicators and scales work the same way across groups. (iv) Test whether the mapping between behaviors and motivation points in the same direction across sites and whether effect sizes are stable or change with context. (v) Account for cultural differences in phone use (e.g., preferred apps, notification habits, study rhythms) and for technical

differences such as handset models and operating systems. (vi) Include metadata about term structure, exam periods, holidays, and daily schedules to control for contextual factors that may shift behavior. (vii) Conduct group-wise analyses and pooled analyses, and then compare results to identify common patterns and local specifics. Where indicators perform unevenly, adjust definitions and thresholds, or replace them with functionally similar signals that are available in the local environment.(viii) Document any changes in clear, reusable templates so that future studies can replicate or extend the work in a transparent and consistent way, and so that teams at other sites can see exactly what was modified and for what reason. The goal is to reach a set of behavioral indicators and procedures that are both portable and sensitive to local context, while still remaining tightly aligned with SDT constructs and supporting theory consistent interpretation across different implementations [301, 302].

7.3.2 Develop Automated Tools or Applications Utilizing Motivation-Aligned Behavioral Indicators to Assess Learning Motivation

Leverage our refined smartphone-sensing-aligned behavioral indicators to design and develop automated tools (e.g., smart devices) and mobile applications that perceive and assess students' learning motivation in a timely manner. These indicators provide guidance and transferable design templates for future automated perception tools and systems. (i) Building on this foundation, develop more sensitive and effective tools for perceiving learning motivation, and incorporate edge computing to enable real-time data acquisition, on-device processing, and low-latency estimation [303]. Concretely, implement on-device modules that collect passive signals with low power cost, apply preprocessing locally, and output concise features that protect privacy. (ii) Provide a simple dashboard for students and instructors that shows weekly trends, uncertainty, and data coverage, while avoiding any judgmental labels. (iii) Offer APIs and software development kits so that learning platforms and third-party tools can read the indicators in a consistent format. (iv) Add configurable policies for when to process locally and when to sync to a secure server, and include offline modes for low-connectivity contexts. (v) Design user flows that let participants opt in, review what is collected, and pause or resume sensing. (vi) Add alerting and summaries that are rate-limited and respectful of attention. (vii) Evaluate the tools in pilot deployments with usability testing, battery impact checks, and privacy reviews. Use these results to improve defaults, refine prompts, and tune thresholds for stable operation in day-to-day use [303].

7.3.3 Advanced Modeling Towards Real-Time Estimation and Prediction

Build on concurrent estimation, refine the learning-motivation model for effective real-time estimation and prediction. (i) Extend LLM-augmented concurrent estimation to an online setting by enabling real-time

data acquisition and streaming ingestion with edge computing [304, 305]; deploy LLM-augmented intelligent agents to orchestrate sensing, context reasoning, and estimation [306–308]. (ii) Design human-interactive, user-friendly interfaces to improve students' acceptance and trust in data collection and feedback [309]. Ultimately deliver accurate individual-level real-time psychometric assessment of learning motivation. (iii) In practice, define clear latency budgets (e.g., seconds to minutes) for feature extraction and estimation. Use sliding windows and small updates so that estimates reflect the most recent behavior while maintaining stability. (iv) Compare different LLM families and sizes under streaming prompts, and test light-weight adaptation methods that can run near the device when possible. Monitor calibration and drift over time, and add simple guards that detect missing data or irregular behavior windows. (v) Introduce agent routines that decide when to sense, when to summarize, and when to request new context, while honoring privacy and power limits [306, 307]. (vi) Report both the estimate and a short rationale that explains the main signals used, written in plain language. For prediction, evaluate next-week and short-horizon forecasts with rolling-origin tests. (vii) Assess the full pipeline for cost, latency, robustness, and user impact, and include human review where decisions could affect support or interventions [304, 305, 308, 309].

7.3.4 Personalization with Learning Analytics, Adaptive Feedback and Intervention

Individual-level, personally tailored motivation assessment should be integrated into learning analytics platforms. Whereas most existing work tracks motivation at the group level over time, future research should pursue targeted, individual-level sensing and assessment [309–311]. (i) Embedding our hybrid approach into learning analytics platforms would enable scalable assessment and, when combined with real-time assessment, support timely feedback and effective personalized interventions. (ii) To achieve this, connect the estimation outputs to the platform's user model so that feedback and recommendations reflect each student's recent state and patterns. Provide small, actionable tips that respect workload and context, such as scheduling a break, planning a focused study block, or reducing digital interruptions. (iii) Allow students to set preferences for feedback frequency, channels, and tone, and offer a clear "quiet mode." (iv) For instructors, provide cohort summaries with privacy-preserving aggregation and simple indicators of who might need support, without exposing raw signals. (v) Track outcomes like engagement, completion, and well-being, and test whether adaptive feedback improves these outcomes relative to standard practice. (vi) Ensure that personalization includes fairness checks so that support is equitable across subgroups and does not amplify existing gaps [309–311].

7.3.5 Ethical and Responsible Deployment

Future research should develop and validate an ethics-and-governance framework for educational settings to enable sensing-based continuous assessment of learning motivation, grounded in purpose limitation, data minimization, and transparency [312–314]. The framework should be clear, simple to follow, and

sustainable across sites. It should state what the system does, what it does not do, and how people can exercise their rights. It should also name responsible roles, describe oversight routines, and define how to review risks before, during, and after deployment. All parts of the pipeline—sensing, preprocessing, modeling, and reporting—should follow the same principles and be auditable. The goal is a system that is useful for learning while protecting students, instructors, and institutions [312–314].

- *Purpose Limitation & Accountability:* Future work will establish and validate an ethics-and-governance framework that strictly limits use to learning-motivation assessment and research improvement. Scope will be defined via pre-registration and an approved use whitelist, with a prohibition list for secondary uses (e.g., grading, discipline, or commercial profiling) and a clear chain of responsibility and approvals. Make the approved scope public in project documents and participant materials. Define who can request changes, who can approve them, and how changes are logged. Require review for any new use, even if it seems minor, and document the decision and rationale. Create simple reporting lines so that concerns can be raised and handled quickly. Train all staff who touch the data or system, and confirm understanding with brief checklists. These steps make the purpose concrete, traceable, and enforceable across the project lifecycle.
- *Data Minimization & Privacy-by-Design:* Adopt edge computing and local preprocessing so that data are de-identified before upload, prefer feature-level over raw data collection, enforce minimal data retention and role-based access, and avoid highly sensitive sensing (e.g., audio or precise location) except where it is clearly justified and explicitly consented to. Specify secure handling for missingness in advance in order to reduce privacy risk while maintaining validity, for example by deciding when to treat missing data as true absence and when to exclude windows entirely. Collect only the signals needed for the stated aims and explain, in study materials and documentation, why each signal is necessary and how it will be used. Hash or tokenize identifiers, drop precise timestamps when they are not needed for analysis, and aggregate data to sufficiently coarse time windows so that individual routines are less identifiable. Use short retention periods by default and keep longer retention only when there is clear scientific or operational justification and when additional approval has been obtained. Apply strict access rules so that each role only sees the minimum subset of data it needs to perform its tasks, and regularly review these rules for appropriateness. Document in detail how missing data are treated and ensure that any imputation strategies or default values do not inadvertently leak private information or create unfair outcomes for particular subgroups. Regularly test the pipeline for re-identification risk, monitor for unintended information leakage, and be prepared to reduce granularity or further aggregate features when required in order to keep privacy risks within acceptable bounds.
- *Transparency, Oversight & Continuous Evaluation:* Provide layered plain-language consent and data-flow visualizations; publish feature sheets that state purpose, limits, error, and bias risks; and enable participant rights (view/correct/delete/withdraw) with human review. Institute independent

oversight (data steward/ethics board), logging and audit trails, periodic fairness/bias and cross-cultural validity checks, and safe feedback policies (thresholds, rate limits, quiet periods). Produce reusable templates (consent forms, data dictionaries, risk checklists) for multi-site sustainability. Offer a short summary and a longer guide so that participants can choose the level of detail they want. Give access portals where participants can see their data, request corrections, or withdraw, and ensure that requests are handled by trained humans. Maintain audit logs for key actions, and review them on a regular schedule. Run fairness checks by subgroup and site, and share high-level results with stakeholders. Define safe defaults for notifications and feedback so that the system does not overwhelm or pressure students. Keep a library of reusable documents and code snippets that help new sites adopt the same high standards [312–314].

Looking ahead, a natural extension of this ethics and governance agenda is to move from an interdisciplinary design that mainly involves students, psychologists, and machine learning researchers toward a more explicitly transdisciplinary process that brings a wider range of perspectives into the core of the work. Future research should therefore include institutional actors such as Institutional Review Boards, university leadership, IT and data protection offices, and, where relevant, governmental or regulatory boards as active partners in the design, implementation, validation, and deployment of sensing based motivation assessment, rather than consulting them only at late approval stages. Involving these stakeholders from the outset would help ensure that the framework is not only theoretically grounded and technically robust but also institutionally legitimate, practically sustainable, and aligned with broader educational and regulatory priorities, because key decisions about acceptable risk, communication, consent, and long term maintenance would be shared and jointly negotiated. In addition, such a transdisciplinary process would make it easier to translate high level ethical principles into concrete procedures, to anticipate institutional constraints, and to design governance mechanisms that can be adopted, audited, and iterated within real higher education systems.

Taken together, these directions outline a pathway from research prototypes to deployable, learner centred systems that can operate at scale in authentic educational environments. The first strand is to embed individualized, sensing based motivation assessment within learning analytics in order to enable targeted, timely feedback and intervention that respond to students' changing motivational states rather than relying only on infrequent, aggregate snapshots. The second, equally important strand is to pair such integration with a rigorous ethics and governance framework that is grounded in purpose limitation, data minimization, transparency, and continuous oversight so that the technical capabilities are always bounded by clear rules and safeguards. This agenda aims to deliver scalable support while safeguarding rights and trust for students, instructors, and institutions, and to show that continuous assessment can be both useful and responsible when it is carefully designed.

A consolidated summary of objectives, design choices, safeguards, and evaluation checkpoints is provided in Table 17 below, which can serve as a reusable, at-a-glance reference for researchers and practitioners

when planning future implementations and for coordinating consistent cross-site adaptations and replications.

Table 17: Future Work: Directions, Overviews, and Potential Applications

Direction	Overview	Potential Applications
Cross-cultural validation	Conduct multi-site, multilingual, and multicultural studies that compare and replicate findings, test measurement invariance, and refine and calibrate smartphone-sensed behavioral indicators.	• Cross-institution and cross-country populations • Multilingual scales and behavior dictionaries • Cross-cultural comparison reports/-dashboards • Baselines for policy/course evaluation
Automated tools & applications	Develop SDKs, mobile apps, and lightweight smart devices that use the refined behavioral indicators, with user-consented, battery-aware edge preprocessing and batch uploads scheduled during Wi-Fi/charging windows.	• Student/teacher motivation-assessing apps • Learning analytics platform plugins • Classroom companion tools
Advanced modeling for real-time estimation and prediction	Enable online concurrent estimation with streaming ingestion and LLM-driven agentic AI to orchestrate sensing and context reasoning; add calibration and uncertainty with abstention rules, and audit prompt robustness and attribution stability.	• Real-time estimation APIs • Live learning-context dashboards • Learning Management System (LMS) feedback interfaces
Personalized analytics & interventions	Integrate person-level estimates into learning analytics, enforce rate limits and quiet periods with human-in-the-loop safeguards, and rigorously evaluate intervention efficacy and safety.	• Student personalized learning analytics and intervention platforms • Teacher/advisor class & individual dashboards/interfaces • Referral/appointment scheduling
Ethical & governance	Establish governance based on purpose limitation and data minimization, with clear consent management, audit logs, fairness/bias checks, the right to withdraw and to request data deletion, and reusable templates for multi-site use.	• Governance toolkit (consent forms, model cards, risk checklists) • Compliance auditing system • Participant self-service portal • Reusable cross-site workflows

7.4 Closing Conclusion

In this dissertation, we proposed a hybrid psychometric approach that integrates passive smartphone sensing with SDT-grounded self-reports to enable the continuous assessment of learning motivation in everyday contexts. The approach combines low-burden behavioral traces with theory-based measures so that behavioral evidence and psychological constructs are used together rather than in isolation. We

operationalized the approach through a coherent five-layer framework and a progressive three-stage study design (association → optimization → estimation). Across the program of studies, we demonstrated an interpretable mapping between behavior and learning motivation, ecologically valid and time-resolved evidence collected in authentic settings, optimization of a compact set of behavioral indicators aligned with learning motivation, and LLM-augmented concurrent estimation at the individual level.

Our work directly addresses the central gap of traditional, episodic self-report methods, which cannot support continuous or real-time assessment and often place additional burden on students. In contrast, the proposed pipeline offers concrete guidance for study design, a transferable set of learning-motivation–aligned behavioral indicators, an operationalizable framework that can be implemented in practice, and privacy-preserving procedures suitable for unobtrusive use in higher-education environments. These elements make it feasible to assess motivation continuously while maintaining interpretability and safeguarding participants.

Looking forward, the same paradigm can guide rigorous cross-cultural validation by applying common protocols across institutions and regions, and it can support the development of intelligent, real-time systems within learning-analytics platforms. With careful governance, such systems can operate on or near the device, integrate ongoing sensing with estimation, and provide individual-level outputs. They can also extend from estimation to short-horizon prediction and enable timely, personalized interventions that are transparent, respectful, and effective, thereby advancing both psychological measurement and educational technology in tandem.

Bibliography

[1] E. L. Deci and R. M. Ryan. “Self-determination theory: A macrotheory of human motivation, development, and health.” In: *Canadian psychology/Psychologie canadienne* 49.3 (2008), p. 182 (cit. on pp. 1, 2).

[2] R. M. Ryan and E. L. Deci. “Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions”. In: *Contemporary Educational Psychology* 61 (2020), p. 101860. issn: 0361-476X. doi: https://doi.org/10.1016/j.cedpsych.2020.101860. url: https://www.sciencedirect.com/science/article/pii/S0361476X20300254 (cit. on pp. 1–3, 37, 38, 40, 41).

[3] T. K. Chiu. “Applying the self-determination theory (SDT) to explain student engagement in online learning during the COVID-19 pandemic”. In: *Journal of research on Technology in Education* 54.sup1 (2022), S14–S30 (cit. on p. 1).

[4] J. Reeve and H. Jang. “What teachers say and do to support students’ autonomy during a learning activity”. In: *Journal of Educational Psychology* 98.1 (2006), pp. 209–218. doi: 10.1037/0022-0663.98.1.209 (cit. on p. 1).

[5] A. E. Black and E. L. Deci. “The effects of instructors’ autonomy support and students’ autonomous motivation on learning organic chemistry: A self-determination theory perspective”. In: *Science Education* 84.6 (2000), pp. 740–756. doi: 10.1002/1098-237X(200011)84:6<740::AID-SCE4>3.0.CO;2-3 (cit. on p. 1).

[6] R. M. Ryan and E. L. Deci. “Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being”. In: *American Psychologist* 55.1 (2000), pp. 68–78. doi: 10.1037/0003-066X.55.1.68 (cit. on pp. 2, 37, 38, 40).

[7] E. L. Deci and R. M. Ryan. “The What and Why of Goal Pursuits: Human Needs and the Self-Determination of Behavior”. In: *Psychological Inquiry* 11.4 (2000), pp. 227–268. doi: 10.1207/S15327965PLI1104_01. url: https://www.tandfonline.com/doi/abs/10.1207/S15327965PLI1104_01 (cit. on p. 2).

[8] R. M. Ryan, C. S. Rigby, and A. K. Przybylski. “The motivational pull of video games: A self-determination theory approach”. In: *Motivation and Emotion* 30.4 (2006), pp. 347–363. doi: 10.1007/s11031-006-9051-8 (cit. on p. 2).

[9] R. M. Ryan and E. L. Deci. *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. New York, NY: The Guilford Press, 2017. isbn: 9781462528769 (cit. on pp. 2, 37).

[10] R. M. Ryan and E. L. Deci. “Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions”. In: *Contemporary Educational Psychology* 25.1 (2000), pp. 54–67. doi: 10.1006/ceps.1999.1020 (cit. on p. 2).

[11] R. M. Ryan et al. “Building a science of motivated persons: Self-determination theory’s empirical approach to human experience and the regulation of behavior.” In: *Motivation science* 7.2 (2021), p. 97 (cit. on p. 2).

[12] R. M. Ryan and E. L. Deci. “Self-determination theory”. In: *Encyclopedia of quality of life and well-being research*. Springer, 2024, pp. 6229–6235 (cit. on p. 2).

[13] R. M. Ryan. *Self-Regulation Questionnaire (SRQ-A)*. https://selfdeterminationtheory.org/self-regulation-questionnaire-srq-a/. Accessed 2025-10-16 (cit. on pp. 2, 4, 19, 54, 61).

[14] B. Chen et al. *Basic Psychological Need Satisfaction and Frustration Scale (BPNSFS)*. https://selfdeterminationtheory.org/basic-psychological-need-satisfaction-and-frustration-scales/. Accessed 2025-10-16 (cit. on pp. 2, 4).

[15] R. M. Ryan. *Intrinsic Motivation Inventory (IMI)*. https://selfdeterminationtheory.org/intrinsic-motivation-inventory/. Accessed 2025-10-16 (cit. on pp. 2, 4, 54, 61).

[16] R. J. Vallerand et al. “The Academic Motivation Scale: A Measure of Intrinsic, Extrinsic, and Amotivation in Education”. In: *Educational and Psychological Measurement* 52.4 (1992), pp. 1003–1017. doi: 10.1177/0013164492052004025 (cit. on pp. 2, 4, 54, 61).

[17] F. Guay, R. J. Vallerand, and C. Blanchard. “On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS)”. In: *Motivation and Emotion* 24.3 (2000), pp. 175–213. doi: 10.1023/A:1005614228250 (cit. on p. 2).

[18] M. Csikszentmihalyi and R. Larson. “Validity and reliability of the experience-sampling method”. In: *The Journal of nervous and mental disease* 175.9 (1987), pp. 526–536 (cit. on p. 2).

[19] J. M. Hektner, J. A. Schmidt, and M. Csikszentmihalyi. *Experience sampling method: Measuring the quality of everyday life*. Sage, 2007 (cit. on p. 2).

[20] K. A. Renninger and S. E. Hidi. *The Cambridge handbook of motivation and learning*. Cambridge University Press, 2019 (cit. on pp. 2, 4).

[21] S. E. Kassab et al. “Measurement of student engagement in health professions education: a review of literature”. In: *BMC Medical Education* 23.1 (2023), p. 354 (cit. on p. 2).

[22] L. P. De Vries, B. M. Baselmans, and M. Bartels. “Smartphone-based ecological momentary assessment of well-being: A systematic review and recommendations for future studies”. In: *Journal of Happiness Studies* 22.5 (2021), pp. 2361–2408 (cit. on p. 2).

[23] P. Srinivas et al. “Context-sensitive ecological momentary assessment: application of user-centered design for improving user satisfaction and engagement during self-report”. In: *JMIR mHealth and uHealth* 7.4 (2019), e10894 (cit. on p. 2).

[24] N. van Berkel, D. Ferreira, and V. Kostakos. “The Experience Sampling Method on Mobile Devices”. In: *ACM Comput. Surv.* 50.6 (2017-12). issn: 0360-0300. doi: 10.1145/3123988. url: https://doi.org/10.1145/3123988 (cit. on p. 3).

[25] G. Eisele et al. “The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population”. In: *Assessment* 29.2 (2022), pp. 136–151 (cit. on p. 3).

[26] A. A. Stone, S. Schneider, and J. M. Smyth. “Evaluation of pressing issues in ecological momentary assessment”. In: *Annual Review of Clinical Psychology* 19.1 (2023), pp. 107–131 (cit. on p. 3).

[27] M. S. Thomas, D. Ansari, and V. C. Knowland. “Annual research review: Educational neuroscience: Progress and prospects”. In: *Journal of child psychology and psychiatry* 60.4 (2019), pp. 477–492 (cit. on p. 3).

[28] A. Apicella et al. “EEG-based measurement system for monitoring student engagement in learning 4.0”. In: *Scientific Reports* 12.1 (2022), p. 5857 (cit. on p. 3).

[29] T. W. Janssen et al. “Opportunities and limitations of mobile neuroimaging technologies in educational neuroscience”. In: *Mind, Brain, and Education* 15.4 (2021), pp. 354–370 (cit. on p. 3).

[30] T. R. Insel. “Digital Phenotyping: Technology for a New Science of Behavior”. In: *JAMA* 318.13 (2017), pp. 1215–1216. doi: 10.1001/jama.2017.11295 (cit. on p. 3).

[31] G. M. Harari, S. D. Gosling, and A. T. Campbell. “Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges”. In: *Perspectives on Psychological Science* 11.6 (2016), pp. 838–854. doi: 10.1177/1745691616650285 (cit. on p. 3).

[32] G. M. Harari et al. “Smartphone sensing methods for studying behavior in everyday life”. In: *Current Opinion in Behavioral Sciences* 18 (2017), pp. 83–90. doi: 10.1016/j.cobeha.2017.07.018 (cit. on p. 3).

[33] V. Scherrer and F. Preckel. “Development of motivational variables and self-esteem during the school career: A meta-analysis of longitudinal studies”. In: *Review of educational research* 89.2 (2019), pp. 211–258 (cit. on p. 3).

[34] F. Guay et al. “Trajectories of self-determined motivation during the secondary school: A growth mixture analysis.” In: *Journal of Educational Psychology* 113.2 (2021), p. 390 (cit. on pp. 3, 56).

[35] D. C. Mohr, M. Zhang, and S. M. Schueller. “Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning”. In: *Annual Review of Clinical Psychology* 13 (2017), pp. 23–47. doi: 10.1146/annurev-clinpsy-032816-044949. url: https://www.annualreviews.org/content/journals/10.1146/annurev-clinpsy-032816-044949 (cit. on p. 3).

[36] R. Wang et al. “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*. 2016, pp. 886–897. doi: 10.1145/2971648.2971740 (cit. on p. 3).

[37] J. Zhou et al. “Predicting Psychotic Relapse in Schizophrenia With Mobile Sensor Data: Routine Cluster Analysis”. In: *JMIR Mhealth Uhealth* 10.4 (2022), e31006. issn: 2291-5222. doi: 10.2196/31006. url: https://mhealth.jmir.org/2022/4/e31006 (cit. on p. 3).

[38] W. Wang et al. “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*. 2020, pp. 1–15 (cit. on p. 3).

[39] R. Wang et al. “SmartGPA: How Smartphones Can Assess and Predict Academic Performance of College Students”. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2015, pp. 295–306. doi: 10.1145/2750858.2804251 (cit. on pp. 3, 5, 43, 55, 56, 61).

[40] R. Wang et al. “Tracking depression dynamics in college students using mobile phone and wearable sensing”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.1 (2018), pp. 1–26 (cit. on pp. 3, 55).

[41] S. A. Nikou and A. A. Economides. “Mobile-Based Assessment: Integrating acceptance and motivational factors into a combined model of Self-Determination Theory and Technology Acceptance”. In: *Computers in Human Behavior* 68 (2017), pp. 83–95 (cit. on pp. 3, 6).

[42] J. L. Howard et al. “Student motivation and associated outcomes: A meta-analysis from self-determination theory”. In: *Perspectives on Psychological Science* 16.6 (2021), pp. 1300–1323 (cit. on p. 3).

[43] P. R. Pintrich et al. “A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).” In: (1991) (cit. on pp. 4, 54, 61).

[44] C. Brandmiller, K. Schnitzler, and H. Dumont. “Teacher perceptions of student motivation and engagement: Longitudinal associations with student outcomes”. In: *European Journal of Psychology of Education* 39.2 (2024), pp. 1397–1420 (cit. on p. 4).

[45] R. M. Ryan and E. L. Deci. *Self-Regulation Questionnaires (complete set)*. https://selfdeterminationtheo org/questionnaires/. Accessed 2025-10-16 (cit. on pp. 4, 19, 54, 61).

[46] M. Hall et al. "A systematic review of momentary assessment designs for mood and anxiety symptoms". In: *Frontiers in Psychology* 12 (2021), p. 642044 (cit. on p. 5).

[47] R. Tourangeau, L. J. Rips, and K. Rasinski. "The psychology of survey response". In: (2000) (cit. on p. 5).

[48] N. M. Bradburn, L. J. Rips, and S. K. Shevell. "Answering autobiographical questions: The impact of memory and inference on surveys". In: *Science* 236.4798 (1987), pp. 157–161 (cit. on p. 5).

[49] P. M. Podsakoff et al. "Common method biases in behavioral research: a critical review of the literature and recommended remedies." In: *Journal of applied psychology* 88.5 (2003), p. 879 (cit. on pp. 5, 54).

[50] A. Dodaj. "Social desirability and self-reports: Testing a content and response-style model of socially desirable responding". In: *Europe's Journal of Psychology* 8.4 (2012), pp. 651–666 (cit. on p. 5).

[51] J. A. Krosnick. "The Palgrave handbook of survey research". In: *Cham: Palgrave Macmillan* (2018) (cit. on p. 5).

[52] N. Schwarz. "Cognitive aspects of survey methodology". In: *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 21.2 (2007), pp. 277–287 (cit. on p. 5).

[53] R. E. McCord and H. M. Matusovich. "Naturalistic observations of metacognition in engineering: Using observational methods to study metacognitive engagement in engineering". In: *Journal of Engineering Education* 108.4 (2019), pp. 481–502 (cit. on p. 5).

[54] D. C. Mohr, M. Zhang, and S. M. Schueller. "Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning". In: *Annual Review of Clinical Psychology* 13 (2017), pp. 23–47. doi: 10.1146/annurev-clinpsy-032816-044949 (cit. on p. 5).

[55] G. M. Harari et al. "Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges". In: *Perspectives on Psychological Science* 11.6 (2016), pp. 838–854 (cit. on pp. 5, 43–45).

[56] B. Cao et al. "DeepMood: Modeling Mobile Phone Typing Dynamics for Mood Detection". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017, pp. 747–755. doi: 10.1145/3097983.3098086 (cit. on p. 5).

[57] E. Birtwistle et al. "Mobile sensing in psychological and educational research: Examples from two application fields". In: *International Journal of Testing* 22.3-4 (2022), pp. 264–288. doi: 10.1080/15305058.2022.2036160 (cit. on p. 5).

[58] J. Melcher, R. Hays, and J. Torous. "Digital Phenotyping for Mental Health of College Students: A Clinical Review". In: *Evidence-Based Mental Health* 23.4 (2020), pp. 161–166. doi: 10.1136/ebmental-2020-300180 (cit. on p. 5).

[59] R. Wang et al. "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*. UbiComp '14. New York, NY, USA: Association for Computing Machinery, 2014, pp. 3–14. doi: 10.1145/2632048.2632054 (cit. on pp. 5, 43, 44, 50, 55, 56, 61).

[60] R. Wang et al. "Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.1 (2018). doi: 10.1145/3191775 (cit. on p. 5).

[61] W. Wang et al. "First-Gen Lens: Assessing Mental Health of First-Generation Students across Their First Year at College Using Mobile Sensing". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6.2 (2022-07). doi: 10.1145/3543194. url: https://doi.org/10.1145/3543194 (cit. on p. 5).

[62] J. F. Huckins et al. "Fusing Mobile Phone Sensing and Brain Imaging to Assess Depression in College Students". In: *Frontiers in Neuroscience* 13 (2019), p. 248. doi: 10.3389/fnins.2019.00248 (cit. on p. 5).

[63] L. Canzian and M. Musolesi. "Trajectories of Depression: Unobtrusive Monitoring of Depressive States by means of Smartphone Mobility Traces Analysis". In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2015, pp. 1293–1304. doi: 10.1145/2750858.2805845 (cit. on p. 5).

[64] H. Zhang et al. "Multi-modal interactive fusion method for detecting teenagers' psychological stress". In: *Journal of Biomedical Informatics* 106 (2020), p. 103427. issn: 1532-0464. doi: https://doi.org/10.1016/j.jbi.2020.103427. url: https://www.sciencedirect.com/science/article/pii/S1532046420300551 (cit. on p. 5).

[65] L. M. Jeno, J.-A. Grytnes, and V. Vandvik. "The effect of a mobile-application tool on biology students' motivation and achievement in species identification: A Self-Determination Theory perspective". In: *Computers & Education* 107 (2017), pp. 1–12 (cit. on p. 6).

[66] J. Dietrich, F. Schmiedek, and J. Moeller. *Academic motivation and emotions are experienced in learning situations, so let's study them. Introduction to the special issue*. 2022 (cit. on p. 6).

[67] K. Liang and S. L. Zeger. "Longitudinal data analysis using generalized linear models". In: *Biometrika* 73.1 (1986), pp. 13–22. doi: 10.1093/biomet/73.1.13 (cit. on pp. 28, 87).

[68] J. W. Hardin and J. M. Hilbe. *Generalized Estimating Equations*. Boca Raton, FL: Chapman and Hall/CRC, 2002. doi: 10.1201/9781420035285 (cit. on pp. 28, 87).

[69] Y. Benjamini, A. M. Krieger, and D. Yekutieli. “Adaptive linear step-up procedures that control the false discovery rate”. In: *Biometrika* 93.3 (2006), pp. 491–507 (cit. on pp. 28, 87).

[70] E. L. Deci and R. M. Ryan. “Conceptualizations of Intrinsic Motivation and Self-Determination”. In: *Intrinsic Motivation and Self-Determination in Human Behavior*. Boston, MA: Springer US, 1985, pp. 11–40. isbn: 978-1-4899-2271-7. doi: 10.1007/978-1-4899-2271-7_2. url: https://doi.org/10.1007/978-1-4899-2271-7_2 (cit. on p. 37).

[71] R. M. Ryan and J. P. Connell. “Perceived locus of causality and internalization: examining reasons for acting in two domains.” In: *Journal of personality and social psychology* 57.5 (1989), p. 749 (cit. on pp. 37, 40).

[72] R. M. Ryan and E. L. Deci. “Intrinsic and extrinsic motivations: Classic definitions and new directions”. In: *Contemporary educational psychology* 25.1 (2000), pp. 54–67 (cit. on p. 37).

[73] S. I. Di Domenico and R. M. Ryan. “The emerging neuroscience of intrinsic motivation: A new frontier in self-determination research”. In: *Frontiers in human neuroscience* 11 (2017), p. 247005 (cit. on pp. 37, 38).

[74] R. J. Vallerand. “Toward a hierarchical model of intrinsic and extrinsic motivation”. In: *Advances in experimental social psychology*. Vol. 29. Elsevier, 1997, pp. 271–360 (cit. on pp. 37, 40).

[75] M. Vansteenkiste et al. “Motivating learning, performance, and persistence: the synergistic effects of intrinsic goal contents and autonomy-supportive contexts.” In: *Journal of personality and social psychology* 87.2 (2004), p. 246 (cit. on pp. 37, 39, 41).

[76] E. L. Deci et al. “Motivation and education: The self-determination perspective”. In: *Educational psychologist* 26.3-4 (1991), pp. 325–346 (cit. on p. 37).

[77] J. Reeve. “Why teachers adopt a controlling motivating style toward students and how they can become more autonomy supportive”. In: *Educational psychologist* 44.3 (2009), pp. 159–175 (cit. on p. 37).

[78] H. A. McGregor et al. “A 2× 2 achievement goal framework”. In: *Journal of Personality and Social Psychology* 80.3 (2001), pp. 501–519 (cit. on pp. 37, 38).

[79] M. Vansteenkiste, C. P. Niemiec, and B. Soenens. “The development of the five mini-theories of self-determination theory: An historical overview, emerging trends, and future directions”. In: *The decade ahead: Theoretical perspectives on motivation and achievement* (2010), pp. 105–165 (cit. on p. 38).

[80] W. S. Grolnick and R. M. Ryan. “Autonomy in children’s learning: an experimental and individual difference investigation.” In: *Journal of personality and social psychology* 52.5 (1987), p. 890 (cit. on pp. 38, 39).

[81] E. L. Deci, R. Koestner, and R. M. Ryan. “A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation.” In: *Psychological bulletin* 125.6 (1999), p. 627 (cit. on p. 38).

[82] B. Soenens and M. Vansteenkiste. “A theoretical upgrade of the concept of parental psychological control: Proposing new insights on the basis of self-determination theory”. In: *Developmental review* 30.1 (2010), pp. 74–99 (cit. on p. 38).

[83] A. Assor, G. Roth, and E. L. Deci. “The emotional costs of parents’ conditional regard: A Self-Determination Theory analysis”. In: *Journal of personality* 72.1 (2004), pp. 47–88 (cit. on p. 38).

[84] N. Weinstein, C. R. DeHaan, and R. M. Ryan. “Attributing autonomous versus introjected motivation to helpers and the recipient experience: Effects on gratitude, attitudes, and well-being”. In: *Motivation and Emotion* 34.4 (2010), pp. 418–431 (cit. on p. 38).

[85] A. H. Olafsen and E. L. Deci. “Self-determination theory and its relation to organizations”. In: *Oxford research encyclopedia of psychology*. 2020 (cit. on p. 38).

[86] M. Vansteenkiste, W. Lens, and E. L. Deci. “Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation”. In: *Educational psychologist* 41.1 (2006), pp. 19–31 (cit. on p. 38).

[87] M. Standage, J. L. Duda, and N. Ntoumanis. “A test of self-determination theory in school physical education”. In: *British journal of educational psychology* 75.3 (2005), pp. 411–433 (cit. on p. 38).

[88] J. De Meyer et al. “The different faces of controlling teaching: implications of a distinction between externally and internally controlling teaching for students’ motivation in physical education”. In: *Physical Education and Sport Pedagogy* 21.6 (2016), pp. 632–652 (cit. on p. 38).

[89] E. A. Locke and G. P. Latham. “Building a practically useful theory of goal setting and task motivation: A 35-year odyssey.” In: *American psychologist* 57.9 (2002), p. 705 (cit. on p. 38).

[90] R. M. Ryan and E. L. Deci. “Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.” In: *American psychologist* 55.1 (2000), p. 68 (cit. on p. 38).

[91] R. M. Ryan and E. L. Deci. “Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions”. In: *Contemporary educational psychology* 61 (2020), p. 101860 (cit. on p. 38).

[92] R. M. Ryan and E. L. Deci. “Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness”. In: *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. New York, NY: Guilford Press, 2017. isbn: 9781462528769 (cit. on p. 38).

[93] D. Oyserman. “Identity-based motivation”. In: *Emerging trends in the social and behavioral sciences* 38 (2015), pp. 1–11 (cit. on pp. 38, 41).

[94] M. Csikszentmihalyi and J. Nakamura. “The Dynamics of Intrinsic Motivation: A Study of Adolescents”. In: *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*. Dordrecht: Springer Netherlands, 2014, pp. 175–197. isbn: 978-94-017-9088-8. doi: 10.1007/978-94-017-9088-8_12. url: https://doi.org/10.1007/978-94-017-9088-8_12 (cit. on p. 39).

[95] J. Litman. “Curiosity and the pleasures of learning: Wanting and liking new information”. In: *Cognition and emotion* 19.6 (2005), pp. 793–814 (cit. on p. 39).

[96] E. L. Deci et al. “Facilitating internalization: The self-determination theory perspective”. In: *Journal of personality* 62.1 (1994), pp. 119–142 (cit. on pp. 39, 40).

[97] C. P. Niemiec and R. M. Ryan. “Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice”. In: *Theory and research in Education* 7.2 (2009), pp. 133–144 (cit. on pp. 39, 40).

[98] E. L. Deci and R. M. Ryan. “The general causality orientations scale: Self-determination in personality”. In: *Journal of research in personality* 19.2 (1985), pp. 109–134 (cit. on p. 40).

[99] M. S. Hagger and K. Hamilton. “General causality orientations in self-determination theory: Meta-analysis and test of a process model”. In: *European Journal of Personality* 35.5 (2021), pp. 710–735 (cit. on p. 40).

[100] M. Gagné. “The role of autonomy support and autonomy orientation in prosocial behavior engagement”. In: *Motivation and emotion* 27.3 (2003), pp. 199–223 (cit. on p. 40).

[101] K. M. Sheldon, R. Ryan, and H. T. Reis. “What makes for a good day? Competence and autonomy in the day and in the person”. In: *Personality and social psychology bulletin* 22.12 (1996), pp. 1270–1279 (cit. on p. 40).

[102] M. Seitl, E. Manuoglu, and A. Hrbáčková. “Daily state of motivation as the effect of appreciation in a diary study”. In: *Current Issues in Personality Psychology* 12.3 (2024), p. 185 (cit. on pp. 40, 41).

[103] K. J. Bartholomew et al. “Self-determination theory and diminished functioning: The role of interpersonal control and psychological need thwarting”. In: *Personality and social psychology bulletin* 37.11 (2011), pp. 1459–1473 (cit. on p. 40).

[104] E. A. Patall, H. Cooper, and J. C. Robinson. “The effects of choice on intrinsic motivation and related outcomes: a meta-analysis of research findings.” In: *Psychological bulletin* 134.2 (2008), p. 270 (cit. on pp. 40–42).

[105] J. Reeve and H. Jang. “What teachers say and do to support students’ autonomy during a learning activity.” In: *Journal of educational psychology* 98.1 (2006), p. 209 (cit. on pp. 40, 41).

[106] J. REEVE. “Why Teachers Adopt a Controlling Motivating Style Toward Students and How They Can Become More Autonomy Supportive”. In: *Educational Psychologist* 44.3 (2009), pp. 159–175. doi: 10.1080/00461520903028990. eprint: https://doi.org/10.1080/00461520903028990. url: https://doi.org/10.1080/00461520903028990 (cit. on p. 40).

[107] C. R. Stefanou et al. “Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership”. In: *Educational psychologist* 39.2 (2004), pp. 97–110 (cit. on p. 40).

[108] H. Jang, J. Reeve, and E. L. Deci. “Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure.” In: *Journal of educational psychology* 102.3 (2010), p. 588 (cit. on pp. 40, 41).

[109] J. Hattie and H. Timperley. “The power of feedback”. In: *Review of educational research* 77.1 (2007), pp. 81–112 (cit. on p. 40).

[110] E. L. Deci. “Effects of externally mediated rewards on intrinsic motivation.” In: *Journal of personality and Social Psychology* 18.1 (1971), p. 105 (cit. on p. 40).

[111] C. Furrer and E. Skinner. “Sense of relatedness as a factor in children’s academic engagement and performance.” In: *Journal of educational psychology* 95.1 (2003), p. 148 (cit. on p. 40).

[112] K. R. Wentzel. “Student motivation in middle school: The role of perceived pedagogical caring.” In: *Journal of educational psychology* 89.3 (1997), p. 411 (cit. on p. 40).

[113] R. F. Baumeister and M. R. Leary. “The need to belong: Desire for interpersonal attachments as a fundamental human motivation”. In: *Interpersonal development* (2017), pp. 57–89 (cit. on p. 40).

[114] R. Pekrun. “The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice”. In: *Educational psychology review* 18.4 (2006), pp. 315–341 (cit. on pp. 40, 42).

[115] R. F. Baumeister et al. “Social exclusion impairs self-regulation.” In: *Journal of personality and social psychology* 88.4 (2005), p. 589 (cit. on p. 40).

[116] J. Reeve et al. “Enhancing students’ engagement by increasing teachers’ autonomy support”. In: *Motivation and emotion* 28.2 (2004), pp. 147–169 (cit. on p. 40).

[117] E. A. Patall et al. “Student autonomy and course value: The unique and cumulative roles of various teacher practices”. In: *Motivation and Emotion* 37.1 (2013), pp. 14–32 (cit. on p. 40).

[118] H. T. Reis et al. “Daily well-being: The role of autonomy, competence, and relatedness”. In: *Relationships, well-being and behaviour*. Routledge, 2018, pp. 317–349 (cit. on pp. 41, 56).

[119] C. K. Enders and D. Tofighi. “Centering predictor variables in cross-sectional multilevel models: a new look at an old issue.” In: *Psychological methods* 12.2 (2007), p. 121 (cit. on pp. 41, 49).

[120] P. J. Curran and D. J. Bauer. “The disaggregation of within-person and between-person effects in longitudinal models of change”. In: *Annual review of psychology* 62.1 (2011), pp. 583–619 (cit. on pp. 41, 49).

[121] N. Bolger and J.-P. Laurenceau. *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford press, 2013 (cit. on pp. 41, 55).

[122] A. M. Rodríguez-Sánchez et al. “Enjoyment and absorption: An electronic diary study on daily flow patterns”. In: *Work & Stress* 25.1 (2011), pp. 75–92 (cit. on p. 41).

[123] P. Valdez. “Circadian rhythms in attention”. In: *The Yale journal of biology and medicine* 92.1 (2019), p. 81 (cit. on p. 41).

[124] E. A. Skinner, T. A. Kindermann, and C. J. Furrer. “A motivational perspective on engagement and disaffection: Conceptualization and assessment of children’s behavioral and emotional participation in academic activities in the classroom”. In: *Educational and psychological measurement* 69.3 (2009), pp. 493–525 (cit. on pp. 41, 42).

[125] E. A. Patall, H. Cooper, and S. R. Wynn. “The effectiveness and relative importance of choice in the classroom.” In: *Journal of educational psychology* 102.4 (2010), p. 896 (cit. on p. 41).

[126] C. P. Niemiec, A. Muñoz, et al. “A need-supportive intervention delivered to English language teachers in Colombia: A pilot investigation based on self-determination theory”. In: *Psychology* 10.07 (2019), p. 1025 (cit. on p. 41).

[127] J. Reeve and S. H. Cheon. “Autonomy-supportive teaching: Its malleability, benefits, and potential to improve educational practice”. In: *Educational psychologist* 56.1 (2021), pp. 54–77 (cit. on p. 41).

[128] J. Reeve and C.-M. Tseng. “Agency as a fourth aspect of students’ engagement during learning activities”. In: *Contemporary educational psychology* 36.4 (2011), pp. 257–267 (cit. on p. 41).

[129] S. A. Karabenick and M. H. Dembo. “Understanding and facilitating self-regulated help seeking”. In: *New directions for teaching and learning* 2011.126 (2011), pp. 33–43 (cit. on p. 41).

[130] C. Ames. “Classrooms: Goals, structures, and student motivation.” In: *Journal of educational psychology* 84.3 (1992), p. 261 (cit. on p. 41).

[131] R. Junco. “In-class multitasking and academic performance”. In: *Comput. Hum. Behav.* 28.6 (2012-11), pp. 2236–2243. issn: 0747-5632. doi: 10.1016/j.chb.2012.06.031. url: https://doi.org/10.1016/j.chb.2012.06.031 (cit. on p. 41).

[132] R. F. Kizilcec, C. Piech, and E. Schneider. “Deconstructing disengagement: analyzing learner subpopulations in massive open online courses”. In: *Proceedings of the third international conference on learning analytics and knowledge*. 2013, pp. 170–179 (cit. on pp. 41, 43, 55, 61).

[133] E. S. Tabanao, M. M. T. Rodrigo, and M. C. Jadud. “Predicting at-risk novice Java programmers through the analysis of online protocols”. In: *Proceedings of the seventh international workshop on Computing education research*. 2011, pp. 85–92 (cit. on p. 41).

[134] M. Leijten and L. Van Waes. “Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes”. In: *Written Communication* 30.3 (2013), pp. 358–392 (cit. on p. 41).

[135] R. Pekrun et al. “The control-value theory of achievement emotions: An integrative approach to emotions in education”. In: *Emotion in education*. Elsevier, 2007, pp. 13–36 (cit. on p. 41).

[136] B. Groot. “Choking Under Pressure: The Effect of Evaluation Threat on Academic Procrastination Behaviour”. In: () (cit. on p. 41).

[137] L. D. Rosen et al. “An empirical examination of the educational impact of text message-induced task switching in the classroom: Educational implications and strategies to enhance learning”. In: *Psicología Educativa. Revista de los Psicólogos de la Educación* 17.2 (2011), pp. 163–177 (cit. on p. 41).

[138] W. Kool et al. “Decision making and the avoidance of cognitive demand.” In: *Journal of experimental psychology: general* 139.4 (2010), p. 665 (cit. on p. 41).

[139] P. Filippello et al. “School refusal and absenteeism: Perception of teacher behaviors, psychological basic needs, and academic achievement”. In: *Frontiers in psychology* 10 (2019), p. 1471 (cit. on p. 41).

[140] F. E. V. Castro, J. Leinonen, and A. Hellas. “Experiences with and lessons learned on deadlines and submission behavior”. In: *Proceedings of the 22nd Koli Calling International Conference on Computing Education Research*. 2022, pp. 1–13 (cit. on pp. 41, 43).

[141] V. Sher, M. Hatala, and D. Gašević. “When do learners study?: an analysis of the time-of-day and weekday-weekend usage patterns of learning management systems from mobile and computers in blended learning”. In: *Journal of Learning Analytics* 9.2 (2022), pp. 1–23 (cit. on p. 41).

[142] M. Saqr, J. Nouri, and U. Fors. “Time to focus on the temporal dimension of learning: A learning analytics study of the temporal patterns of students’ interactions and self-regulation”. In: *International Journal of Technology Enhanced Learning* 11.4 (2019), pp. 398–412 (cit. on p. 41).

[143] T.-T. Goh. “Learning management system log analytics: the role of persistence and consistency of engagement behaviour on academic success”. In: *Journal of Computers in Education* (2025), pp. 1–24 (cit. on p. 41).

[144] V. Sher, M. Hatala, and D. Gašević. “Analyzing the consistency in within-activity learning patterns in blended learning”. In: *Proceedings of the tenth international conference on learning analytics & knowledge*. 2020, pp. 1–10 (cit. on p. 41).

[145] S. R. Alghamdi et al. “Analytics of Temporal Patterns of Self-regulated Learners: A Time Series Approach”. In: *Proceedings of the 15th International Learning Analytics and Knowledge Conference*. LAK ’25. New York, NY, USA: Association for Computing Machinery, 2025, pp. 283–292. isbn: 9798400707018. doi: 10.1145/3706468.3706504. url: https://doi.org/10.1145/3706468.3706504 (cit. on p. 41).

[146] T. Dvorak and M. Jia. “Do the Timeliness, Regularity, and Intensity of Online Work Habits Predict Academic Performance?” In: *Journal of Learning Analytics* 3 (2016-12), pp. 318–330. doi: 10.18608/jla.2016.33.15 (cit. on p. 41).

[147] C. Smith. “The Impact of homework deadline times on college student performance and Stress: A Quasi-Experiment in Business Statistics”. In: *Journal of Statistics and Data Science Education* (2025), pp. 1–10 (cit. on p. 41).

[148] M. Credé, S. G. Roch, and U. M. Kieszczynka. “Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics”. In: *Review of Educational Research* 80.2 (2010), pp. 272–295 (cit. on p. 41).

[149] Z. Zhao, P. Deng, and J. Zhou. “Identifying longitudinal attendance patterns through student subpopulation distribution comparison”. In: *Proceedings of the 15th International Conference on Educational Data Mining*. 2022, p. 640 (cit. on p. 41).

[150] A. Delavande et al. *Timetables, Attendance and Academic Achievement in Higher Education*. Tech. rep. IZA Discussion Papers, 2025 (cit. on p. 41).

[151] S. Larabi-Marie-Sainte et al. “The impact of timetable on student’s absences and performance”. In: *Plos one* 16.6 (2021), e0253256 (cit. on p. 41).

[152] A. Sheshadri et al. “Predicting student performance based on online study habits: a study of blended courses”. In: *arXiv preprint arXiv:1904.07331* (2019) (cit. on p. 41).

[153] W. Zhou, E. Simpson, and D. P. Domizi. “Google Docs in an out-of-class collaborative writing activity.” In: *International Journal of Teaching and Learning in Higher Education* 24.3 (2012), pp. 359–375 (cit. on p. 41).

[154] D. J. Nicol and D. Macfarlane-Dick. “Formative assessment and self-regulated learning: A model and seven principles of good feedback practice”. In: *Studies in higher education* 31.2 (2006), pp. 199–218 (cit. on p. 41).

[155] D. W. Johnson, R. T. Johnson, and K. A. Smith. “Cooperative learning returns to college what evidence is there that it works?” In: *Change: the magazine of higher learning* 30.4 (1998), pp. 26–35 (cit. on p. 41).

[156] S. J. Karau and K. D. Williams. “Social loafing: A meta-analytic review and theoretical integration.” In: *Journal of personality and social psychology* 65.4 (1993), p. 681 (cit. on p. 41).

[157] H. Jang, E. J. Kim, and J. Reeve. “Longitudinal test of self-determination theory’s motivation mediation model in a naturally occurring classroom context.” In: *Journal of Educational psychology* 104.4 (2012), p. 1175 (cit. on pp. 41, 54).

[158] A. Wigfield and J. S. Eccles. “Expectancy–value theory of achievement motivation”. In: *Contemporary educational psychology* 25.1 (2000), pp. 68–81 (cit. on p. 41).

[159] S. Yu et al. “Self-determined motivation to choose college majors, its antecedents, and outcomes: A cross-cultural investigation”. In: *Journal of Vocational Behavior* 108 (2018), pp. 132–150 (cit. on p. 41).

[160] S. M. Barnett and S. J. Ceci. “When and where do we apply what we learn?: A taxonomy for far transfer.” In: *Psychological bulletin* 128.4 (2002), p. 612 (cit. on p. 41).

[161] N. R. Council. *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, DC: The National Academies Press, 2000. isbn: 978-0-309-07036-2. doi: 10.17226/9853. url: https://nap.nationalacademies.org/catalog/9853/how-people-learn-brain-mind-experience-and-school-expanded-edition (cit. on p. 41).

[162] S. Hajian. “Transfer of learning and teaching: A review of transfer theories and effective instructional practices.” In: *IAFOR Journal of education* 7.1 (2019), pp. 93–111 (cit. on p. 41).

[163] K. M. Sheldon and A. J. Elliot. “Goal striving, need satisfaction, and longitudinal well-being: the self-concordance model.” In: *Journal of personality and social psychology* 76.3 (1999), p. 482 (cit. on pp. 41, 42).

[164] R. J. Vallerand, M. S. Fortier, and F. Guay. “Self-determination and persistence in a real-life setting: toward a motivational model of high school dropout.” In: *Journal of Personality and Social psychology* 72.5 (1997), p. 1161 (cit. on p. 41).

[165] A. E. Black and E. L. Deci. “The effects of instructors’ autonomy support and students’ autonomous motivation on learning organic chemistry: A self-determination theory perspective”. In: *Science education* 84.6 (2000), pp. 740–756 (cit. on pp. 41, 42).

[166] M. O. Johansen, S. Eliassen, and L. M. Jeno. “The bright and dark side of autonomy: How autonomy support and thwarting relate to student motivation and academic functioning”. In: *Frontiers in Education*. Vol. 8. Frontiers Media SA. 2023, p. 1153647 (cit. on p. 41).

[167] S. Heikkinen et al. “A longitudinal study of interplay between student engagement and self-regulation”. In: *International Journal of Educational Technology in Higher Education* 22.1 (2025), p. 21 (cit. on p. 41).

[168] N.-J. Akpinar, A. Ramdas, and U. Acar. “Analyzing student strategies in blended courses using clickstream data”. In: *arXiv preprint arXiv:2006.00421* (2020) (cit. on p. 42).

[169] V. Paajanen. “LMS log activity as a predictor of learning success on an undergraduate flipped classroom course of cellular biology”. In: *Proceedings http://ceur-ws. org ISSN* 1613 (2022), p. 0073 (cit. on p. 42).

[170] L. von Keyserlingk et al. “Students’ study activities before and after exam deadlines as predictors of performance in STEM courses: A multi-source data analysis”. In: *Learning and Individual Differences* 117 (2025), p. 102598 (cit. on p. 42).

[171] H. W. Marsh and A. J. Martin. “Academic self-concept and academic achievement: Relations and causal ordering”. In: *British journal of educational psychology* 81.1 (2011), pp. 59–77 (cit. on p. 42).

[172] B. J. Zimmerman. “Becoming a self-regulated learner: An overview”. In: *Theory into practice* 41.2 (2002), pp. 64–70 (cit. on p. 42).

[173] B. Schmitz and B. S. Wiese. “New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data”. In: *Contemporary educational psychology* 31.1 (2006), pp. 64–96 (cit. on pp. 42, 54).

[174] A. Bandura. “Self-efficacy: toward a unifying theory of behavioral change.” In: *Psychological review* 84.2 (1977), p. 191 (cit. on p. 42).

[175] S. Hidi and K. A. Renninger. “The four-phase model of interest development”. In: *Educational psychologist* 41.2 (2006), pp. 111–127 (cit. on p. 42).

[176] R. J. Vallerand and G. Reid. “On the causal effects of perceived competence on intrinsic motivation: A test of cognitive evaluation theory”. In: *Journal of Sport and exercise Psychology* 6.1 (1984), pp. 94–102 (cit. on p. 42).

[177] K. R. Kim and E. H. Seo. “The relationship between procrastination and academic performance: A meta-analysis”. In: *Personality and individual differences* 82 (2015), pp. 26–33 (cit. on pp. 42, 43).

[178] F. Sana, T. Weston, and N. J. Cepeda. “Laptop multitasking hinders classroom learning for both users and nearby peers”. In: *Computers & Education* 62 (2013), pp. 24–31 (cit. on p. 42).

[179] S. Park. “Analysis of time-on-task, behavior experiences, and performance in two online courses with different authentic learning tasks”. In: *International Review of Research in Open and Distributed Learning* 18.2 (2017), pp. 213–233 (cit. on p. 42).

[180] B. Chen, S. Knight, and A. Wise. “Critical issues in designing and implementing temporal analytics.” In: *Journal of Learning Analytics* (2018) (cit. on pp. 43, 46, 48).

[181] N. Goel et al. “Circadian rhythms, sleep deprivation, and human performance”. In: *Progress in molecular biology and translational science* 119 (2013), pp. 155–190 (cit. on p. 43).

[182] K. J. Mehta. “Morning wake-time and the time of teaching/assessment session can influence test score”. In: *Humanities and Social Sciences Communications* 12.1 (2025), pp. 1–9 (cit. on p. 43).

[183] S. Li, J. Du, and J. Sun. “Unfolding the learning behaviour patterns of MOOC learners with different levels of achievement”. In: *International Journal of Educational Technology in Higher Education* 19.1 (2022), p. 22 (cit. on p. 43).

[184] E. M. Altmann and J. G. Trafton. “Memory for goals: An activation-based model”. In: *Cognitive science* 26.1 (2002), pp. 39–83 (cit. on p. 43).

[185] V. M. González and G. Mark. “”Constant, constant, multi-tasking craziness”managing multiple working spheres”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2004, pp. 113–120 (cit. on p. 43).

[186] G. Mark, D. Gudith, and U. Klocke. “The cost of interrupted work: more speed and stress”. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 2008, pp. 107–110 (cit. on p. 43).

[187] R. Junco and S. R. Cotten. “No A 4 U: The relationship between multitasking and academic performance”. In: *Computers & Education* 59.2 (2012), pp. 505–514 (cit. on p. 43).

[188] R. E. Landrum and R. A. Gurung. “Is Latency to Test Deadline a Predictor of Student Test Performance?” In: *Psychology Learning & Teaching* 12.3 (2013), pp. 227–235 (cit. on p. 43).

[189] N. D. Lane et al. “A survey of mobile phone sensing”. In: *IEEE Communications magazine* 48.9 (2010), pp. 140–150 (cit. on pp. 43, 44, 55).

[190] N. Eagle and A. Pentland. “Reality mining: sensing complex social systems”. In: *Personal and ubiquitous computing* 10.4 (2006), pp. 255–268 (cit. on p. 43).

[191] J. Torous et al. “New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research”. In: *JMIR mental health* 3.2 (2016), e5165 (cit. on p. 43).

[192] P. K. Doyle-Baker et al. “Smartphone GPS locations of students’ movements to and from campus”. In: *ISPRS International Journal of Geo-Information* 10.8 (2021), p. 517 (cit. on p. 43).

[193] B. Tabuenca, D. Börner, and M. Kalz. “Effects of an ambient learning display on noise levels and perceived learning in a secondary school”. In: *IEEE Transactions on Learning Technologies* 14.1 (2021), pp. 69–80 (cit. on p. 43).

[194] A. Stopczynski et al. “Measuring large-scale social networks with high resolution”. In: *PloS one* 9.4 (2014), e95978 (cit. on pp. 43, 50).

[195] T. Kim et al. “Sociometric badges: Using sensor technology to capture new forms of collaboration”. In: *Journal of Organizational Behavior* 33.3 (2012), pp. 412–427 (cit. on p. 43).

[196] H. Zhang et al. “Passive social sensing with smartphones: a systematic review”. In: *Computing* 105.1 (2023), pp. 29–51 (cit. on p. 43).

[197] M. Lan and K. F. Hew. “Examining learning engagement in MOOCs: A self-determination theoretical perspective using mixed method”. In: *International Journal of Educational Technology in Higher Education* 17.1 (2020), p. 7 (cit. on p. 43).

[198] J.-P. Onnela et al. “Beiwe: A data collection platform for high-throughput digital phenotyping”. In: *Journal of Open Source Software* 6.68 (2021), p. 3417 (cit. on pp. 44, 45).

[199] D. Ferreira, V. Kostakos, and A. K. Dey. “AWARE: mobile context instrumentation framework”. In: *Frontiers in ICT* 2 (2015), p. 6 (cit. on p. 45).

[200] G. Cardone et al. “MSF: An efficient mobile phone sensing framework”. In: *International Journal of Distributed Sensor Networks* 9.3 (2013), p. 538937 (cit. on p. 45).

[201] Y. Altshuler et al. “Incremental learning with accuracy prediction of social and individual properties from mobile-phone data”. In: *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*. IEEE. 2012, pp. 969–974 (cit. on p. 45).

[202] D. T. Wagner, A. Rice, and A. R. Beresford. “Device analyzer: Understanding smartphone usage”. In: *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Springer. 2013, pp. 195–208 (cit. on pp. 45, 49).

[203] M. Böhmer et al. “Falling asleep with Angry Birds, Facebook and Kindle: a large scale study on mobile application usage”. In: *Proceedings of the 13th international conference on Human computer interaction with mobile devices and services*. 2011, pp. 47–56 (cit. on p. 45).

[204] A. Sahami Shirazi et al. “Large-scale assessment of mobile notifications”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’14. Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 3055–3064. isbn: 9781450324731. doi: 10.1145/2556288.2557189. url: https://doi.org/10.1145/2556288.2557189 (cit. on p. 45).

[205] K. Geyer et al. “Open-source smartphone app and tools for measuring, quantifying, and visualizing technology use”. In: *Behavior Research Methods* 54.1 (2022), pp. 1–12 (cit. on p. 45).

[206] T. Deng et al. “Measuring smartphone usage and task switching with log tracking and self-reports”. In: *Mobile Media & Communication* 7.1 (2019), pp. 3–23 (cit. on p. 45).

[207] A. Carroll and G. Heiser. “An analysis of power consumption in a smartphone”. In: *2010 USENIX Annual Technical Conference (USENIX ATC 10)*. 2010 (cit. on p. 45).

[208] J. Paek, J. Kim, and R. Govindan. “Energy-efficient rate-adaptive GPS-based positioning for smartphones”. In: *Proceedings of the 8th international conference on Mobile systems, applications, and services*. 2010, pp. 299–314 (cit. on p. 45).

[209] K. K. Rachuri et al. “EmotionSense: a mobile phones based adaptive platform for experimental social psychology research”. In: *Proceedings of the 12th ACM international conference on Ubiquitous computing*. 2010, pp. 281–290 (cit. on p. 45).

[210] Z. Yan et al. “Energy-efficient continuous activity recognition on mobile phones: An activity-adaptive approach”. In: *2012 16th international symposium on wearable computers*. Ieee. 2012, pp. 17–24 (cit. on p. 45).

[211] R. R. Jagesar, J. A. Vorstman, and M. J. Kas. “Requirements and operational guidelines for secure and sustainable digital phenotyping: Design and development study”. In: *Journal of Medical Internet Research* 23.4 (2021), e20996 (cit. on pp. 45, 46).

[212] T. R. Insel. “Digital phenotyping: technology for a new science of behavior”. In: *Jama* 318.13 (2017), pp. 1215–1216 (cit. on p. 45).

[213] J.-P. Onnela and S. L. Rauch. “Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health”. In: *Neuropsychopharmacology* 41.7 (2016), pp. 1691–1696. doi: 10.1038/npp.2016.7 (cit. on pp. 45, 46).

[214] G. M. Harari. “A process-oriented approach to respecting privacy in the context of mobile phone tracking”. In: *Current opinion in psychology* 31 (2020), pp. 141–147 (cit. on p. 46).

[215] C. De Terwangne. “Principles relating to processing of personal data”. In: *The EU general data protection (GDPR): a commentary*. Oxford University Press, 2020, pp. 309–320 (cit. on p. 46).

[216] H. Lu et al. “The jigsaw continuous sensing engine for mobile phone applications”. In: *Proceedings of the 8th ACM conference on embedded networked sensor systems*. 2010, pp. 71–84 (cit. on pp. 46, 48).

[217] A. Ferrari et al. “Trends in human activity recognition using smartphones”. In: *Journal of Reliable Intelligent Environments* 7.3 (2021), pp. 189–213 (cit. on p. 46).

[218] J. S. Lee et al. “Smartphone and wearable device-based Digital phenotyping to Understand Substance use and its Syndemics”. In: *Journal of Medical Toxicology* 20.2 (2024), pp. 205–214 (cit. on p. 46).

[219] A. Bulling, U. Blanke, and B. Schiele. “A tutorial on human activity recognition using body-worn inertial sensors”. In: *ACM Computing Surveys (CSUR)* 46.3 (2014), pp. 1–33 (cit. on pp. 46, 49).

[220] Z. Papamitsiou and A. A. Economides. “Temporal learning analytics visualizations for increasing awareness during assessment”. In: *International Journal of Educational Technology in Higher Education* 12.3 (2015), pp. 129–147 (cit. on pp. 46, 49).

[221] A. Maslennikova, D. Rotelli, and A. Monreale. “Session-Based Time-Window Identification in Virtual Learning Environments.” In: *Journal of Learning Analytics* 10.3 (2023), pp. 7–27 (cit. on p. 46).

[222] D. C. Mohr, M. Zhang, and S. M. Schueller. “Personal sensing: understanding mental health using ubiquitous sensors and machine learning”. In: *Annual review of clinical psychology* 13.1 (2017), pp. 23–47 (cit. on p. 46).

[223] V. P. Cornet and R. J. Holden. “Systematic review of smartphone-based passive sensing for health and wellbeing”. In: *Journal of biomedical informatics* 77 (2018), pp. 120–132 (cit. on pp. 46, 55, 57).

[224] J. F. Huckins et al. “Mental health and behavior of college students during the early phases of the COVID-19 pandemic: Longitudinal smartphone and ecological momentary assessment study”. In: *Journal of medical Internet research* 22.6 (2020), e20185 (cit. on pp. 46, 55, 57, 61).

[225] A. Trifan, M. Oliveira, J. L. Oliveira, et al. “Passive sensing of health outcomes through smartphones: systematic review of current solutions and possible limitations”. In: *JMIR mHealth and uHealth* 7.8 (2019), e12649 (cit. on pp. 46, 55, 61).

[226] N. Eagle and A. S. Pentland. “Eigenbehaviors: Identifying structure in routine”. In: *Behavioral ecology and sociobiology* 63.7 (2009), pp. 1057–1066 (cit. on p. 48).

[227] T. Huynh, M. Fritz, and B. Schiele. “Discovery of activity patterns using topic models”. In: *Proceedings of the 10th international conference on Ubiquitous computing*. 2008, pp. 10–19 (cit. on p. 48).

[228] S. Knight, A. F. Wise, and B. Chen. “Time for change: Why learning analytics needs temporal analysis”. In: *Journal of Learning Analytics* 4.3 (2017), pp. 7–17 (cit. on pp. 48, 49).

[229] A. Sameh et al. “Digital phenotypes and digital biomarkers for health and diseases: a systematic review of machine learning approaches utilizing passive non-invasive signals collected via wearable devices and smartphones”. In: *Artificial Intelligence Review* 58.2 (2024), p. 66 (cit. on p. 49).

[230] D. Pollreisz and N. TaheriNejad. “Detection and removal of motion artifacts in PPG signals”. In: *Mobile Networks and Applications* 27.2 (2022), pp. 728–738 (cit. on p. 49).

[231] G. Clifford et al. “Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms”. In: *Physiological measurement* 33.9 (2012), p. 1419 (cit. on p. 49).

[232] Y. Xu et al. “Effects of data preprocessing methods on addressing location uncertainty in mobile signaling data”. In: *Annals of the American Association of Geographers* 111.2 (2020), pp. 515–539 (cit. on p. 49).

[233] E. L. Hamaker, R. M. Kuiper, and R. P. Grasman. “A critique of the cross-lagged panel model.” In: *Psychological methods* 20.1 (2015), p. 102 (cit. on p. 49).

[234] H. Falaki et al. “A first look at traffic on smartphones”. In: *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. 2010, pp. 281–287 (cit. on p. 49).

[235] M. Pielot, K. Church, and R. De Oliveira. “An in-situ study of mobile phone notifications”. In: *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. 2014, pp. 233–242 (cit. on p. 49).

[236] A. Mehrotra et al. “My phone and me: understanding people’s receptivity to mobile notifications”. In: *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2016, pp. 1021–1032 (cit. on p. 49).

[237] C. Song et al. “Limits of predictability in human mobility”. In: *Science* 327.5968 (2010), pp. 1018–1021 (cit. on p. 50).

[238] Y. Wang, A. Yalcin, and C. VandeWeerd. “An entropy-based approach to the study of human mobility and behavior in private homes”. In: *PLoS one* 15.12 (2020), e0243503 (cit. on p. 50).

[239] H. Zhang, R. Dantu, and J. Cangussu. “Quantifying reciprocity in social networks”. In: *2009 International Conference on Computational Science and Engineering*. Vol. 4. IEEE. 2009, pp. 1031–1035 (cit. on p. 50).

[240] J. B. McQueen. “Some methods of classification and analysis of multivariate observations”. In: *Proc. of 5th Berkeley Symposium on Math. Stat. and Prob.* 1967, pp. 281–297 (cit. on p. 50).

[241] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum likelihood from incomplete data via the EM algorithm”. In: *Journal of the royal statistical society: series B (methodological)* 39.1 (1977), pp. 1–22 (cit. on p. 50).

[242] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake. “Finite mixture models”. In: *Annual review of statistics and its application* 6.1 (2019), pp. 355–378 (cit. on p. 50).

[243] R. J. Campello, D. Moulavi, and J. Sander. “Density-based clustering based on hierarchical density estimates”. In: *Pacific-Asia conference on knowledge discovery and data mining*. Springer. 2013, pp. 160–172 (cit. on p. 50).

[244] U. Von Luxburg. “A tutorial on spectral clustering”. In: *Statistics and computing* 17.4 (2007), pp. 395–416 (cit. on p. 50).

[245] S. M. Lundberg and S.-I. Lee. “A unified approach to interpreting model predictions”. In: *Advances in neural information processing systems* 30 (2017) (cit. on p. 51).

[246] S. Mishra et al. “Cross-task generalization via natural language crowdsourcing instructions”. In: *arXiv preprint arXiv:2104.08773* (2021) (cit. on p. 52).

[247] V. Sanh et al. “Multitask prompted training enables zero-shot task generalization”. In: *arXiv preprint arXiv:2110.08207* (2021) (cit. on p. 52).

[248] L. Ouyang et al. “Training language models to follow instructions with human feedback”. In: *Advances in neural information processing systems* 35 (2022), pp. 27730–27744 (cit. on p. 52).

[249] J. Wei et al. “Chain-of-thought prompting elicits reasoning in large language models”. In: *Advances in neural information processing systems* 35 (2022), pp. 24824–24837 (cit. on p. 52).

[250] T. Kojima et al. “Large language models are zero-shot reasoners”. In: *Advances in neural information processing systems* 35 (2022), pp. 22199–22213 (cit. on p. 52).

[251] S. Yao et al. “React: Synergizing reasoning and acting in language models”. In: *The eleventh international conference on learning representations*. 2022 (cit. on p. 52).

[252] Z. C. Lipton. “The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery.” In: *Queue* 16.3 (2018), pp. 31–57 (cit. on p. 52).

[253] S. J. Pan and Q. Yang. “A survey on transfer learning”. In: *IEEE Transactions on knowledge and data engineering* 22.10 (2009), pp. 1345–1359 (cit. on p. 52).

[254] T. Curran et al. “A three-wave longitudinal test of self-determination theory’s mediation model of engagement and disaffection in youth sport”. In: *Journal of Sport and Exercise Psychology* 38.1 (2016), pp. 15–29 (cit. on p. 54).

[255] H. Bellhäuser, C. Dignath, and M. Theobald. “Daily automated feedback enhances self-regulated learning: a longitudinal randomized field experiment”. In: *Frontiers in Psychology* 14 (2023), p. 1125873 (cit. on p. 54).

[256] R. J. Vandenberg and C. E. Lance. “A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research”. In: *Organizational research methods* 3.1 (2000), pp. 4–70 (cit. on p. 54).

[257] M. K. Smith et al. “The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices”. In: *CBE—Life Sciences Education* 12.4 (2013), pp. 618–627 (cit. on pp. 54, 61).

[258] M. Stains et al. “Anatomy of STEM teaching in North American universities”. In: *Science* 359.6383 (2018), pp. 1468–1470 (cit. on pp. 54, 61).

[259] B. DiCicco-Bloom and B. F. Crabtree. “The qualitative research interview”. In: *Medical education* 40.4 (2006), pp. 314–321 (cit. on pp. 54, 61).

[260] N. K. Denzin. “A theoretical introduction to sociological methods”. In: *Chicago, IL: Aldlne* (1970) (cit. on p. 55).

[261] T. D. Jick. “Mixing qualitative and quantitative methods: Triangulation in action”. In: *Administrative science quarterly* 24.4 (1979), pp. 602–611 (cit. on p. 55).

[262] J. W. Cox and J. Hassard. “Triangulation in organizational research: A re-presentation”. In: *Organization* 12.1 (2005), pp. 109–133 (cit. on p. 55).

[263] A. D. Ho and T. J. Kane. “The Reliability of Classroom Observations by School Personnel. Research Paper. MET Project.” In: *Bill & Melinda Gates Foundation* (2013) (cit. on p. 55).

[264] S. Slade and P. Prinsloo. “Learning analytics: Ethical issues and dilemmas”. In: *American Behavioral Scientist* 57.10 (2013), pp. 1510–1529 (cit. on p. 55).

[265] S. Shiffman, A. A. Stone, and M. R. Hufford. “Ecological momentary assessment”. In: *Annu. Rev. Clin. Psychol.* 4.1 (2008), pp. 1–32 (cit. on pp. 55, 61).

[266] T. J. Trull and U. Ebner-Priemer. “Ambulatory assessment”. In: *Annual review of clinical psychology* 9.1 (2013), pp. 151–176 (cit. on pp. 55, 61).

[267] C. Wrzus and A. B. Neubauer. “Ecological momentary assessment: A meta-analysis on designs, samples, and compliance across research fields”. In: *Assessment* 30.3 (2023), pp. 825–846 (cit. on pp. 55, 61).

[268] A. L. Howard and M. Lamb. “Compliance trends in a 14-week ecological momentary assessment study of undergraduate alcohol drinkers”. In: *Assessment* 31.2 (2024), pp. 277–290 (cit. on pp. 55, 61).

[269] C. N. Scollon, C. Kim-Prieto, and E. Diener. “Experience sampling: Promises and pitfalls, strengths and weaknesses”. In: *Journal of Happiness studies* 4.1 (2003), pp. 5–34 (cit. on p. 55).

[270] M. Csikszentmihalyi. *Handbook of research methods for studying daily life*. Guilford Publications, 2013 (cit. on p. 55).

[271] E. Dejonckheere and Y. Erbas. “Designing an experience sampling study”. In: (2021) (cit. on p. 55).

[272] G. M. Harari et al. “Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges”. In: *Perspectives on Psychological Science* 11.6 (2016), pp. 838–854. doi: 10.1177/1745691616650285. url: https://journals.sagepub.com/doi/10.1177/1745691616650285 (cit. on p. 55).

[273] S. Tonkin et al. “Evaluating declines in compliance with ecological momentary assessment in longitudinal health behavior research: Analyses from a clinical trial”. In: *Journal of Medical Internet Research* 25 (2023), e43826 (cit. on p. 55).

[274] D. A. Rohani et al. “Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review”. In: *JMIR mHealth and uHealth* 6.8 (2018), e9691 (cit. on p. 55).

[275] M. Bauer et al. “Smartphones in mental health: a critical review of background issues, current status and future concerns”. In: *International journal of bipolar disorders* 8.1 (2020), p. 2 (cit. on pp. 55, 61).

[276] S. Saeb et al. “Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study”. In: *Journal of medical Internet research* 17.7 (2015), e4273 (cit. on pp. 55, 61).

[277] D. Ben-Zeev et al. “Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health.” In: *Psychiatric rehabilitation journal* 38.3 (2015), p. 218 (cit. on pp. 55, 61).

[278] E. Smets et al. “Large-scale wearable data reveal digital phenotypes for daily-life stress detection”. In: *NPJ digital medicine* 1.1 (2018), p. 67 (cit. on pp. 55, 61).

[279] A. Anderson et al. “Engaging with massive online courses”. In: *Proceedings of the 23rd international conference on World wide web*. 2014, pp. 687–698 (cit. on pp. 55, 61).

[280] T. Sinha et al. “Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions”. In: *arXiv preprint arXiv:1407.7131* (2014) (cit. on pp. 55, 61).

[281] R. S. Baker, T. Martin, and L. M. Rossi. “Educational data mining and learning analytics”. In: *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (2016), pp. 379–396 (cit. on pp. 55, 61).

[282] S. C. Yeo et al. “Early morning university classes are associated with impaired sleep and academic performance”. In: *Nature Human Behaviour* 7.4 (2023), pp. 502–514 (cit. on pp. 55, 61).

[283] S. Sinche et al. “Analysis of student academic performance using human-in-the-loop cyber-physical systems”. In: *Telecom*. Vol. 1. 1. MDPI. 2020, pp. 18–31 (cit. on pp. 55, 61).

[284] V. Kassarnig et al. “Class attendance, peer similarity, and academic performance in a large field study”. In: *PloS one* 12.11 (2017), e0187078 (cit. on pp. 55, 61).

[285] D. L. Mack et al. “Mental health and behavior of college students during the COVID-19 pandemic: longitudinal mobile smartphone and ecological momentary assessment study, part II”. In: *Journal of medical Internet research* 23.6 (2021), e28892 (cit. on p. 55).

[286] C. Anderl, M. K. Hofer, and F. S. Chen. “Directly-measured smartphone screen time predicts well-being and feelings of social connectedness”. In: *Journal of Social and Personal Relationships* 41.5 (2024), pp. 1073–1090 (cit. on p. 55).

[287] A. Madan et al. “Sensing the ”Health State”of a Community”. In: *IEEE Pervasive Computing* 11.4 (2012-10), pp. 36–45. issn: 1536-1268. doi: 10.1109/MPRV.2011.79. url: https://doi.org/10.1109/MPRV.2011.79 (cit. on p. 56).

[288] C. Wu et al. “Improving prediction of real-time loneliness and companionship type using geosocial features of personal smartphone data”. In: *Smart Health* 20 (2021), p. 100180 (cit. on pp. 56, 61).

[289] A. Bogomolov et al. “Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits”. In: *Proceedings of the 22nd ACM International Conference on Multimedia*. MM ’14. Orlando, Florida, USA: Association for Computing Machinery, 2014, pp. 477–486. isbn: 9781450330633. doi: 10.1145/2647868.2654933. url: https://doi.org/10.1145/2647868.2654933 (cit. on p. 56).

[290] S. Jafarlou et al. “Objective monitoring of loneliness levels using smart devices: A multi-device approach for mental health applications”. In: *Plos one* 19.6 (2024), e0298949 (cit. on p. 56).

[291] A. H. Bradley and A. L. Howard. “Stress and mood associations with smartphone use in university students: a 12-week longitudinal study”. In: *Clinical Psychological Science* 11.5 (2023), pp. 921–941 (cit. on p. 56).

[292] L. Canzian and M. Musolesi. “Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis”. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp ’15. Osaka, Japan: Association for Computing Machinery, 2015, pp. 1293–1304. isbn: 9781450335744. doi: 10.1145/2750858.2805845. url: https://doi.org/10.1145/2750858.2805845 (cit. on p. 56).

[293] A. Choi, D. Lottridge, and J. Warren. “Personalised modelling of routine variability and affective states”. In: *npj Digital Medicine* 8.1 (2025), p. 597 (cit. on p. 56).

[294] S. Nepal et al. “Capturing the college experience: A four-year mobile sensing study of mental health, resilience and behavior of college students during the pandemic”. In: *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 8.1 (2024), pp. 1–37 (cit. on p. 57).

[295] T.-Y. Lee et al. “Dynamic Bidirectional Associations Between Global Positioning System Mobility and Ecological Momentary Assessment of Mood Symptoms in Mood Disorders: Prospective Cohort Study”. In: *Journal of Medical Internet Research* 26 (2024), e55635 (cit. on p. 57).

[296] L. Yi et al. “Measuring environmental and behavioral drivers of chronic diseases using smartphone-based digital phenotyping: intensive longitudinal observational mHealth substudy embedded in 2 prospective cohorts of adults”. In: *JMIR Public Health and Surveillance* 10 (2024), e55170 (cit. on p. 57).

[297] C. Slade, R. M. Benzo, and P. Washington. “Design Guidelines for Improving Mobile Sensing Data Collection: Prospective Mixed Methods Study”. In: *Journal of Medical Internet Research* 26 (2024), e55694 (cit. on p. 57).

[298] I. Jolliffe. “Principal component analysis”. In: *International encyclopedia of statistical science*. Springer, 2011, pp. 1094–1096 (cit. on p. 89).

[299] P. J. Rousseeuw. “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”. In: *Journal of computational and applied mathematics* 20 (1987), pp. 53–65 (cit. on p. 102).

[300] R. F. Woolson. “Wilcoxon signed-rank test”. In: *Wiley encyclopedia of clinical trials* (2007), pp. 1–3 (cit. on p. 104).

[301] L. V. Phan et al. “Mobile Sensing around the Globe: Considerations for Cross-Cultural Research”. In: *Technology and Measurement around the Globe*. Ed. by L. Tay, S. E. Woo, and T. Behrend. Educational and Psychological Testing in a Global Context. Cambridge University Press, 2023, pp. 176–210 (cit. on p. 128).

[302] J.-M. Andres-Bray, S. Hutt, and R. S. Baker. "Exploring Cross-Country Prediction Model Generalizability in MOOCs". In: *Proceedings of the Tenth ACM Conference on Learning@ Scale*. 2023, pp. 183–194 (cit. on p. 128).

[303] K. Huckvale, S. Venkatesh, and H. Christensen. "Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety". In: *NPJ digital medicine* 2.1 (2019), p. 88 (cit. on p. 128).

[304] F. Li and C. Wang. "Artificial intelligence and edge computing for teaching quality evaluation based on 5G-enabled wireless communication technology". In: *Journal of Cloud Computing* 12.1 (2023), p. 45 (cit. on p. 129).

[305] Q. Chen et al. "Educational 5G edge computing: Framework and experimental study". In: *Electronics* 11.17 (2022), p. 2727 (cit. on p. 129).

[306] K. Rogers et al. "Playing Dumb to Get Smart: Creating and Evaluating an LLM-based Teachable Agent within University Computer Science Classes". In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. CHI '25. New York, NY, USA: Association for Computing Machinery, 2025. isbn: 9798400713941. doi: 10.1145/3706598.3713644. url: https://doi.org/10.1145/3706598.3713644 (cit. on p. 129).

[307] L. Dai et al. "Agent4EDU: Advancing AI for Education with Agentic Workflows". In: *Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Education*. ICAIE '24. New York, NY, USA: Association for Computing Machinery, 2025, pp. 180–185. isbn: 9798400712692. doi: 10.1145/3722237.3722268. url: https://doi.org/10.1145/3722237.3722268 (cit. on p. 129).

[308] C. Gao et al. "Large language models empowered agent-based modeling and simulation: A survey and perspectives". In: *Humanities and Social Sciences Communications* 11.1 (2024), pp. 1–24 (cit. on p. 129).

[309] T. Susnjak, G. S. Ramaswami, and A. Mathrani. "Learning analytics dashboard: a tool for providing actionable insights to learners". In: *International Journal of Educational Technology in Higher Education* 19.1 (2022), p. 12 (cit. on p. 129).

[310] L. Paulsen and E. Lindsay. "Learning analytics dashboards are increasingly becoming about learning and not just analytics-A systematic review". In: *Education and Information Technologies* 29.11 (2024), pp. 14279–14308 (cit. on p. 129).

[311] N. Valle et al. "Staying on target: A systematic literature review on learner-facing learning analytics dashboards". In: *British Journal of Educational Technology* 52.4 (2021), pp. 1724–1748 (cit. on p. 129).

[312] W. Holmes, F. Miao, et al. *Guidance for generative AI in education and research*. Unesco Publishing, 2023 (cit. on pp. 129–131).

[313] Q. Liu and M. Khalil. “Understanding privacy and data protection issues in learning analytics using a systematic review”. In: *British Journal of Educational Technology* 54.6 (2023), pp. 1715–1747 (cit. on pp. 129–131).

[314] I. Rets, C. Herodotou, and A. Gillespie. “Six Practical Recommendations Enabling Ethical Use of Predictive Learning Analytics in Distance Education.” In: *Journal of Learning Analytics* 10.1 (2023), pp. 149–167 (cit. on pp. 129–131).