Research Article

Vanessa W. Vongkulluksn*, Kui Xie

Multilevel Latent State-Trait Models with Experience Sampling Data: An Illustrative Case of Examining Situational Engagement

https://doi.org/10.1515/edu-2022-0016
received July 12, 2021; accepted April 28, 2022.

Abstract: Learning processes often occur at a situational level. Changes in learning context have implications on how students are motivated or are able to cognitively process information.

To study such situational phenomena, Experience Sampling Method (ESM) can help assess psychological variables in the moment and in context. However, data collected via ESM is voluminous and imbalanced. Special types of statistical modeling are needed to handle this unique data structure in order to maximize its potential for scientific discovery. The purpose of this paper is to illustrate how Latent State-Trait modeling used within a multilevel framework can help model complex data as derived by ESM. A study of situational engagement is presented as an illustrative case. We describe methodological considerations which facilitated the following analyses: (1) *Decomposition of trait-level and state-level engagement*; (2) *Group differences in variance decomposition*, and (3) *Predicting state component of engagement*. Discussions include the relative advantages and disadvantages of ESM and multilevel Latent State-Trait modeling in facilitating situational psychological research.

Keywords: Experiential Sampling; Situational Engagement; Latent Variable Modeling; Latent State-Trait Modeling.

1 Introduction

There is increasing recognition that many psychological variables function at a situational level. That is, learning processes such as motivation and cognition differ across contexts depending on learning tasks and environments (Lavigne & Vallerand, 2010; Xie, Heddy, & Vongkulluksn, 2019). Additionally, the permeation of technology means that learning contexts have never been more diverse, from classroom settings to mobile learning environments that are adaptive to the learner and the learning task (Crompton, 2013). Therefore, scholars have recently suggested an emergent need for a person-in-context orientation in education research (Sinatra et al., 2015). This means investigating how learning processes are influenced by both contextual features of the learning environment and person-specific factors. The person-in-context research orientation often requires complex data that represent fluctuations in psychological variables such as motivational and cognitive responses to different learning contexts.

Data collection methods have been created to help assess psychological variables in the moment and in context, facilitating the person-in-context research orientation. One of the most prominent is the experience sampling method (ESM), which entails repeatedly administering surveys as respondents are embedded within the context in which learning occurs (Csikszentmihalyi & Larson, 2014). The rise of mobile technology has made ESM data collection easier and less intrusive (Xie, Greene, & Heddy, 2019; Xie, Heddy, & Vongkulluksn, 2019). ESM gathers the type of complex data needed to understand learning phenomena in a situational manner. However, data collected via ESM is voluminous and imbalanced. ESM data is collected across numerous time points which, depending to the scheduling method, can be at inconsistent temporal intervals. Special types of statistical modeling are needed to handle this unique data structure in order to maximize its potential for scientific discovery.

The purpose of this paper is to illustrate how Latent State-Trait modeling used within a multilevel framework (Geiser, 2021; Geiser et al., 2013) can help model complex data as derived by ESM. We make the case that the

***Corresponding author: Vanessa W. Vongkulluksn,** University of Nevada, Las Vegas, Nevada, United States, E-mail: vanessa.vongkulluksn@unlv.edu
Kui Xie, The Ohio State University, Columbus, Ohio, United States

combination of ESM data collection facilitated by mobile technology (Xie, Greene, & Heddy, 2019; Xie, Heddy, & Vongkulluksn, 2019) and multilevel Latent State-Trait analysis offers an innovative way to facilitate person-in-context research within diverse learning environments. A study of situational engagement is presented as an illustrative case.

2 Person-in-Context Research and Experience Sampling Method

Person-in-context education research focuses the research lens on the interaction between the individual learner and the particular learning context where learning activities occur. Sinatra and associates (2015) described the person-in-context research orientation as existing within a continuum of engagement measurement and research focus. Person-oriented research focuses on examining individuals' motivation and learning, whereas context-oriented research focuses on describing characteristics of the learning environment. Person-in-context research exists in the middle of the continuum where the focus of research is examining individuals as being embedded within the learning context and the unique interactions that occur between the two. This research orientation gets at learning processes at the situational level. Learning is not seen as only influenced by global processes such as students' overall motivation to learn or teachers' setup of the learning environment. Rather, there is a reciprocation between the person and the context as demonstrated by how students' learning responses change depending on the interactions of contextual features with their personal characteristics (Bandura, 2001; Xie, Heddy, & Vongkulluksn, 2019).

At the very heart of the person-in-context research orientation is the recognition that psychological constructs exist at both the "trait" and the "state" levels (Curran & Bauer, 2011; Geiser, 2021). That is, each learner may tend to be motivated and engaged in a certain way in school or within a content domain over time (trait motivation or engagement) but also has a specific response to a particular learning task (state motivation or engagement). It's important to keep in mind that the label "trait" in this case does not imply immutability, only that these characteristics tend to remain relatively stable over time. In our previous work (Xie, Heddy, & Vongkulluksn, 2019), we examined how students' self-efficacy – or the belief in their ability to perform well on a specific task (Bandura, 1997) – has both trait- and state-level components. When thus decomposed, trait-level and state-level self-efficacy interact with contextual features in different ways to influence learning engagement. Examining psychological variables like self-efficacy at both trait- and state-levels at the same time expand theories on learning because we can better pinpoint in-the-moment processes versus person-level processes. This helps us design learning environments that consciously target both positive momentary responses to instruction as well as adaptive development of individuals' characteristics.

Researchers' ability to compartmentalize trait-level vs. state-level components of psychological variables depends on collecting intensive longitudinal data that represent their fine-grained fluctuations in context. Experience-Sampling Method (ESM) is a technique used to collect data as people perform relevant tasks, and thus is able to capture data that reflects what is happening at the moment (Zirkel et al., 2015). ESM is a situational approach to data collection; it is a more sensitive method of examining motivation and cognition during learning than traditional self-report methods that rely on either projection (i.e., data gathered before learning events occur) or retrospection (i.e., data gathered after learning events occurred). A collection of recent studies has shown that ESM allows for a more nuanced view of students' motivation and cognition as they relate to the context of learning. For example, Schmidt and colleagues (2008, 2013, 2015, 2018) used ESM data to discern motivational differences across contexts, such as an increase in intrinsic motivation when students read outside school versus a tendency toward goal-directed behaviors when they read in school (Shumow et al., 2008). ESM is thus a method well-suited to facilitate person-in-context research (Sinatra et al., 2015), uncovering the context-dependent nature of motivation and cognition.

There are three main types of ESM sampling techniques: random, fixed, and event-based sampling. (Zirkel, Garcia, & Murphy, 2015). In random sampling, a specific number of data collection events are specified and participants are assessed randomly throughout a given data collection period until the desired number is reached. In fixed sampling, data collection is set for a specific time of the day or week during the data collection period (e.g., every day at 3 pm or every Friday at 3 pm), and the data collection intervals are thus fixed. In event-based sampling, data collection occurs during or after specific events, such as the target learning task. In our previous work, we found that event-based sampling was an effective method for improving sampling accuracy, reducing extraneous prompts, and reducing missing data in a study of out-of-classroom learning behaviors (Xie, Greene,

& Heddy, 2019). While event-based sampling captured about equivalent numbers of data points compared to fixed sampling, event-based sampling had a much higher capturing rate (84.7% vs. 31%). Event-based sampling was more accurate at collecting data during relevant activities, whereas fixed sampling prompted students more often with less accuracy. Event-based sampling is shown to be a highly effective data collection method, especially for examining unstructured environments like out-of-classroom learning.

However, two problematic aspects arise with event-based experience sampling. First, event-based sampling requires a high degree of coordination when conducted outside structured environments like the classroom. Generally, participants in ESM will respond to surveys at times when they are prompted by the researcher through some type of alarm device. In fixed sampling, this is an easy matter of setting the alarm for a specific time each day or week. In event-based sampling, the alarm needs to be programed for likely times students will engage in relevant learning activities. Our previous research solved this issue using the *ESM-Mobile approach*, using mobile devices to manage scheduling of ESM prompts and survey delivery. Students schedule a time on the ESM Mobile app for when they are likely to engage in relevant learning activities (e.g., study for a specific class) and the app prompts them to answer a brief survey during the pre-scheduled time. With increased accessibility and connectivity of mobile devices, ESM research has become possible even when the learning contexts of interest are diverse and unstructured.

Second, while all ESM sampling methods produce intensive longitudinal data, data collected via event-sampling method tend to be additionally complex because of its imbalanced nature. The data is imbalanced because intervals between time points vary (Geiser, 2021). The ESM-Mobile approach provides the flexibility needed for students to schedule ESM data collection during the most relevant times according to their personal schedules. However, this means intervals between data collection points will vary widely between adjacent time points within students as well as between students. Additionally, the data is also imbalanced because there is likely not the same number of data collection points across students. For example, in a research study on out-of-classroom learning, some students may study every night five days a week, while other students may only study once per week. Some students may have a routine study schedule, while others may study at convenient times. Therefore, the frequency and interval of ESM events based on their study habits would differ quite substantially from each other.

Finally, ESM inherently produces missing data, either because students fail to respond to the prompt or because they are not engaging in the relevant task as previously planned. Although event-based sampling increase sampling accuracy and thereby reduce missingness, data collected through this method still tend to have a larger amount of missingness than other types of psychological measurement.

3 Comparison of Three Contemporary Statistical Modeling Approaches

Due to the above-mentioned complexities, researchers utilizing ESM data – especially event-based ESM data – need to consider specific types of statistical modeling which leverage the unique affordances of the intensive longitudinal data while overcoming data structure issues. We discuss relative merits of some statistical modeling approaches below.

3.1 Linear Regressions within Hierarchical Linear Modeling (HLM) Framework

Intensive longitudinal data collected via ESM assess certain psychological variables from the same individuals across multiple time points. This repeated-measures data has an inherent hierarchical structure, with observations nested within individuals (Geiser et al., 2013; Rabe-Hesketh & Skrondal, 2012; Raudenbush & Bryk, 2002). Within the hierarchical linear modeling framework, linear regressions can account for within-individual dependence by splitting the residuals into (1) the random effect component representing differences between individuals' mean scores and the sample mean, and (2) within-subject residual representing the random deviation of each observation from the individual's mean score. For researchers interested in trait- vs. state-level functioning of a particular psychological variable, within versus between effects of a predictor variable can be modeled to compare how each component may differentially influence the outcome (for model specification, please refer to Rabe-Hesketh & Skrondal, 2012, p. 152). For example, self-efficacy can be decomposed into trait-level self-efficacy representing individuals' mean scores across observations and state-level self-efficacy representing the deviations between each observation and the trait-level mean (Xie, Heddy, & Vongkulluksn, 2019). The within

Table 1: Comparison of Three Longitudinal Statistical Modeling Approaches.

	HLM Linear Regression	Single-Level Latent State-Trait SEM	Multilevel Latent State-Trait SEM
Complex Nature of Intensive Longitudinal Data			
Large Number of Data Points	More compact modeling without need to model each data point individually	Need to specify measurement and structural models for each time point	More compact modeling without need to model each data point individually
Imbalanced Data	Implicitly handled through nested data modeling. Allows for unequal number of time points and time intervals	Requires the same number of data collection points and constant time intervals across individuals	Implicitly handled through nested data modeling. Allows for unequal number of time points and time intervals
Missing Data	Implicitly handled through nested data modeling	Can use FIML to handle missingness with Missing at Random or Missing Completely at Random	Implicitly handled through nested data modeling
Additional Considerations			
Ease of Modeling	Only one dependent variable may be modeled at one time. Straightforward modeling and interpretation.	More flexibility in modeling, while still requiring constant time intervals across individuals	High flexibility in modeling, allowing for multiple independent and dependent variables as well as unequal number of time points and time intervals
Modeling Measurement Error	Does not estimate measurement error	Can specify measurement model and estimate measurement error	Can specify measurement model and estimate measurement error

and between effects of self-efficacy on students' learning outcomes can then be discerned.

Modeling longitudinal data with linear regressions within the HLM framework is advantageous because the imbalanced longitudinal data structure is implicitly handled through the specification of random effect at the individual level (Table 1). That is, explicit modeling is not required for each observation individually, making model specification more compact (Geiser et al., 2013). Likewise, missingness of observations due to non-response or sampling inaccuracy is also implicitly handled through this nested data modeling. Another advantage of HLM linear regression is its straightforward modeling due to the ease with which predictors can be included and their slope parameters interpreted (Gefen et al., 2000). However, the tradeoff is the restrictive modeling capabilities of linear regression, which allows unidirectional relationships for one dependent variable at a time. This is additionally restrictive for studies seeking to examine trait- vs. state-level characteristics of specific variables since it means that such decomposition could only take place for predictor and not outcome variables. An additional disadvantage of the HLM linear regression approach is its inability to model and estimate measurement error, which may lead to uncertainty about the amount of variation due to unreliable measures.

3.2 Single-Level Latent State-Trait Structural Equation Modeling

Longitudinal structural equation modeling is another powerful approach to represent longitudinal data. Unlike HLM linear regression, longitudinal structural equation modeling may include measurement models and associated latent variables, which distinguish between true individual differences and variability due to measurement error (Little, 2013). In longitudinal designs, this methodological capacity is crucial since researchers need to discern how much of the observed variability is due to true intra-individual changes over time (Geiser, 2021).

Within the universe of longitudinal SEM, a specific approach which explicitly addresses longitudinal stability vs. variability is Latent State-Trait (LST) structural equation models (Geiser et al., 2013; Steyer et al., 1992, 2015). LST class of models specify that each observed variable is a function of a latent *trait* variable characterizing person-level effects for the ith observed indicator at time t (ξ_{it} [psi]), a latent *state* variable characterizing occasion-level residual fluctuations (ζ_{it} [zeta]), and a *residual* term representing random measurement error (ε_{it} [epsilon]):

$$Y_{it} = \xi_{it} + \zeta_{it} + \varepsilon_{it}$$

[General LST Equation]

In this theorization, the means of both the latent state and error residual variables are set at zero, as they are defined as types of residuals (Geiser et al., 2013). All latent trait variables are also specified to be uncorrelated with latent state residuals and with all measurement errors. These specifications allow LST models to estimate the extent to which psychological variable is a function of trait-level characteristic vs. state-level response.

One of the simplest types of LST model that leverages multiple indicators of the target variable is the Singletrait-Multistate (STMS) model (Geiser, et al., 2013; Geiser, 2021). STMS models specify a common latent trait factor and one latent state residual factor shared by all indicators measured at the same occasion. So, each observed variable is a linear function of an intercept parameter, a common latent trait (ξ), a common latent state residual (ζ_t), and indicator specific measurement error:

$$Y_{it} = \alpha_i + \lambda_i \xi + \gamma_i \zeta_t + \varepsilon_{it}$$

[STMS Equation]

The hierarchical structure of the STMS model specifies that the second-order trait factor represents the person-level component of the target variable, whereas the first-order state factor represents the occasion-specific component.

Another model closely related to the STMS model is the Multitrait-Multistate (MTMS) model (Geiser et a., 2013; Geiser, 2021). The MTMS relaxes the assumption in the STMS model that all observed indicators share one common trait factor. This assumption may not be tenable in many practical measurement situations when the indicators measure some related but distinct aspects of a specific construct. The MTMS model specifies a different latent trait factor for each indicator. The same specification is made for the latent state factor as the STMS model, in which one state factor is shared by all observations with scaling differences γ_i [gamma]. Each observed variable is thus specified by the function:

$$Y_{it} = \xi_i + \gamma_i \zeta_t + \varepsilon_{it}$$

[MTMS Equation]

The latent trait factors are specified as correlated with one another, and such correlations can be interpreted as indicating the extent to which the indicators measure homogenous aspects of the underlying construct. All trait factor loadings are set to 1, while the means of each latent trait factor and state residual factor loadings are freely estimated.

Single-level LST models like the classic STMS and MTMS models have the advantage over HLM linear regression models because they are able to estimate how much of the observed variability is due to measurement error (Table 1). In addition, STMS and MTMS models allow for more flexibility in modeling compared to HLM linear regression since multiple predictors and outcomes can be added and estimated. Specifically, the target variable with decomposed trait- and state-level variance can be modeled either as a predictor of another outcome variable or as an outcome variable itself influenced by a set of predictors. On the flip side, these longitudinal SEM models require the same number of data collection points and constant time intervals across individuals. While this requirement may work for traditional longitudinal data collection and fixed-sampling ESM data, data collected via event-based ESM sampling will not work for this modeling approach. When considering the use of single-level LST models, there is thus a tradeoff between using a data collection method that can more accurately capture students during relevant activities like event-based sampling vs. the ability to more accurately and flexibly model longitudinal processes.

3.3 Multilevel Latent State-Trait Structural Equation Modeling

Multilevel LST models leverage the relative advantages of both HLM linear regression and single-level LST models. Multilevel LST models such as multilevel STMS and MTMS models have the ability to estimate measurement errors as well as allowing for an unbalanced data structure (Table 1). The specification of multilevel STMS and MTMS models estimate latent state factors and measurement error at the occasion level (level 1). Latent trait factors are estimated as random intercept parameters similar to HLM linear regression models and they are modeled at the person level (level 2).

In multilevel STMS models, three latent effects are modeled at level 1: (1) the common latent state factor with variance $var(\zeta)$ and loadings γ_i; (2) measurement error variance $var(\varepsilon_i)$; and (3) the time variable with slope coefficients set at zero with the assumption that trait factors do not change with time (Geiser et al., 2013; Geiser, 2021). At level 2, the time variable is specified as having random intercept parameters for each indicator, representing the latent trait factor. The STMS model places the assumption that all random intercepts are linear functions of a common trait factor. Thus, only one latent trait mean, variance parameter, intercept, and indicator-specific factor loadings are estimated at level 2. Multilevel

MTMS models are similarly specified, but indicator-specific trait factors (random intercepts) are estimated, each with a distinct mean and variance parameters. Thus specified, multilevel STMS and MTMS models combined the ability to estimate measurement error of single-level LST models and the capability to model unbalanced longitudinal data of HLM linear regressions (Table 1). Therefore, multilevel STMS and MTMS models are capable of modeling intensive longitudinal data, such as data that are derived from event-based sampling ESM.

3.4 Person-Level vs. Situation-Specific Engagement: An Illustrative Example

In order to illustrate how multilevel STMS and MTMS models may be used in combination with event-based ESM for education research, the following section will present a study of situational learning engagement.

Learning engagement is a multifaceted construct, including behavioral, cognitive, affective, and social dimensions (Wang et al., 2016). In this study, we operationalize situational engagement as a combination of behavioral and cognitive engagement. Behavioral engagement is defined as students' involvement in academic activities, comprising of factors like effort, persistence, and concentration (Buhs & Ladd, 2001; Wang et al., 2016). Cognitive engagement is defined as the use of information processing strategies, such as elaborating on information using prior knowledge and monitoring one's own understanding (Greene, 2015; Wang et al., 2016; Xie, Vongkulluksn, Lu, & Cheng, 2020). Engagement is an important indicator of successful learning episodes, having been shown to be closely tied to learning benefits such as increased achievement (e.g., Archambault et al., 2012; Xie, Hensley, Law, & Sun, 2019), interest (e.g., Heddy & Sinatra, 2017; Zhu et al., 2009), and academic identification (e.g., Walker et al., 2006; Wang & Eccles, 2012).

Recently, there is recognition that there is a need to study engagement from a person-in-context research orientation (Sinatra, Heddy, & Lombardi, 2015; Schmidt, Rosenberg, & Beymer, 2018; Xie, Greene, & Heddy, 2019). That is, research which examines how engagement is situationally tied to the interaction between personal and contextual characteristics will allow for a more in-depth understanding of this important learning variable. Engagement exists at both the person-level grain size in which students are typically engaged in academic activities, as well as the occasion-level grain size as influenced by context. Arguably, engagement at this specific, occasion-level grain size is more proximal to the learning process and is where teachers' practices are most influential (Schmidt, Rosenbery, & Beymer, 2018). But, while research in this area has increased in recent years, there is still much less person-in-context engagement research compared to person-oriented and context-oriented studies. The lack in this type of research is related to the difficulty of collecting intensive longitudinal data combined with the difficulty of statistically representing such data.

Our purpose is to illustrate how the ESM-mobile approach combined with multilevel LST structural equation modeling can address these difficulties, and engagement is the perfect construct to examine through this method. In the following section, we will describe the data collection procedures as facilitated by *ESM-Mobile* and multilevel LST modeling which illustrate the following:

1. *Decomposition of trait-level and state-level engagement*: How much variance in engagement is at the trait-level vs. state-level?
2. *Group differences in variance decomposition:* Does the variance decomposition of engagement differ for students with low vs. high intrinsic motivation?
3. *Predicting state component of engagement*: To what extent does students' occasion-specific workload predict state-level engagement?

4 Methods

4.1 Participants and General Procedures

The participants for this study were 57 undergraduate students in a large public university in a mid-western state. Students were recruited from undergraduate courses in learning and motivation strategies in Fall 2019 and thus were from various colleges across campus. At an initial meeting at the beginning of the semester, the researchers met with potential participants to help students download and learn how to use the *ESM-Mobile* app on their smartphones. Students also took pre-surveys on their initial level of intrinsic motivation for attending college. Two weeks in the middle and two weeks at the end of the semester were designated as ESM data collection weeks. These time points were selected because they coincide with midterm and final exam schedules, representing likely times that students will study for their courses. As an incentive, participants were eligible to win 1 of 15 $50 Amazon gift cards.

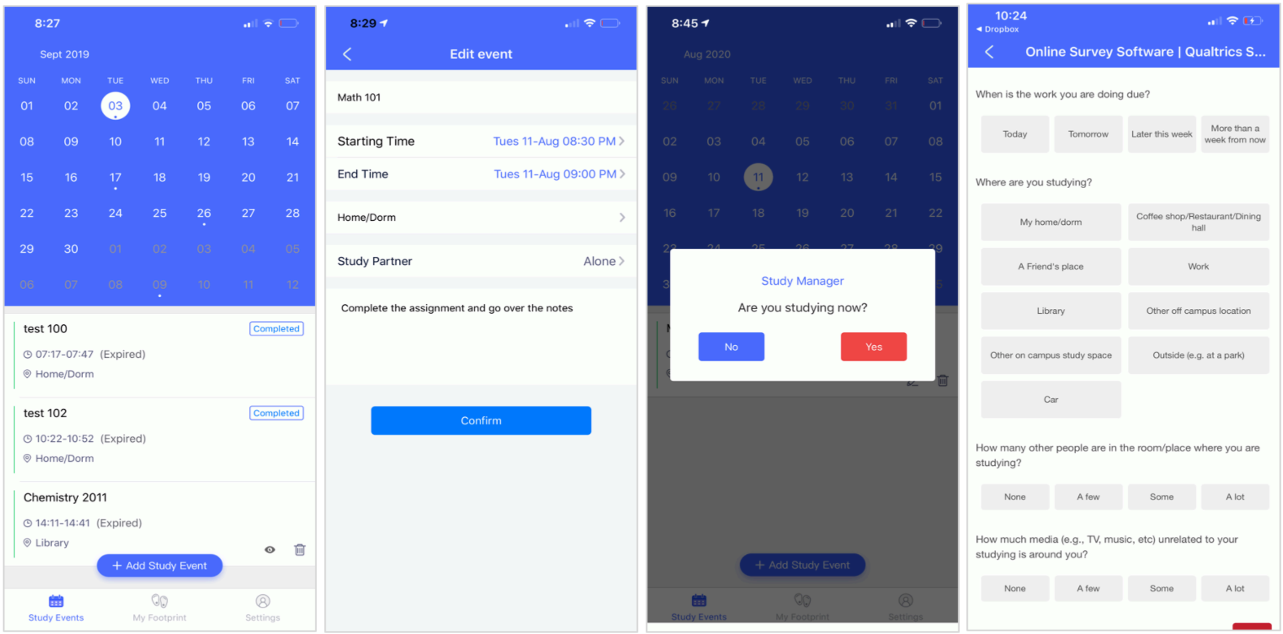

Figure 1: ESM-Mobile Screenshots.

4.2 ESM Procedures

The ESM-Mobile app was specifically built to deliver experience sampling surveys and to examine students' out-of-classroom engagement (see Figure 1; More details on Apple App Store: https://apps.apple.com/us/app/study-scheduler/id1473217961). *ESM-Mobile* runs on both iOS and android platforms, and is designed for mobile devices such as smartphones and tablets. ESM-Mobile was built with an event-based sampling approach; Participants plan study events in the app's calendar interface and the app prompts them during those times. The ESM prompt first asks if the student is studying as planned. If so, the ESM app directs students to provide details of their engagement. The app collects participants' responses linked with in-the-moment contextual information, collected through both the app (e.g., time, location) as well as via self-report by students. For this study, ESM surveys included items measuring engagement and workload (as a type of student-reported contextual information).

In total, 802 events were scheduled, 368 prompts received no response, 35 prompts received an initial response that the student was not studying as planned, and 399 prompts reached students while they were studying. Out of 399 accurately sampled prompts, 327 ESM surveys were completed (82%). The average number of prompts completed was about 5 per participant, with a median of 4 and a range of 20.

4.3 Measures

The measures used for this study were collected both via a pre-survey delivered electronically via Qualtrics and ESM surveys delivered through the ESM-Mobile app.

Intrinsic Motivation. Students responded to four pre-survey items measuring their intrinsic motivation for academic activities from Ratelle et al. (2007). Students use a 6-point Likert scale (1=strongly disagree to 6= strongly agree) to indicate their level of agreement to statements answering the stem question: Why are you in college? A sample item is: Because I experience pleasure and satisfaction while learning new things. The Cronbach's alpha for the scale in this study was 0.88.

Situational Engagement. In-the-moment engagement was measured via ESM-Mobile using three items, one (eng1) assessing behavioral engagement from Schmidt, Rosenberg, & Beymer (2018) and two (eng2 and eng3) assessing cognitive engagement from Greene (2015). The items are:
- eng1: How well are you concentrating on the study task?
- eng2: I use my related prior knowledge to help me learn while studying.
- eng3: I am aware of what material I did or did not understand.

Table 2: Descriptive Statistics.

	Mean	SD	Pearson's Correlations					
			1.	2.	3.	4.	5.	6.
1. Engage1 (eng1)	4.24	1.29	1.00					
2. Engage2 (eng2)	4.36	1.19	0.48 [0.40,0.56]	1.00				
3. Engage3 (eng3)	4.50	1.07	0.41 [0.32,0.50]	0.56 [0,48,0.63]	1.00			
4. Academic Workload	4.31	1.36	0.25 [0.14,0.34]	0.23 [0.12,0.33]	0.24 [0.13,0.34]	1.00		
5. Personal Workload	3.58	1.53	0.06 [-0.05,0.16]	0.00 [-0.11,0.11]	0.01 [-0.10,0.12]	0.28 [0.17,0.37]	1.00	
6. Intrinsic Motivation‡	4.80	0.76	0.31 [0.21,0.41]	0.17 [0.06,0.27]	0.19 [0.08,0.29]	0.25 [0.15,0.35]	0.22 [0.12,0.32]	1.00

Notes. For Pearson's correlations, 95% confidence intervals are provided within brackets.
Engagement items, academic workload, and personal workload were collected at the momentary-level via ESM-Mobile. Intrinsic motivation was collected during a pre-survey facilitated by Qualtrics.
‡ Correlations of momentary variables with intrinsic motivation were at the person-level, where the mean of each variable was calculated and then assessed for its correlation with intrinsic motivation.

Students report their agreement to these statements using a 6-point Likert scale. The Cronbach's alpha for the scale was 0.73.

Workload. One contextual information collected for this study via ESM-Mobile is students' academic and personal workload. For academic workload, students were asked: Which of the following statements best describes your academic workload for this week? For personal workload: Which of the following statements best describes your personal/family commitment for this week? Response categories range from 1=very light, 2 =light, 3 =average, 4=heavy, to 5=very heavy.

4.4 Decomposition of Trait and State Engagement

To examine the variance decomposition of engagement at the trait vs. state level, we use three situational engagement indicators collected via ESM-Mobile. The mean, standard deviation, and Pearson's correlations among all measures in the study are presented in Table 2. Repeated measures, within-person correlations calculated based on Bland and Altman (1995) are presented in Table 3. Pearson's correlations showed that all situational engagement items are all positively and significantly correlated with one another and with situational academic workload. Situational personal workload is not significantly correlated with engagement items, but is positively correlated with academic workload. Intrinsic motivation is positively and significantly correlated with the mean score of all other indices. Within-person correlations also showed positive and significant relationships between engage2 with engage1 and engage3.

We modeled situational engagement and decomposed trait vs. state components with both multilevel STMS and MTMS models using Mplus version 8 (annotated code files can be found in the Appendix). Fit indices showed that the multilevel STMS model did not fit the data well, with $\chi^2(6)$=72.176, $p(\chi^2)$=0.0000, AIC=2739.518, BIC=2784.997, RMSEA=0.184, CFI=0.529, TLI=0.294, Level 1 SRMR=0.110, Level 2 SRMR=0.158. Specifically, RMSEA is well above the cutoff criteria for good fit of 0.06-0.08 (Hu & Bentler, 1999; Kline, 2011). The chi-square p-value is significant indicating rejection of the null hypothesis that there are no discrepancies between the population and the model covariance (Barrett, 2007; Kline, 2011). This result may be because the STMS model assumption of a common trait factor shared by all indicators may be too restrictive, especially when the indicators covered both behavioral and cognitive aspects of engagement.

Table 3: Within-Person Correlations.

	1.	2.	3.	4.	5.
1. Engage1 (eng1)	1.00				
2. Engage2 (eng2)	0.32 [0.21,0.43]	1.00			
3. Engage3 (eng3)	0.10 [-0.02,0.22]	0.24 [0.13,0.35]	1.00		
4. Academic Workload	0.10 [-0.02,0.22]	0.03 [-0.09,0.15]	0.07 [-0.05,0.19]	1.00	
5. Personal Workload	-0.02 [-0.13,0.11]	-0.00 [-0.12,0.12]	0.01 [-0.11,0.13]	0.01 [-0.11,0.13]	1.00

Notes. 95% confidence intervals are provided within brackets.
Engagement items, academic workload, and personal workload were collected at the momentary-level via ESM-Mobile.

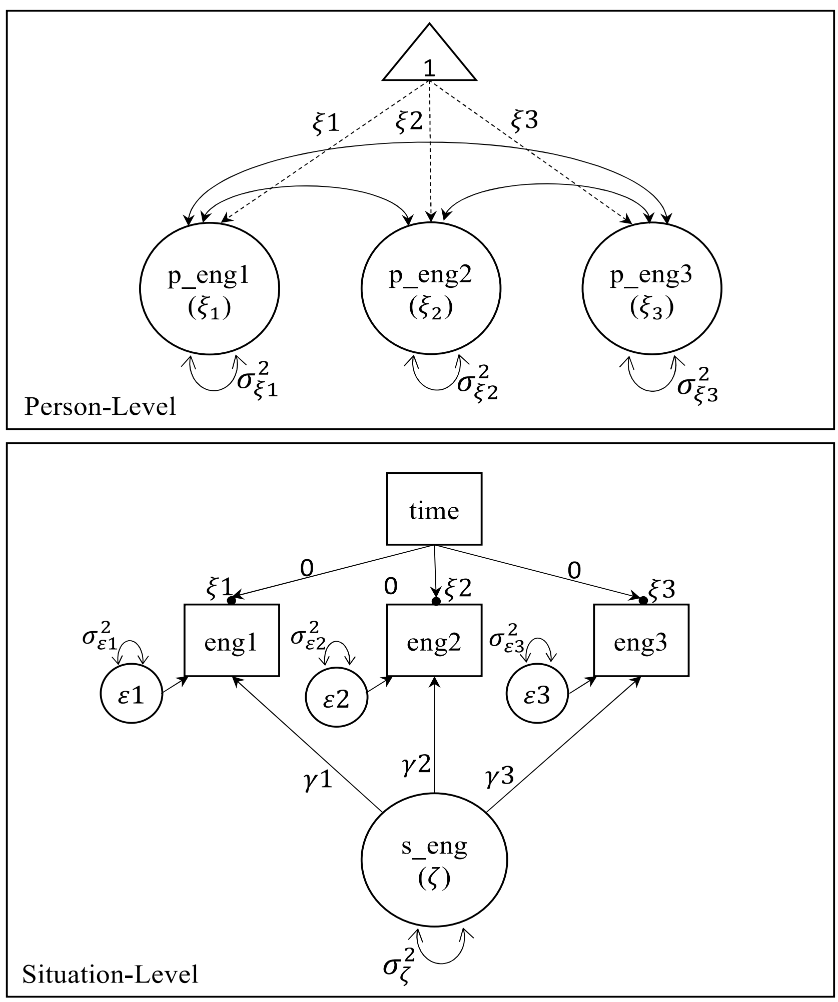

Figure 2: Multilevel MTMS Model of Engagement.

Table 4: Parameter estimates for the multilevel MTMS model of engagement.

Label	Parameter	Unstandardized		Standardized	
		Estimate	SE	Estimate	SE
Level 1 – Within Occasions					
State loadings	$\gamma 1$	1.000	-	0.507	0.094
	$\gamma 2$	1.197	0.391	0.668	0.112
	$\gamma 3$	0.598	0.162	0.369	0.079
State variance	$var(\zeta)$	0.240	0.094	1.000	0.000
Error variances	$var(\varepsilon 1)$	0.693	0.097	0.743	0.095
	$var(\varepsilon 2)$	0.426	0.116	0.553	0.150
	$var(\varepsilon 3)$	0.546	0.054	0.058	0.058
Level 2 – Between Occasions					
Trait means	$\xi 1$	4.263	0.129	5.117	0.644
	$\xi 2$	4.449	0.122	5.524	0.672
	$\xi 3$	4.579	0.112	6.216	0.755
Trait variances	$var(\xi 1)$	0.694	0.171	1.000	0.000
	$var(\xi 2)$	0.649	0.155	1.000	0.000
	$var(\xi 3)$	0.543	0.130	1.000	0.000
Trait covariances	$\xi 1,\xi 2$	0.502	0.139	0.748	0.081
& correlations	$\xi 1,\xi 3$	0.454	0.125	0.740	0.090
	$\xi 2,\xi 3$	0.560	0.130	0.944	0.034

The multilevel MTMS model which instead specified indicator-specific trait factors (Figure 2) was shown to fit the data much better, with χ^2 (3)=2.364, p(χ^2)=0.5004, AIC=2675.705, BIC=2732.554, RMSEA=0.000, CFI=1.000, TLI=1.000, Level 1 SRMR=0.036, Level 2 SRMR=0.011. The p-value associated with the chi-square becomes non-significant, and both SRMR at levels 1 and 2 were lower indicating smaller sizes of standardized residuals. We thus proceeded with the multilevel MTMS model for interpretation and built on this model for the remaining analysis.

Results show that the trait factor means across indicators range from 4.263 for eng1 to 4.579 for eng3, which is slightly above the median score with the range being 1 to 6 (Table 4). The indicator-specific trait correlations range from cor(ξ1,ξ3)=0.740 to cor(ξ2,ξ3)=0.944. Notably, the correlations between the two cognitive engagement items (eng2 & eng3) were much higher than between these two items and the behavioral engagement item (eng1). The result of some trait correlations being much less than 1 corroborated the result that the STMS model fit the data poorly. Mplus also outputted the intraclass correlations (ICCs), which are 0.427 for eng1, 0.458 for eng2, and 0.470 for eng3. The ICCs represents the proportion of variability in the observed variable that is due to differences across individuals, or how “trait-like” engagement is (also referred to as the indicator consistency coefficient; Geiser et al., 2013; Geiser, 2021). The results showed that about 43 to 47% of the variance in measured engagement can be attributed to trait-level differences.

Additionally, we can calculate the amount of the observed indicator variance attributable to situation-specific influences (or indicator occasion-specificity coefficient; Geiser et al., 2013; Geiser, 2021), using the following equation:

$$\text{Occasion-specificity } (Y_i) = \frac{\gamma_i^2 var(\zeta)}{var(\xi_i)+\gamma_i^2 var(\zeta)+var(\varepsilon_i)}$$

For example, the occasion-specificity for eng2 is:

$$\text{Occasion-specificity (eng2)} = \frac{1.197^2 * 0.240}{0.649 + 1.197^2 * 0.240 + 0.426} = 0.242$$

Table 5: Variance Decomposition Coefficients for the MTMS model of engagement.

Indicator	Consistency (ICC)	Occasion-Specificity	Reliability
eng1	0.427	0.148	0.574
eng2	0.458	0.242	0.700
eng3	0.470	0.073	0.535

Finally, the indicator reliability coefficient can be calculated to discern the amount of the indicator variance that is reliably measured by the indicator. This can also be thought of as the amount of reliable variance after accounting for measurement error. The indicator reliability coefficient can be calculated using the following equation:

$$\text{Reliability } (Y_i) = \frac{var(\xi_i)+\gamma_i^2 var(\zeta)}{var(\xi_i)+\gamma_i^2 var(\zeta)+var(\varepsilon_i)}$$

For example, the reliability coefficient for eng2 is:

$$\text{Reliability (eng2)} = \frac{0.649+1.197^2 * 0.240}{0.649+ 1.197^2 * 0.240+ 0.426} = 0.700$$

With the indicator consistency, occasion-specific, and reliability coefficients thus derived, we display these parameters in Table 5.

Results show that the variance in engagement can be attributed to both trait- and state-level differences. Higher consistency coefficients compared to occasion-specificity coefficients for all three items showed that a relatively higher amount of variance can be attributed to differences across students rather than intraindividual differences. In particular, the state-level variance for eng3 (I am aware of what material I did or did not understand) only attributed to about 7.3% of the observed total variance. The extent to which students use cognitive strategies to monitor their understanding may be especially due to person-level or trait-like tendency that do not vary much across context. Of note, the state-level standardized loading was also low for this item (γ_3=0.300) while its trait correlations with eng1 and eng3 are high at 0.737 and 0.950, respectively. Typically, low standardized loadings may be grounds for the removal of the item from the measurement model (Kline, 2011). However, in Latent State-Trait models, the extent to which an item contributes to the measurement of the latent construct has to do with its functioning both at the state and trait levels. In this MTMS model, it is clear that although eng3 is a more distal contributor to the latent state factor, it has a close relationship with other engagement indicators at the trait level. This trend is consistent with its low occasion-specificity coefficient showing low variance at the state level. Such an example demonstrates that while a battery of indicators may represent the underlying characteristics of engagement, the extent to which each indicator varies at the trait- versus state- level differs and should be subject for further research. Some indicators – such as eng3 in this study – may not vary much at the state level and thus contribute less to the definition of the state latent factor, but may have strong ties with other indicators at the trait level.

At the same time, some indicators showed substantial variation across situations. For example, 24.2% of the variance in eng2 (I use my related prior knowledge to help me learn while studying) can be attributed to state-level differences. The extent to which student use related prior knowledge may be due to contextual features such as the teachers' stated directions or the learning task at hand.

4.5 Group Differences in Variance Decomposition

We further illustrate how the MTMS model can answer additional research questions related to group differences in trait- vs. state-level variance decomposition. For this purpose, we chose a motivation variable – intrinsic motivation – as a type of student-level characteristic that may differentiate the variance decomposition of engagement. Intrinsic motivation specifies the extent to which students' engagement in academic activities is due to self-determined reasons such as their own interest and enjoyment (Deci & Ryan, 2008). Intrinsic motivation has been shown to be intricately linked with students' investment and engagement in academic activities (Authors, 2020; Authors, 2021; Deci & Ryan, 2000). In addition to influencing levels of engagement, we hypothesize that students who are lowly vs. highly intrinsically motivated may differ in the extent to which their engagement is stable across time. Intrinsic motivation scores were collected at the beginning of the semester and students were separated into low vs. high intrinsic motivation groups using a mean-split of average scores across items. In total, 28 students were designated as the low intrinsic motivation group and another 29 as the high intrinsic motivation group.

A similar MTMS model was specified allowing all parameters to vary between groups. This model fit the data well, with χ^2 (6)=3.106, p(χ^2)=0.7955, AIC=2663.981, BIC=2777.680, RMSEA=0.000, CFI=1.000, TLI=1.000, Level 1 SRMR=0.017, Level 2 SRMR=0.026. Multiple group SEM analysis showed that the model with freely varying parameters fit significantly better than one constraining parameters to be equal across groups ($\Delta\chi^2$ =41.723, Δdf=15, p<0.001). MTMS parameters by groups is presented in Table 6 and the variance decomposition coefficients in Table 7.

As expected, results showed that students in the high intrinsic motivation group had higher trait factor means across indicators. For example, high intrinsically motivated students had an average trait factor mean of 4.499 for eng1 compared to 4.026 for the low intrinsic motivation group. On the other hand, indicator-specific trait correlations were higher for the low motivation group (0.767 to 0.988) compared to the high motivation group (0.649 to 0.894). One interpretation of this result is that lowly motivated students' indices of engagement tend to occur together. For example, when one engagement indicator is low, others tend to be low as well. In contrast, engagement indicators covary to a lesser extent for highly motivated students. There may be times when one engagement indicator is low and another is high. It may be the case that highly motivated students are more selective in the way they engage with the learning materials and may not always display all indices of engagement together.

Variance decomposition coefficients showed that students with high intrinsic motivation had a slightly higher consistency coefficient for eng1 (behavioral engagement indicator) compared to the low motivation group. This means that for this group more of the variance in behavioral engagement can be attributed to trait-level differences. In contrast, the high intrinsic motivation group had lower consistency coefficients for eng2 and eng3 (cognitive engagement indicators). Less of the variance in cognitive engagement resides at the trait-level for highly motivated students. Interestingly, the occasion-specificity coefficients for cognitive engagement items were also low for this group. Relatedly, the indicators for cognitive engagement were less reliably measured for students with high intrinsic motivation. One interpretation of this result is students with high intrinsic motivation may engage in a variety of cognitive engagement strategies not captured by these indicators (Greene, 2015) and select different cognitive strategies for different learning tasks. Additional research which further examines engagement patterns across differently motivated students within various learning contexts is needed to confirm and explain these trends.

4.6 Predicting Situation-Specific Component of Engagement

Another type of valuable research is to understand how contextual factors influence the state-level component of engagement. That is, partialling out trait-level differences in how students typically engage across tasks, what factors in the learning environment help students become more situationally engaged in a learning task. To illustrate how this can be done using ESM and mutilevel-MTMS modeling, we measured students' personal and academic workload as a contextual factor that may influence the state component of engagement. Recall that workload was measured in-situ as a type of student-reported contextual factor.

We began with the MTMS model of engagement similar to that for the last research questions. Personal and academic workload were specified as observed predictors (ß1 and ß2 [betas]) of the common latent state factor (s_eng or ζ, Figure 2). Results show that academic workload positively and significantly predicts the state-level engagement (Table 8). That is, the higher academic workload reported by students, the higher their state-level situational engagement was. This result was in line with previous research, which showed that proximity to a deadline was positively associated with increased cognitive engagement (Xie, Heddy, & Vongkulluksn, 2019). As students perceive that there is more academic work to be done, they may be more apt to concentrate on learning tasks and use cognitive strategies for learning. On the other hand, personal workload – or the level of non-academic commitment – was not associated with state-level engagement. This type of predictive modeling presents a novel way of conceptualizing the influence of contextual factors on learning engagement. Rather than positing that contextual factors like workload may influence overall engagement, multilevel LST models are able to discern their effects on just the state-level component of engagement.

5 Implications and Limitations

This paper demonstrates the ways in which ESM data collection combined with multilevel LST structural equation modeling – such as STMS and MTMS models – can help facilitate our understanding of how learning processes occur at a situational level. Such research represents a person-in-context orientation, which disentangles the interaction between student characteristics and contextual characteristics in their influence on learning outcomes

Table 6: Parameter estimates for the multilevel MTMS model of engagement across student groups.

Label	Parameter	Low Intrinsic Motivation		High Intrinsic Motivation	
		Estimate (SE)	**Std.Estimate (SE)**	**Estimate (SE)**	**Std.Estimate (SE)**
Level 1 – Within Occasions					
State loadings	$\gamma 1$	1.000	0.511 (0.086)	1.000	0.546 (0.462)
	$\gamma 2$	1.298 (0.332)	0.781 (0.095)	0.897 (1.501)	0.505 (0.427)
	$\gamma 3$	0.865 (0.196)	0.600 (0.087)	0.198 (0.289)	0.118 (0.133)
State variance	$var(\zeta)$	0.255 (0.098)	1.000 (0.000)	0.265 (0.451)	1.000 (0.000)
Error variances	$var(\varepsilon 1)$	0.723 (0.103)	0.739 (0.087)	0.624 (0.452)	0.702 (0.505)
	$var(\varepsilon 2)$	0.274 (0.103)	0.390 (0.149)	0.624 (0.364)	0.745 (0.431)
	$var(\varepsilon 3)$	0.338 (0.059)	0.640 (0.104)	0.737 (0.093)	0.986 (0.031)
Level 2 – Between Occasions					
Trait means	$\xi 1$	4.026 (0.178)	4.957 (0.868)	4.499 (0.178)	5.631 (1.073)
	$\xi 2$	4.324 (0.182)	4.954 (0.833)	4.581 (0.159)	6.590 (1.216)
	$\xi 3$	4.472 (0.164)	5.683 (0.916)	4.652 (0.152)	6.946 (1.328)
Trait variances	$var(\xi 1)$	0.660 (0.224)	1.000 (0.000)	0.638 (0.238)	1.000 (0.000)
	$var(\xi 2)$	0.762 (0.246)	1.000 (0.000)	0.483 (0.179)	1.000 (0.000)
	$var(\xi 3)$	0.619 (0.194)	1.000 (0.000)	0.449 (0.172)	1.000 (0.000)
Trait covariances & correlations	$\xi 1,\xi 2$	0.598 (0.210)	0.843 (0.083)	0.360 (0.168)	0.649 (0.157)
	$\xi 1,\xi 3$	0.490 (0.180)	0.767 (0.107)	0.418 (0.167)	0.782 (0.158)
	$\xi 2,\xi 3$	0.678 (0.209)	0.988 (0.026)	0.416 (0.151)	0.894 (0.088)

(Sinatra, Heddy, & Lombardi, 2015). The key advantage of ESM is its ability to capture motivation and cognition processes in situ, linked with contextual data that can be collected as it is happening. This data collection method avoids biases that may stem from prospective speculation or retrospective recollection. Multilevel LST modeling takes advantage of such situationally authentic, intensive longitudinal data by allowing researchers to flexibly decompose learning-related variables into trait-level and state-level components. Synergistically, ESM combined with multilevel LST modeling can answer a wider range of research questions, especially those focused on the relationship among learning-related variables within particular contexts in which learning takes place. This approach is particularly useful for examining unstructured learning environments, where students may engage in relevant tasks at inconsistent times and/or within shifting contexts.

Table 7: Variance Decomposition Coefficients for the MTMS model of engagement.

Indicator	Low Intrinsic Motivation			High Intrinsic Motivation		
	Consistency	Occasion-Specificity	Reliability	Consistency	Occasion-Specificity	Reliability
eng1	0.403	0.156	0.559	0.418	0.174	0.591
eng2	0.520	0.293	0.813	0.366	0.162	0.527
eng3	0.539	0.166	0.706	0.375	0.009	0.384

Table 8: Parameter estimates for the multilevel MTMS model of engagement with situation-specific work load.

Label	Parameter	Unstandardized		Standardized	
		Estimate	SE	Estimate	SE
Level 1 – Within Occasions					
State loadings	$\gamma 1$	1.00	-	0.531	0.085
	$\gamma 2$	1.109	0.302	0.649	0.092
	$\gamma 3$	0.616	0.163	0.399	0.078
Academic load coefficient	ß1	0.089	0.044	0.233	0.100
Personal load coefficient	ß2	-0.008	0.033	-0.022	0.098
State variance	$var(\zeta)$	0.253	0.087	0.948	0.045
Error variances	$var(\varepsilon 1)$	0.678	0.091	0.718	0.090
	$var(\varepsilon 2)$	0.451	0.093	0.579	0.119
	$var(\varepsilon 3)$	0.535	0.053	0.841	0.062
Level 2 – Between Occasions					
Trait means	$\xi 1$	3.914	0.245	4.896	0.667
	$\xi 2$	4.068	0.245	5.365	0.702
	$\xi 3$	4.362	0.173	6.061	0.756
Trait variances	$var(\xi 1)$	0.639	0.162	1.000	0.000
	$var(\xi 2)$	0.575	0.143	1.000	0.000
	$var(\xi 3)$	0.518	0.127	1.000	0.000
Trait covariances &	$\xi 1,\xi 2$	0.438	0.129	0.723	0.089
correlations	$\xi 1,\xi 3$	0.415	0.120	0.721	0.097
	$\xi 2,\xi 3$	0.512	0.123	0.938	0.037

Related to our data collection strategy, one potential concern for collecting data via ESM is the possibility of selection bias related to the extent to which students self-schedule ESM data collection and respond to ESM prompts. The data collected for this study showed that at least 49.8% of the prompts reached students when they were engaging in relevant learning tasks (sampling accuracy) and students completed the ESM survey 40.8% out of the original 802 scheduled prompts. There is a concern about whether the engagement data collected represent a selected group of students, although the direction of bias is unclear. For example, students who are typically self-regulated and highly engaged may be more likely to schedule study events in advance. At the same time, students who are highly engaged in a learning task may not choose to interrupt their studies and respond to an ESM survey. In our study, we take several measures to boost response rates and thereby reduce selection effects. First, we chose only on few weeks within the semester to collect data, rather than a more prolonged data collection. This helps reduce the fatigue students may feel towards scheduling and answering ESM surveys. Second, we integrate much of our data collection efforts within particular classrooms. The researchers visited students' class as well as asked instructors to relay reminders for students to schedule study events in the ESM mobile app. We note that there will inevitably be some selection bias in most survey research. However, we counterbalance this concern with the situationally-authentic data we were able to collect.

Related to our analytic strategy, we caution that STMS and MTMS models as specified in this paper do not take into account how trait-level constructs may change. That is, our model assumes that the interindividual differences in engagement do not change over the course of our data collection period; students' trait-level tendency to engage in academic tasks over time were thought to remain constant. This assumption is reasonable in our research context which spanned about half an academic semester, but may not be tenable for longitudinal studies which cover longer periods. For instance, over a period of a year or two, students may be expected to have a change in their engagement tendencies. Some students may become more focused in their studies and will typically report higher engagement across learning tasks. Others may report declining engagement tendencies. If such trait-level changes are expected, models which account for these "growths" should be used. An example of such a model is the latent growth curve model, which can also handle multiple indicators of a certain trait and maybe specified in a multilevel context (Geiser, 2021).

Relatedly, the STMS and MTMS models illustrated here do not account for "spillover" effects at the state level (level 1). Motivational and cognitive constructs measured at one time point are thought to influence the same constructs measured at the next adjacent time point (Little, 2013). For example, the way in which students engage in a learning task at one time point may have spillover effects on their engagement measured at the next time point. This effect is thought to be more likely when measurement occasions are close together (Geiser, 2021). If a student is engaged in a certain manner one day, it is expected that s/he will have a similar engagement pattern if measured again during the same day or on the next day. For person-in-context research, we can also consider that when the same construct is measured within the same context, spillover effects may also be more likely. Consider a structured learning environment like a particular classroom. When entering the same learning space, students may experience more spillover effects on their motivation and cognition from a previous learning task. Students may tend to engage in a certain manner when learning in a specific class, with a certain teacher, and the same set of similar rules and structures (Schmidt et al., 2018). Students may also be expected to pick up where they left off when they return to the same workspace. In the context of this study, we measured learning engagement at relatively spaced-out intervals. On average, students in our study answered one to two ESM prompts per week. Additionally, the learning task examined was out-of-classroom studying, which was expected to occur in a variety of contexts (different learning tasks in different settings). Therefore, spillover effect from one measurement occasion to another was assumed to be minimal. However, we can explicitly test such effects with autoregressive paths in our SEM models, in which one construct measured at time t is assumed to be influenced by the same construct measured at time t-1. It should be noted that in order to examine trait-change or autoregressive effects, a constant number of time points and time intervals are needed across individuals. That is, the same restrictions would apply as when single-level LST models are specified (Table 1). There is a necessary trade-off between flexibility in modeling complex, intensive longitudinal data and the ability to model change through time.

Another aspect to consider is the number of level-1 (time points) and level-2 (students) sample size needed for unbiased parameter estimates. In multilevel modeling, it is generally recommended that the addition of level-2 units is more beneficial for reducing parameter bias and increasing power for detecting statistically significant effects at level 2 (Cheung & Au, 2005; Muthen, 1991). A simulation study by Hox and associates (2010) showed that the empirical coverage of 95% confidence intervals for level-2 loadings and variances were satisfactory (>0.90) for multilevel confirmatory factor models using the Maximum Likelihood (ML) estimator with 50 groups and 5 level-1 observations – characteristics which closely match the current study. However, scholars have pointed out that the question of adequate sample size is complex and is dependent on many model characteristics (Preacher et al., 2010). Therefore, future simulation studies may need to examine required level-1 and level-2 sample sizes for accurate parameter estimation with multilevel latent state-trait models. We also note that sample size considerations may be balanced with the focus of specific research questions – whether level-1 or level-2 effects are being examined. Given the estimation accuracy, differences in research context should also play a part in informing whether limited resources should be directed at recruiting more participants (level-2 units) or observing additional time points for each participant (level-1 units).

Finally, another disadvantage of multilevel STMS or MTMS models is it does not allow researchers to test measurement invariance assumptions. The multilevel STMS and MTMS model make a number of assumptions about the time invariance of intercepts, factor loadings and variances (i.e., strict measurement equivalence; Geiser, 2013). In single level LST models, these assumptions can be tested by examining explicitly whether these parameters differed from time 1 to time 2 to time 3, and so on. However, within a multilevel framework, testing measurement invariance assumption is not possible. The time invariance of parameters allows for more parsimonious modeling and makes the multilevel specification possible. This issue demonstrates the tradeoff between model parsimony and more rigorous assumptions. It should be noted that such an assumption is often made in longitudinal research. Nonetheless, multilevel LST modeling might best be used with well-validated measurements, especially ones which have been shown to measure a construct in a similar way across time.

While the aforementioned assumptions are stringent, there are many research contexts in which these assumptions are reasonable. On the other hand, the ESM approach combined with multilevel LST modeling provides the flexibility needed to conduct a variety of person-in-context research. Application of such methods would allow educational researchers to examine nuances in how learning processes occur in context, adding a much needed situational lens towards improving learning and instruction.

Conflict of Interest Statement: There is no disclosure to be made for potential conflicts of interest for this research.

Ethical Statement: The research reported herein involved human participants who provided informed consent. The Institutional Review Board approved our research procedures and we closely adhered to their guidelines for research with human subjects when conducting this study.

Acknowledgement: The study reported in this paper is based upon work in the project entitled, *"The Influence of Contextual Features on Learning Engagement in Out-of-Classroom Settings for Academically At-Risk College Students"*, funded by the Spencer Foundation (Grant#201900220). The grant organization had no involvement in the design, data collection, data analysis, writing, or decision on article submission. The conclusions and recommendations within this article do not necessarily reflect the views of the Spencer Foundation.

References

Archambault, I., Janosz, M., & Chouinard, R. (2012). Teacher beliefs as predictors of adolescents' cognitive engagement and achievement in mathematics. *Journal of Educational Research*, *105*(5), 319–328. https://doi.org/10.1080/00220671.2011.629694

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Bandura, A. (2001). Social Cognitive Theory : An Agentic Perspective. *Annu. Rev. Psychol.*, *21*, 331–343.

Bland, J. M., & Altman, D. G. (1995). Correlation between subjects. *BMJ*, *310*(March), 633.

Buhs, E. S., & Ladd, G. W. (2001). Peer rejection as an antecedent of young children's school adjustment: an examination of mediating processes. *Developmental Psychology*, *37*(4), 550–560. https://doi.org/10.1037/0012-1649.37.4.550

Cheung, M. W. L., & Au, K. (2005). Applications of multilevel structural equation modeling to cross-cultural research. *Structural Equation Modeling*, *12*(4), 598–619. https://doi.org/10.1207/s15328007sem1204_5

Crompton, H. (2013). Mobile learning: New approach, new theory. In Z. . Berge & L. Muilenburg (Eds.), *Handbook of mobile learning* (Vol. 1, pp. 47–58). New York: Routledge.

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology* (pp. 35–54). Springer.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural Equation Modeling and Regression: Guidelines for Research Practice. *Communications of the Association for Information Systems*, *4*(August). https://doi.org/10.17705/1cais.00407

Geiser, C. (2021). *Longitudinal Structural Equation Modeling with Mplus: A Latent State-Trait Perspective*. Guilford Press.

Geiser, C., Bishop, J., Lockhart, G., Shiffman, S., & Grenard, J. L. (2013). Analyzing latent state-trait and multiple-indicator latent growth curve models as multilevel structural equation models. *Frontiers in Psychology*, *4*(DEC), 1–23. https://doi.org/10.3389/fpsyg.2013.00975

Greene, B. A. (2015). Measuring Cognitive Engagement With Self-Report Scales: Reflections From Over 20 Years of Research. *Educational Psychologist*, *50*(1), 14–30. https://doi.org/10.1080/00461520.2014.989230

Heddy, B. C., & Sinatra, G. M. (2017). Transformative parents: Facilitating transformative experiences and interest with a parent involvement intervention. *Science Education*, *101*(5), 765–786. https://doi.org/10.1002/sce.21292

Hox, J. J., Maas, C. J. M., & Brinkhuis, M. J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, *64*(2), 157–170. https://doi.org/10.1111/j.1467-9574.2009.00445.x

Lavigne, G. L., & Vallerand, R. J. (2010). The dynamic processes of influence between contextual and situational motivation: A test of the hierarchical model in a science education setting. *Journal of Applied Social Psychology*, *40*(9), 2343–2359. https://doi.org/10.1111/j.1559-1816.2010.00661.x

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford press.

Muthen, B. (1991). Multilevel Factor Analysis of Class and Student Achievement Components Author (s): Bengt O . Muthén Source : Journal of Educational Measurement , Vol . 28 , No . 4 (Winter , 1991), pp . 338-354 Published by : National Council on Measurement in Educati. *Journal of Education Measurement*, *28*(4), 338–354.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233. https://doi.org/10.1037/a0020141

Rabe-Hesketh, S., & Skrondal, A. (2012). Multilevel and longitudinal modeling using Stata. 3rd. *College Station, Tex: Stata Press Publication*.

Ratelle, C. F., Guay, F., Vallerand, R. J., Larose, S., & Senécal, C. (2007). Autonomous, Controlled, and Amotivated Types of Academic Motivation: A Person-Oriented Analysis. *Journal of Educational Psychology*, *99*(4), 734–746. https://doi.org/10.1037/0022-0663.99.4.734

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). sage.

Schmidt, J. A., Rosenberg, J. M., & Beymer, P. N. (2018). A person-in-context approach to student engagement in science: Examining learning activities and choice. *Journal of Research in Science Teaching*, *55*(1), 19–43. https://doi.org/10.1002/tea.21409

Schmidt, J., Kackar-Cam, H., Strati, A., & Shumow, L. (2015). The Role of Challenge in Students' Engagement and Competence in High School Science Classrooms: Hispanic and Non-Hispanic Whites Compared. *NCSSS Journal*, *20*(1), 20–26.

Shumow, L., Schmidt, J. A., & Kackar, H. (2008). Reading In Class \& Out of Class: An Experience Sampling Method Study. *Middle Grades Research Journal*, *3*(3).

Shumow, L., Schmidt, J. A., & Zaleski, D. J. (2013). *Multiple Perspectives on Student Learning , Engagement , and Motivation in High School Biology Labs Published by : University of North Carolina Press Stable URL : https://www.jstor.org/stable/23351976 REFERENCES Linked references are available on JSTOR fo*. *96*(3), 232–252.

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The Challenges of Defining and Measuring Student Engagement in Science. *Educational Psychologist*, *50*(1). https://doi.org/10.1080/00461520.2014.1002924

Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*.

Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits—Revised. *Annual Review of Clinical Psychology*, *11*, 71–98.

Walker, C. O., Greene, B. A., & Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences*, *16*(1), 1–12. https://doi.org/10.1016/j.lindif.2005.06.004

Wang, M. Te, & Eccles, J. S. (2012). Social Support Matters: Longitudinal Effects of Social Support on Three Dimensions of School Engagement From Middle to High School. *Child Development*, *83*(3), 877–895. https://doi.org/10.1111/j.1467-8624.2012.01745.x

Wang, M. Te, Fredricks, J. A., Ye, F., Hofkens, T. L., & Linn, J. S. (2016). The Math and Science Engagement Scales: Scale development, Validation, And psychometric properties. *Learning and Instruction*, *43*, 16–26. https://doi.org/10.1016/j.learninstruc.2016.01.008

Xie, K., Heddy, B. C., & Greene, B. A. (2019). Affordances of using mobile technology to support experience-sampling method in examining college students' engagement. *Computers & Education*, *128*, 183-198.

Xie, K., Heddy, B. C., & Vongkulluksn, V. W. (2019). Examining engagement in context using experience-sampling method with mobile technology. *Contemporary Educational Psychology*, *59*, 101788.

Xie, K., Hensley, L. C., Law, V., & Sun, Z. (2019). Self-regulation as a function of perceived leadership and cohesion in small group online collaborative learning. *British Journal of Educational Technology*, *50*(1), 456-468.

Xie, K., Vongkulluksn, V. W., Lu, L., & Cheng, S. L. (2020). A person-centered approach to examining high-school students' motivation, engagement and academic performance. *Contemporary Educational Psychology*, *62*, 101877.

Zhu, X., Chen, A., Ennis, C., Sun, H., Hopple, C., Bonello, M., Bae, M., & Kim, S. (2009). Situational interest, cognitive engagement, and achievement in physical education. *Contemporary Educational Psychology*, *34*(3), 221–229. https://doi.org/10.1016/j.cedpsych.2009.05.002

Zirkel, S., Garcia, J. A., & Murphy, M. C. (2015). Experience-Sampling Research Methods and Their Potential for Education Research. *Educational Researcher*, *44*(1), 7–16. https://doi.org/10.3102/0013189X14566879

Appendix A

Multilevel Multitrait-Multistate (MTMS) Model Mplus Code

```
Title: Multilevel Multitrait-Multistate Model
Data:
  File is [file path];
Variable:
  Names are
    user_id eng1 eng2 eng3 time;          ! Naming Variables
  Missing are all (-9999);
  usevar = user_id eng1 eng2 eng3 time;
  cluster =user_id;                       ! Specifying cluster variable (person-level ID)
  within =time;                           ! Specifying time as level 1 variable
Analysis: type = twolevel;
estimator = ML;
model:
%within%                                  !Level-1 model specification below this line
zeta by eng1@1 eng2 eng3;                 ! Latent state factor specified by indicators
eng1 eng2 eng3 on time@0;                 ! Regression of indicators on time, set at zero
eng1-eng3*;                               ! Estimating error variances
%between%                                 ! Level-2 model specification below this line
ksi1 by eng1@1;                           ! Indicator-specific trait factor, loadings set at 1
ksi2 by eng2@1;
ksi3 by eng3@1;
[eng1-eng3@0];                            ! Intercepts set to zero
[ksi1-ksi3*];                             ! Freely estimated trait factor means for each indicator
ksi1-ksi3 with ksi1-ksi3*;                ! Freely estimated trait factor covariances
eng1-eng3@0;                              ! Level-2 residual variances set to zero
output: sampstat stdyx;                   ! Output sample statistics and standardized parameters
```

Appendix B

MTMS Model with Parameters Constrained Across Groups Mplus Code

```
Title:  Multilevel Multitrait-Multistate Model with Constrained
Parameters
Data:
  File is [file path];
Variable:
  Names are
     user_id eng1 eng2 eng3 intgrp time;              ! Naming Variables
  Missing are all (-9999) ;
usevar = user_id eng1 eng2 eng3 intgrp time;
  cluster =user_id;                                   ! Specifying cluster variable (person-level ID)
  within =time;                                       ! Specifying time as level 1 variable
  grouping is intgrp (0=low, 1=high);                 ! Specifying grouping variable
Analysis: type = twolevel;
estimator = ML;
model:
%within%                                              !Level-1 model specification below this line
zeta by eng1@1 eng2 eng3;                             ! Latent state factor specified by indicators
eng1 eng2 eng3 on time@0;                             ! Regression of indicators on time, set at zero
eng1-eng3*;                                           ! Estimating error variances
%between%                                             ! Level-2 model specification below this line
ksi1 by eng1@1;                                       ! Indicator-specific trait factor, loadings set at 1
ksi2 by eng2@1;
ksi3 by eng3@1;
[eng1-eng3@0];                                        ! Intercepts set to zero
[ksi1-ksi3*]                                          ! Freely estimated trait factor means for each indicator
ksi1-ksi3 with ksi1-ksi3*;                            ! Freely estimated trait factor covariances
eng1-eng3@0;                                          ! Level-2 residual variances set to zero

Model low:                                            ! Model specific for intgp=0=low
%within%
zeta by eng1@1
 eng2 (1)                                             ! Constraining loadings of indicators on state factor
 eng3 (2);
eng1 eng2 eng3 on time@0;
eng1* (3);                                            ! Constraining error variances
eng2* (4);
```

```
eng3* (5);
zeta* (6);                                  ! Constraining latent state factor variance
%between%
ksi1 by eng1@1;
ksi2 by eng2@1;
ksi3 by eng3@1;
[eng1-eng3@0];
[ksi1*] (7);                                ! Constraining trait factor means of each indicator
[ksi2*] (8);
[ksi3*] (9);
ksi1 with ksi2 (10);                        ! Constraining trait factor covariances
ksi1 with ksi3 (11);
ksi2 with ksi3 (12);
ksi1* (13);                                 ! Constraining latent trait factor variances
ksi2* (14);
ksi3* (15);
eng1-eng3@0;

Model high:                                 ! Model specific for intgp=1=high
%within%
zeta by eng1@1
 eng2 (1)                                   ! Constraining loadings of indicators on state factor
 eng3 (2);
eng1 eng2 eng3 on time@0;
eng1* (3);                                  ! Constraining error variances
eng2* (4);
eng3* (5);
zeta* (6);                                  ! Constraining latent state factor variance
%between%
ksi1 by eng1@1;
ksi2 by eng2@1;
ksi3 by eng3@1;
[eng1-eng3@0];
[ksi1*] (7);                                ! Constraining trait factor means of each indicator
[ksi2*] (8);
[ksi3*] (9);
ksi1 with ksi2 (10);                        ! Constraining trait factor covariances
ksi1 with ksi3 (11);
ksi2 with ksi3 (12);
```

```
ksi1* (13);                              ! Constraining latent trait factor variances
ksi2* (14);
ksi3* (15);
eng1-eng3@0;

output: sampstat stdyx;                  ! Output sample statistics and standardized parameters
```

Appendix C

MTMS Model with Level-1 Predictor Mplus Code

```
Title: Multilevel Multitrait Multistate Model with Level-1
Predictor (Workload)
Data:
  File is [file path];
Variable:
  Names are
user_id eng1 eng2 eng3 acaload    perload time;   ! Naming Variables
  Missing are all (-9999);
usevar =  user_id eng1 eng2 eng3 acaload perload time;
  cluster =user_id;                      ! Specifying cluster variable (person-level ID)
  within =time acaload perload;          ! Specifying time and workload as level 1 variables
Analysis: type = twolevel;
estimator = ML;
model:
%within%                                 ! Level-1 model specification below this line
zeta by eng1@1 eng2 eng3;                ! Latent state factor specified by indicators
eng1 eng2 eng3 on time@0;                ! Regression of indicators on time, set at zero
eng1-eng3*;                              ! Estimating error variances
zeta on acaload;                         ! Regression of latent state factor on academic load
zeta on perload;                         ! Regression of latent state factor on personal load
%between%                                ! Level-2 model specification below this line
ksi1 by eng1@1;                          ! Indicator-specific trait factor, loadings set at 1
ksi2 by eng2@1;
ksi3 by eng3@1;
[eng1-eng3@0];                           ! Intercepts set to zero
[ksi1-ksi3*];                            ! Freely estimated trait factor means for each indicator
ksi1-ksi3 with ksi1-ksi3*;               ! Freely estimated trait factor covariances
eng1-eng3@0;                             ! Level-2 residual variances set to zero
output: sampstat stdyx;                  ! Output sample statistics and standardized parameters
```