DOI: https://doi.org/10.57125/FED/2023.06.25.02

How to cite: Chiu, M. S. (2023). Detected affect during online mathematics learning mediate self-report motivation changes: Examining a motivation-affect regulation model. *Futurity Education, 3*(2), 21–42. https://doi.org/10.57125/FED/2023.06.25.02

Detected Affect During Online Mathematics Learning Mediate Self-Report Motivation Changes: Examining a Motivation-Affect Regulation Model

Mei Shiu Chiu

Department of Education, National Chengchi University, 64, Zhinan Rd. Sec. 2, Taipei, 11605, Taiwan, http://orcid.org/0000-0002-2929-5151

Corresponding author: *chium@nccu.edu.tw.*

Received: March 7, 2023 | **Accepted:** April 25, 2023 | **Published:** June 25, 2023

Abstract: This study investigates relationships between detected affect during online learning and self-reports concerning motivational constructs in mathematics. Students reported their self-concept, interest, and values in mathematics before and after learning mathematics on an online intelligent tutoring system for one academic year over the autumn and spring semesters. Students' experiences of engaged concentration, confusion, frustration, and boredom were detected using log files from the system. The participants included over 8,000 2nd- to 8th-grade students in the United States who used an intelligent tutoring system, Reasoning Mind, during the academic year of 2017-2018. Path analyses with mediating effects indicate that students initially perceiving high self-concept are less frustrated in both autumn and spring. Students interested in mathematics experience initial disengagement and later confusion in online mathematical problem-solving. This result calls for cross-platform support from teachers and system designs, especially for students interested in traditional mathematical teaching and learning. For theory, this study posits the Motivation-Affect Regulation

(MAR) model, highlighting the mediating effects of motivation effect. Results find that the earlier self-concept can be mediated by fewer frustrations in online learning and lead to later self-concept. Interest triggers early disengagement and late confusion during online learning. Value is not mediated by affect. In methodology, structural equation modelling with mediating models offers opportunities to investigate more complex relationships when common measures are collected across time. Both offline pedagogical designs and online learning system designs need to consider the gaps between different formats of learning. Students with a high interest in learning mathematics in real classrooms may fail to adjust themselves easily to online mathematical problem- solving and experience disengagement in the early stages (one semester) and confusion in the stages (after one semester).

Keywords: affective states, intelligent tutoring system, mathematics learning; mediating effects, online learning.

Introduction

Traditionally, mathematics education researchers and educational psychologists use self-report questionnaires, scales, or (oral or written) text to study attitudes, beliefs, emotional responses, or motivation (Blikstein & Worsley, 2016; Gomez-Chacon, 2000; Hannula, 2002; McLeod, 1994; Pintrich, 2003). For topic-oriented methods, one of the most traditionally, perhaps also widely, used self-report tools to study affect is Likert-type scales for domain-specific learning (e.g., mathematics) (D'Mello et al., 2017). Multimodal data with learning analytics and personalised artefacts elect as suitable process-oriented methods (Blikstein & Worsley, 2016). Among the diverse multimodal methods, even though the self-report, the learner output, or physical measures are useful tools (Pachman et al., 2016), the advance of educational technology gives birth to real-time affect (automatically) detected by machine learning algorithms using online, behavioural log-file data, which appear to be the most efficient method. Besides, both self-report motivation and detected affect are related to achievement in theory (McLeod, 1994) and empirical studies (Marsh & Hau, 2004; Wang & Baker, 2018). The self-report and automatically detected measures may compensate each other. The self-report measures are more based on conscious expressions (e.g., stating "I like mathematics.") and detected measures are more based on unconscious physical behaviour (e.g., clicking on hints when failing to solve problems on an online learning system). This further evidences criterion validity for both the self-report affect and the detected affect.

This study is exploratory in nature and related to three fields: mathematics education, psychology, and educational technology. The term use of affect-related constructs appears to be different in these three fields. This study chooses to use 'motivation' to represent "domain-specific non-real-time affect regarding mathematical learning" (i.e., topic orientation) and 'affect' to represent "the real-time affect during mathematical problem-solving" (i.e., process orientation) (D'Mello & Mills, 2014). In terms of methodology, this study focuses on (offline Likert-type scale) "self-report motivation" and (online automatically) "detected affect."

Relationships between online behaviours and self-reports have been an emerging research field in learning analytics (van Halem et al., 2020). However, the past studies on longitudinal, interactive relationships between the self-report motivation and the detected affect appear to only explore their simple relationships (i.e., correlations) (a relevant study omitted for blind review, 2018). This study attempts to investigate their relationships by fully utilising a longitudinal, experimental design: the pre-test self-report motivation, the detected affect during online learning in the first and second semesters, and then the post-test self-report otivation (cf. Model 1 in Figure 2). This design offers an opportunity to investigate the longitudinal development of both self-report motivation and detected affect, and their interactions overtimes.

Research Problem

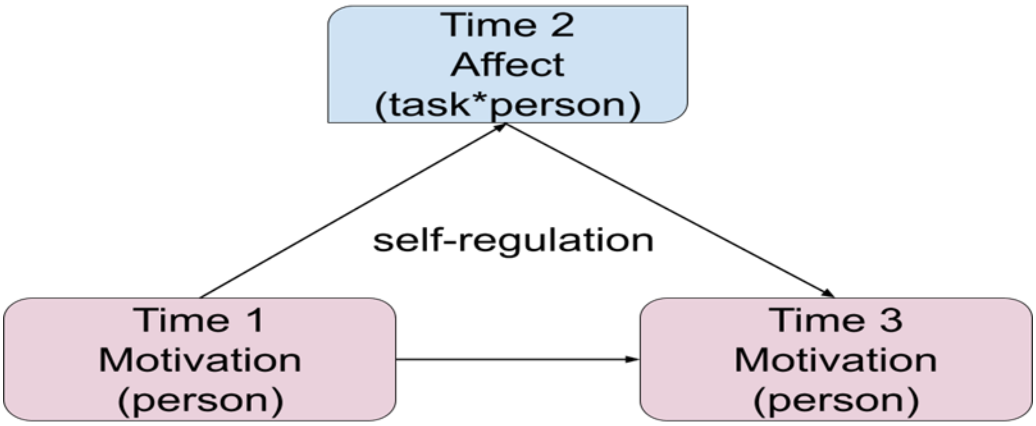
During learning, students may regulate their person-level motivation and situation-level affect (Efklides, 2011). The person level is the trait-like, general, or metacognitive knowledge (of person/self, tasks, and strategies), affect, and motivation. This is normally a top-down process. Motivation (at the person level) plays a role in affect during problem-solving situations (at the person*task level). Affect in turn predicts a later academic (e.g. language) performance (Gillet et al., 2013).

The task person level includes explicit knowledge, conscious analytic process, and unconscious process. All these directly relate to behaviour during hands-on, online task processing. This is a bottom-up process. Affect during problem-solving "updates" motivation (Baars et al., 2017) through the process of "feedback" (Efklides, 2011, p. 20). The use of "update" implies that affect at the task person level rewrites motivation at the person level after a timely experience related to motivation.

The bi-direction motivation effects and the mutual affect suggest a reciprocal relationship between motivations and affect over time. That is, motivation impacts affect and the affect updates motivation, which per se is a self-regulated learning (SRL) process. Figure 1 hypothesizes an SRL process as the reciprocal relationship between motivation (at the personal level) and affect (at the task personal level). This motivation-affect regulation (MAR) model serves as the theoretical basis of this study.

Figure 1

The Motivation-Affect Regulation (MAR) Model



Source: author own development

Research Focus

Diverse Constructs of Motivation. Different types of motivation such as self-efficacy or competence beliefs (self-concept), control beliefs, interests, values, and goals are identified in Educational and psychological researches (Pintrich, 2003). Among diverse types of motivation, the self-concept is mostly researched, perhaps due to its strong relationships with achievements (Marsh & Hau, 2004). The self-concept has (meta) cognitive, motivational, and affective characteristics in the "metacognitive and affective model of self-regulated learning" (the MASRL model) (Efklides, 2011).

The interest and value are part of motivation or affect in the MASRL model (Efklides, 2011). The two constructs are related to course or career choices (Cerinsek et al., 2013; Maltese & Tai, 2010).

Although the interest and value share some commonality, if distinguishing them deliberately, the interest is a major part of intrinsic motivation in response to the subject per se (e.g., mathematics and science) or the temporary situation (e.g., pleasant learning materials and game-based learning) (Høgheim & Reber, 2015). The value, on the other hand, focuses on 'extrinsic motivation', responding to social appraisals (e.g., reputable colleges and higher-paid jobs) (Cerasoli et al., 2014). Therefore, these three constructs are the focused self-report motivation in this study.

Longitudinal Relationships of Motivation. The knowledge of longitudinal relationships within self-report motivation may provide insight into its longitudinal relationship with detected affect. Researchers on self-report motivation often use structural equation modelling (SEM) algorithms to examine their longitudinal relationships. The longitudinal relationships are normally stronger for the same motivation measures lasting over time (i.e., parallel paths) than for different affective states (i.e., cross paths). For example, prior interest and self-concept predict their respective later ones (parallel paths betas = 0.51 to 0.57) better than they predict each other (cross- path betas = 0.04 to 0/12) (Marsh et al., 2005). This pattern also occurs for students' prior digital learning preference, wish for digital schoolwork, and schoolwork engagement to predict their later ones over three years of middle school (parallel paths betas = 0.48 to 0.56) (Hietajärvi et al., 2020). The cross-paths from digital learning preference of previous years (grades 7-8) to predict the next years school engagement (years 8-9) are much smaller (cross-path betas = both 0.15) or non-significant.

Diverse Constructs of Detected Affect. The detected affect slightly corresponds to the emotional constructs (e.g., frustration) of affect in mathematical problem-solving (McLeod, 1994). In order to promote adaptive teaching, intelligent tutoring systems aim using diverse data sources and machine learning algorithms to (automatically) detect learners' affect.

The diverse data sources include learners' dialogues with online pedagogical agents, body posture through pressure (for the back and seat) on a sensor, user behavioural log-file data, human judges, and synchronous observation (D'mello & Graesser, 2007; Ocumpaugh et al., 2015). The machine learning techniques (e.g., feature engineering and predictive algorithms) normally map the physical, text, and log- file data to human judgment or observations to define and finally automatically detect affect. The detected constructs include the affect (e.g., anger, engaged concentration, and surprise) or behaviours (e.g., being off-task and gaming the system) (Baker & Rossi, 2013). D'Mello's (2013) the meta-analysis on affect, learning, and technology reveals that the most researched affective constructs are engaged in the concentration/flow, boredom, confusion, and frustration; besides, engaged concentration/flow, boredom, confusion are the most frequently occurring affective constructs, which also the focused detected affective constructs of this study.

Longitudinal Relationships of Detected Affect. Few studies investigate pure longitudinal relationships of detected affect. The studies to date appear to focus on linking detected affect to specific online learning systems or problem-solving contexts. For example, boredom reduces slightly from early to later online learning stages in a personalised, self-paced system. Engaged concentration leads to confusion when cognitive disequilibrium occurs, which further may link to frustration and then boredom (if a failure occurs), found by self-report methodology for online learning (D'Mello & Graesser, 2012). Confusion and frustration rise while studying worked examples, prompting students to correct, compare and explain errors (Richey et al., 2019). This leads to better learning gains, although confusion and frustration are negatively related to pre-test, post-test, or delayed post-test achievements in mathematics.

Longitudinal Relationships between Self-Report Motivation and Detected Affect

Relatively few studies focus on relationships between self-report motivation and detected affect. Following a similar research design to this study (pre-test motivation survey, one-year online learning (with detected affect), and post-test motivation survey), correlation analysis finds that self-concept relates to student learning outcomes and behaviours (e.g., correctness, completion time, and actions) more than interest and value (Slater et al., 2018). Another study uses middle school students' detected affect in the ASSISTments system to relate to their high-school self-report motivation (Ocumpaugh et al., 2016).

Correlation analysis finds that STEM career self-efficacy positively relates to engaged concentration, negatively to confusion, and has no significant relationship with either boredom or frustration. STEM career interest has a negative correlation with confusion but non-significant correlations with engaged concentration, frustration, and boredom.

These studies appear to use correlational analysis only, failing to take into account time, which has been embedded in the data structure. This study aims to supplement this research gap by considering time as a structural issue examined by the SEM methodology.

Research Aim and Hypotheses

A special form of longitudinal relations is mediating effects. Mediating effects are popularly used in psychological research because mediating effects explain psychological phenomena focusing on processing (Chiu, 2022). Using the mediating effect methodology can partially assess the effectiveness of an intervention (e.g., the detected affect in online learning in this study). Three models, therefore, are formulated to examine the mediating effect. To explain using this study's measures and follow the conventional use of effect notations (Baron & Kenny 1986), a mediating effect is verified if meeting the following two conditions;

1.The later (post-test) motivation is impacted by its earlier (pre-test) motivation. In Figure 2, this is represented by the path c0 in Model 0.

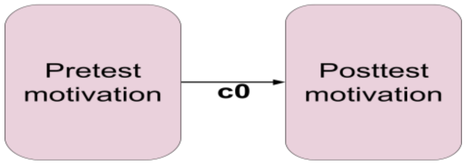
2.When adding a mediator (e.g., affect in Model 1 and Model 2 of Figure 2), the impact of an earlier motivation on its later one reduces (i.e., $c1 < c0$; $c2 < c0$).

Three popular algorithms to investigate mediating effects are regression (Shrout & Bolger, 2002), multilevel analysis (Kenny et al., 2003), and path analysis of SEM methodology (Preacher & Hayes, 2008). This study chose to use SEM because one single SEM analysis can obtain the three path parameters in the mediating models directly (i.e., A and B in the Models 1-2 of Figure 2).

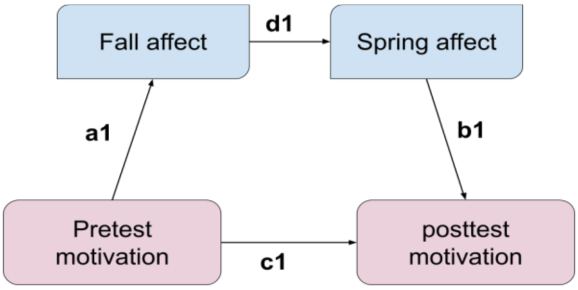
Figure 2

The Three Hypothetical Models

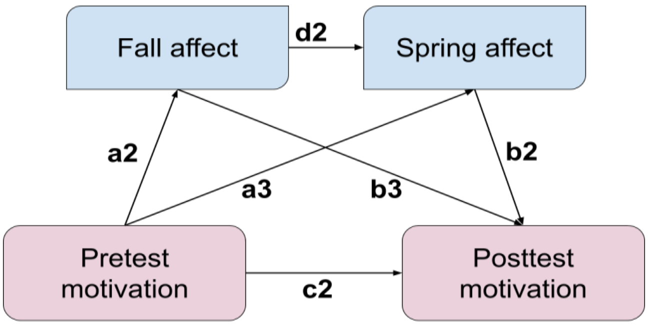
Model 0



Model 1



Model 2



Note. Model 0 is to set the scene that earlier motivation predicts the later motivation, for comparison with Model 1 and Model 2. Model 1 and Model 2 examine the mediating effect of affect on motivation change, the focus of this study.

Another reason for using SEM is that this study uses one mediator occurring at two time points (i.e., detected affect in fall and spring; Figure 2). Past studies normally use one mediator or several parallel mediators (Preacher & Hayes, 2008). Placing one mediator occurring at two time points allows an opportunity to investigate the following effects simultaneously.

1.Direct effects of prior motivation on later one (c0, c1, and c2), prior motivation on autumn-semester affect (a1 and a2), and spring-semester affect on later motivation (b1 and b2).

2.Direct effects of autumn-semester affect on spring-semester affect (d1 and d2). 3.Mediating effects of both fall affect and spring one (a1*d1*b1 and a2*d2*b2).

4.Direct effects of prior motivation on spring affect (a3).

5.Mediating effects of spring affect from prior motivation to later one (a3*b2). Effects 2-5 appear to be new in the literature.

Conceptually, this study aims to test the three models in Figure 2 for the different constructs of self- report motivation and detected affect, separately.

Model 0. Early motivation predicts later one.

Model 1 (concise model). Student affect over two phases (fall semester and spring semester) mediates the effect that early motivation predicts later one.

Model 2 (full model). The detected affect of both phases mediates the effect that early self-report motivation predicts later one.

Research Methodology

General Background

Participants and Data Source. The participants included over 8,000 2nd- to 8th-grade students in the United States who used an intelligent tutoring system, Reasoning Mind, during the academic year of 2017-2018 (including fall and spring terms) (Ocumpaugh et al., 2013). All the users of the system were included and so no sampling methods were used. The students filled in an online survey about three motivational constructs (self- concept, interest, and value). This study was not preregistered and was a secondary analysis of identified data, which were exempt from ethics committee review in both the countries of the original researchers and the author.

The system was designed for pre-K through grade 7 students to learn mathematics, utilizing a story- based curricular design and offering opportunities for a blended learning pedagogy. Teachers could use the system to provide students with self-paced, personalized learning experiences in a virtual city. A rewarding system provided students with game-based problem-solving experiences. A pedagogical agent guided students through the virtual city and learning processes. The learning modes in the system included guided study, solving additional related problems for mastery, collaborative games, and summative assessment. Students spent most of their time on guided study, with a sequence of mathematical objectives and problems of different levels of difficulty. Teachers could walk around the classroom to work with individual students or groups. The system also provided homework print for students to practice problems based on their prior achievements diagnosed by the platform.

Self-Report Motivation for Mathematical Learning. Students reported their motivations for mathematical learning on three scales (factors or constructs): self-concept (or concept), interest, and value. Each subscale has four items (Table 1).

Table 1

The Survey Items and Factor Loadings by Confirmatory Factor Analysis

Items	Factor Loadings
Concept (or Self-Concept)	
1. I would like math more if it wasn't so hard.	0.644
2. I make a real effort in math, but it seems harder for me than for other students.	0.640
3. Nobody's perfect, but I'm just not good at math.	0.614

4. Some topics in math are just so hard that I know from the start I'll never understand them.	0.676
Interest	
1. It is important to me to be a good mathematician.	0.661
2. I enjoy working on math problems.	0.708
3. Math is one of the things that is important to me personally.	0.740
4. I would even give up some of my spare time to learn new topics in math.	0.511
Value	
1. Math will help me in my daily life.	0.701
2. I need math to learn other subjects.	0.571
3. I will need math to get into college.	0.587
4. I will need math to get the job I want.	0.645

Note. The scale for self-concept (or concept) is 1 = Not at all like me to 4 = exactly like me; the scale is reversed coded in the later analysis because all the items are negative wording. The scale for interest is 1 = Not at all like me to 4 = exactly like me. The scale for value is 1 = strongly disagree to 4 = strongly agree.

Students were surveyed at the start and end of the school year. As such, the measure names in this study are added a prefix "pre" to represent a pre-test measure (e.g., pre concept) and "post" to represent a post-test one (e.g., post interest). These measures are given in Table 2 in items 1-6.

Table 2

Means, Standard Deviations (SDs), and Correlations Between the Measures

measures	mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1.preConcept	***2.493***	0.830													
2.preInterest	***3.148***	0.693	*0.244*												
3.preValue	3.467	0.501	0.017	*0.503*											
4.postConcept	***2.575***	0.861	*0.501*	*0.219*	0.042										
5.postInterest	***3.030***	0.749	*0.212*	*0.594*	*0.358*	*0.246*									
6.postValue	3.453	0.511	*0.121*	*0.354*	*0.416*	0.041	*0.575*								
7.EngagementFall	***0.715***	0.157	0.04	**-0.092**	-0.01	-0.059	**-0.142**	-0.037							
8.ConfusionFall	***0.023***	0.033	*-0.101*	0.001	-0.019	-0.011	0.026	-0.031	*-0.689*						
9.FrustrationFall	0.008	0.011	*-0.08*	0.057	-0.018	-0.063	**0.085**	0.031	*-0.407*	*0.192*					
10.BoredomFall	***0.106***	0.053	*-0.111*	0.015	-0.051	-0.064	0.058	-0.002	*-0.547*	*0.120*	*0.504*				
11.EngagementSpring	***0.741***	0.141	0.076	-0.036	0.008	-0.021	-0.059	0.007	*0.718*	*-0.504*	*-0.380*	*-0.479*			

12.ConfusionSpring	***0.010***	0.021	0.002	**0.079**	0.024	0.046	**0.099**	0.013	*-0.442*	*0.437*	*0.093*	0.078	*-0.378*		
13.FrustrationSpring	0.008	0.013	*-0.171*	-0.04	-0.041	*-0.108*	-0.007	0.031	*-0.350*	*0.290*	*0.313*	*0.286*	*-0.363*	0.061	
14.BoredomSpring	***0.114***	0.075	*-0.199*	-0.025	-0.051	*-0.092*	0.03	-0.034	*-0.638*	*0.463*	*0.393*	*0.524*	*-0.724*	*0.182*	*0.421*

Note. The underlined figures are not significant at $p < 0.05$. Italic (green) figures indicate their direction and significance support the prediction that same-valance measures have positive correlations, and bold (red) ones violate. The bold, italic (blue) means are different between two times in the same measure (e.g., preConcept is different from postConcept).

After the self-report scales were initially designed, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were performed to select proper items. Using both EFA and CFA is a recommended practice for developing psychological scales (Morin et al., 2017). Participants' responses were coded to let higher scores represent higher degrees in the meaning of the constructs. Next, based on the EFA and CFA results, mean scores of the items belonging to the same factors were calculated and used to examine the hypotheses. The following describes the three phases of scale development.

Initial Design. Four subscales were initially developed on the basis of related literature (Marsh et al., 2005; Ryan & Ryan, 2005). In total, there were 18 items.

Self-concept had five items (e.g., "I would like math more if it wasn't so hard.") with a scale of 1 = *Notat all like me* to 4 = *exactly like me*.

Class interest had four items. Two problems asked for the importance of learning math (e.g., "How important is it for you to learn a lot in math classes?") with a scale of 1 = *Not at all important* to 5 = *very important*. Two problems asked for opinions (e.g. "How much do you look forward to math classes?" with a scale 1 = *Not at all* to 5 = *very much*. All items were deleted after item selection (the next section).

Domain interest had five items (e.g., "I enjoy working on math problems." Its scale is 1 = *Not at all likeme* to 4 = *exactly like me*).

Value comprised four items (e.g., "Math will help me in my daily life.") with a scale of 1 = *strongly disagree* to 4 = *strongly agree*.

Exploratory Factor Analysis (EFA). EFA initially explored the likely factor structures of the 18 items using the pre-test data. A screen test revealed 3 to 4 factors were likely proper solutions. The 3-factor solution presented a more proper solution than the 4-factor solution after scrutinizing the 3- and 4- factor solutions and the original scale designs (Marsh et al., 2005). The reasons include that the 4-factor solution had one factor with only one item, but, in the 3-factor solution, each factor had 4 to 6 items with factor loading > 0.500.

Both interest and value only had 4 items with non-overlapping, larger-than-0.500 factor loading, and the same scaling method. To be consistent, the self-concept also contained four items, excluding one item with the lowest factor loading from the original five items. This procedure resulted in a self-reported motivational scale with 12 items in total (i.e., 4 items * 3 factors = 12 items).

Confirmatory Factor Analysis (CFA). A CFA validated the 3-factor model (with 12 items) identified by the EFA using the post-test data. The fit indices revealed the 3-factor model was a good fit to the empirical data ($\chi2$ (df) = 559.142 (51), $p < 0.0005$; RMSEA = 0.045; SRMR = 0.038; CFI = 0.964;

and TLI = 0.953. (The "Data Analysis/Hypothesis testing" section presents the full name, meaning, and use of the fit indices in justifying a model fit.)

The CFA factor loadings of the items for their respective factors are from 0.511 to 0.740 (Table 1). This study had a large sample size and all the factor loadings were larger than 0.500, the generally accepted size for factor loadings (Morin et al., 2017). The internal reliability coefficients (Cronbach's alpha) of the three subscales were 0.734 (self-concept), 0.741 (interest), and 0.715 (value). A Cronbach's alpha > 0.700 is mostly acceptable in the literature (Cortina, 1993). Overall, the self-reported motivational scale had desirable construct validity and internal reliability.

Detected Affect

A Machine Learning Procedure to Detect Online Affect. As a later development, the Reasoning Mind system introduced a procedure to detect students' affective states and behaviour online (Ocumpaugh et al., 2015). Trained coders observe students for up to 20 seconds and record students' affective states and behaviours on a synchronized device. Predictive algorithms, developed using standard cross-validation techniques, are developed to predict the observation records by features identified using online learning log data with feature engineering. Thus, students' online affective states and behaviours can be automatically inferred using online learning log-file data.

The use of the procedure as a labelling system in doing classroom observations has been used to successfully detect diverse types of student affective states. Ocumpaugh et al. (2013) find that the learning system is quite engaging, with students spending 66% in engaged concentration and dividing the remainder of their time (34%) between boredom, confusion, and frustration.

The Four Detected Affects. This study focused on four affective measures: engaged concentration, confusion, frustration, and boredom. The data were obtained from the automated affect detectors that were trained on the procedure with observations. The distribution percentages of affective states in the procedure with observations were: engaged concentration (65.9%), confusion (2.0%), frustration (1.1%), boredom (13.7%), and NA (17.3%). The distribution percentages in the new data from the automated detectors were engaged concentration (78.8%), confusion (31.1%), frustration (< 0.001%), and boredom (0.84%). Due to the extremely low incidence of frustration and boredom, the data for these two states were re-threshold to match the distribution of the procedure with observations. Accordingly, the top 1.1% and 13.7% of the affect predictions were marked as active for frustration and boredom, respectively.

As stated in the 'participants' section, this study used students' experiences of online learning in autumn and spring semesters. The mean scores of the four affective constructs for each student were calculated for each semester separately. As such, there were eight affective measures (= 4 detected affects* 2 semesters) in total. To give each measure a unique name in this study, the words "autumn" and "spring" were added for each detected affective construct (e.g., "Engaged Autumn" and "Confusion Spring"). Measures 7-4 in Table 2 present the names of the eight detected affect measures.

Data Analysis

Data Preparation. The initial dataset contained the four detected affect measures of 10,300,358 student actions at a 20-second grain size. The mean scores of the action-level measures formed their student-level measures for the fall and spring terms separately. Only 8,074 students have both fall and spring data, formed as a subset of data for further combination with the self-report motivation data.

The survey dataset contained 377,856 answers (209,610 for the pretest and 168,246 for the posttest). The answer-level dataset was transformed to a student-level one for the pretest and posttest separately. Only students with both posttest and posttest were retained (i.e., by inner join), resulting in a subset of data of 5,435 students. This dataset was used for EFA and CFA to determine the final items used in this study (Table 1). Detailed EFA and CFA procedures are presented in the Method/Measures/Self-report motivation section.

The datasets generated from the automated affect detectors and self-reported measures were combined using an inner join. This procedure selected a student sample that had affect data in both the autumn and spring semesters and motivation data in both the pre-test and post-test. The final dataset contained 792 students for later data exploration and analysis.

Data Exploration. Pearson correlation explored simple relationships between all the measures (Hypothesis 1). Because many correlational tests were simultaneously performed, Benjamini-Hochberg procedure was performed to correct multiple comparisons (Benjamini & Hochberg, 1995).

Hypothesis Testing. Path analysis with mediating effects using SEM methodology examined the mediating effects of detected affect on self-report motivation. Figure 2 presents the detailed models examined.

Generally, evaluating model fit to empirical data in SEM methodology applies multiple indices. Typically, used indices and criteria include the root-mean-square error of approximation (RMSEA) < 0.080, standardized root mean residual (SRMR) < 0.080, comparative fit index (CFI) > 0.900, and Tucker-Lewis index (TLI) > 0.900 (Hair et al., 2010). That is, for both RMSEA and SRMR, a smaller value indicates a better model fit, while for both CFI and TLI; a larger value indicates a better model fit. A traditional index but rarely used one is a non-significant chi-square ($\chi 2$), because a significant $\chi 2$ would be easily obtained if the sample size is large (Bollen & Long, 1993). As a convention, this paper would present $\chi 2$ and its related statistics. The reason is that most highly used criteria (e.g., RMSEA, CFI, and TLI) use $\chi 2$ as a base to adjust themselves such as by sample size, degree of freedom, and penalty for a complex model (Chiu, 2020).

Model 0 and Model 2 are just-identified or saturated models, which have the same number of free parameters as the total number of path coefficients and measure variances. This will create a zero degree freedom (Raykov et al., 2013). As such, all the fit indices, as described in the previous paragraph, cannot be obtained. Even with this, the path coefficients obtained for just-identified models are still trustworthy (Heyder et al., 2017).

Statistical packages in the free R language environment were used to perform factor analysis (EFA and CFA for self-report motivation item selection), Cronbach's alpha, correlation, and hypothesis testing. R codes with outputs were released on two Kaggle notebooks (https://www.kaggle.com/code/meishiuchiu1/geniesurveyefa-cfa-cronbachalpha-r-datanouserid; https://www.kaggle.com/code/meishiuchiu1/geniesurveyaffect-corr-mediating-r).

Research Results

Data Exploration: Paired T-Test and Correlations

Paired t-test examines whether differences occur between the same motivations or affect measures over two time periods. The results reveal that students change their motivations and affects over time, except for value and frustration (Table 2). This may serve as an additional reason for using the posited model (Figure 2) to address the changes in motivation and affect in the academic year.

Pearson's correlations in Table 2 explore the bivariate relationships between all the measures. The results reveal that the measures of the same valances (either positive or negative) and the same measures obtained in two times (e.g., interest in pre-test and post-test; frustration in fall and spring) are positively correlated for both motivation and affect measures, separately. There are, however, low correlations between self-report motivation and detected affect.

Motivation. Self-concept, interest, and values are positive-valence constructs. All three constructs are significantly, positively correlated, except for three non-significant ones (Table 2). Generally, the highest correlations occur between the pre-test and post-test for all three motivation measures: self-concept ($r = 0.501$), interest ($r = 0.594$), and value ($r = 0.415$).

Affect. Engaged concentration is the only positive-valence affect. Engaged concentration has negative correlations with the other negative-valence affect measures (confusion, frustration, and boredom; Table 2). The three negative-valence affect measures are positively correlated with each other, except for two non-significant correlations. Positive correlations occur for the same affect detected at different times (e.g., positive correlation between boredom in fall and that in spring).

Motivation and Affect. The correlations between motivation and affect are very low (-0.199 to 0.099; Table 2) and only 12 out of the 48 correlations are significant. Self-concept is negatively correlated with negative-valence affect measures, though with many non-significant correlations. Interest generally has non-significant correlations with affect; all of its significant correlations violate the general principle of positive correlations between same-valence measures (e.g., -0.092 pre interest with engaged autumn) and vice versa (e.g., 0.099 post interest with confusion spring). No significant correlation occurs for value.

Model Fit Evaluation

Table 3 presents the fit index values for Model 0 and Model 1, each model with a pair of motivation and affect (Figure 2). Model 2, like Model 0, has a zero degree of freedom, and fit index values are not used to assess the models. Only two models of Model 1 reveal bad model fit values; that is, the models of self- concept with frustration (RMSEA = 0.104; TLI = 0.844) and boredom (RMSEA = 0.114; TLI = 0.879).

Table 3

Fit Index Values for the Models

	Mediator	$\chi2$	df	*p*	RMSEA	SRMR	CFI	TLI
Concept								
Model 0		0	0	NA	0	0	1	1
Model 1	engagement	6.525	2	0.038	0.053	0.017	0.994	0.983
	confusion	2.674	2	0.263	0.053	0.017	0.994	0.983
	frustration	19.247	2	0.000	**0.104**	0.044	0.948	**0.843**

	boredom	22.467	2	0.000	**0.114**	0.042	0.960	**0.879**
Interest								
Model 0		0	0	NA	0	0	1	1
Model 1	engagement	10.749	2	0.005	0.074	0.019	0.991	0.972
	confusion	6.102	2	0.047	0.051	0.025	0.992	0.976
	frustration	5.778	2	0.056	0.049	0.022	0.991	0.973
	boredom	2.182	2	0.336	0.011	0.012	1.000	0.999
Value								
Model 0		0	0	NA	0	0	1	1
Model 1	engagement	2.891	2	0.236	0.024	0.01	0.999	0.996
	confusion	1.674	2	0.433	0.000	0.011	1.000	1.003
	frustration	1.726	2	0.422	0.000	0.012	1.000	1.004
	boredom	1.447	2	0.485	0.000	0.010	1.000	1.000

Note. **Bold** (red) figures indicate their direction and significance violate predictions. χ2 = chi-square (or minimum function test) statistic; df = degrees of freedom; RMSEA = root mean square error of approximation; SRMR = standardised root mean square residual; CFI = comparative fit index; TLI = Tucker– Lewis index.

Due to some undesirable fit index values for Model 1 (the concise mediating effect from a particular affect over two phases), the following only focuses on the results based on Model 2 (the full mediating effects via the both phases affect). Besides, Model 2 provides more mediating effects, as addressed in the next section, than Model 1. Table 4 presents the path coefficients based on Models 0 and 2

Table 4

Path Parameters for Models 0 and 2 (Figure 2)

		Model 0					Model 2					
	Mediator	c0	a2	a3	b2	b3	c2	d2	a2*d2*b2	a2*b3	a3*b2	total2
Concept		0.501										
	engagement		0.040	0.047	-0.001	-0.074	0.504	*0.716*	0.000	-0.003	0.000	0.500
	confusion		*-0.101*	0.047	0.033	0.025	0.503	*0.442*	-0.001	-0.003	0.002	0.501
	frustration		*-0.080*	*-0.146*	-0.018	-0.018	*0.496*	*0.301*	0.000	0.001	0.003	0.501
	boredom		*-0.111*	*-0.143*	0.017	-0.017	0.502	*0.508*	-0.001	0.002	-0.002	0.501
Interest		0.594										
	engagement		**-0.092**	0.029	0.053	**-0.126**	*0.584*	*0.720*	-0.033	*0.012*	0.002	0.594
	confusion		0.001	**0.079**	0.051	0.002	*0.590*	*0.436*	0.000	0.000	0.004	0.594

	frustration		0.058	-0.058	0.000	0.051	*0.591*	*0.316*	0.000	0.003	0.000	0.594
	boredom		0.014	-0.033	0.028	0.034	0.594	*0.526*	0.000	0.000	-0.001	0.594
Value		0.415										
	engagement		-0.011	0.014	0.054	-0.074	*0.414*	*0.717*	0.000	0.001	0.001	0.415
	confusion		-0.019	0.033	0.016	-0.028	*0.414*	*0.437*	0.000	0.001	0.001	0.415
	frustration		-0.017	-0.035	0.041	0.027	0.417	*0.312*	0.000	0.000	-0.001	0.415
	boredom		-0.051	-0.024	-0.027	0.035	0.416	*0.523*	0.001	-0.002	0.001	0.415

Note. The path parameter estimates are completely standardised solutions. The underlined figures are not significant at $p < 0.05$. Italic (green) figures indicate their direction and significance support predictions, and bold (red) ones violate.

Mediating Effects Examination 1: Inferred Mediating

A criterion for a mediating effect is the effect of a motivation from pre-test to post-test (Model 0) reduces when including mediators (Model 2), using this study as an example (cf. The Present Study). That is, a mediating effect occurs if c0 in Model 0 is larger than c2 in Model 2 (i.e., c0 > c2; Figure 2) although a direct multiplication of several related relationships is also an indicator (e.g., a2*b3, Table 4).

Given the above criterion, six pairs of motivation and affect show mediating effects (i.e., c0 > c2; Figure 2; Table 4). That is, the predictive effect from pre-test to post-test self-concept (c0 = 0.501) reduces when frustration works as a mediator (c2 = 0.496); the predictive effect from pre-test to post-test interest (c0 = 0.594) reduces when engaged concentration (c2 = 0.584), confusion (c2 = 0.590), and frustration (c2

= 0.591) work as mediators; the predictive effect from pre-test to post-test value (c0 = 0.415) reduces when engaged concentration (c2 = 0.414) and confusion (c2 = 0.414) work as mediators.

Mediating Effects Examination 2: Direct Mediating

The direct effects between motivation and affect (based on a2 and a3; Model 2 of Figure 2) slightly objectivise the above inferred mediating effects (mainly based on c0 and c2). This emerges as a direct measure of a mediating effect obtained by multiplying all the effects relating to one mediator in its related paths. (Model 2 in Figure 2 shows the related paths.) The only significant mediating effect is the model for interest mediated by engaged concentration in fall (a2 b3 = -0.092 -0.126 = 0.012). This result implies that engaged concentration in fall mediating the effect of self-concept from pretest to post-test is the most salient mediating effect in this study.

Compared with the pre-test interest and value, the pre-test self-concept directly predicts the most affect measures (Table 4). Pre-test self-concept negatively predicts confusion (a2 = -0.101), frustration (- 0.080), and boredom (-0.111) in fall as well as frustration (a3 = -0.146) and boredom (-0.143) in spring.

Interest has three direct predictive effects (Table 4). Pre-test interest negatively predicts engaged concentration in fall (a2 = -0.092), which further negatively predicts post-test interest (b3 = -0.126). Pre- test interest positively predicts confusion in spring (a3 = 0.079). Value has no significant direct effect on any detected affect.

Longitudinal Development within Motivation and Affect

Among the three motivation constructs, interest has the largest path coefficients from pre-test to post-test (0.584 to 0.594), followed by self-concept (0.496 to 0.504) and value (0.414 to 0.416; c2 in Table 4). The results imply that the degrees of stability of the three motivation constructs over time are the self- concept, the interest, and the value, in descending order. (Similar trends occur in Model 0.)

Among the four affect measures, engaged concentration has the largest path coefficients from fall to spring (0.716 to 0.720), followed by boredom (0.508 to 0.526), confusion (0.436 to 0.442), and frustration (0.301 to 0.316), in descending order (d2 in Table 4). The results also imply the degrees of stability among the four affect measures over time.

Discussion

The MAR Model

This study posits the MAR model (Figure 1) as a theoretical basis with a reference to related SRL theories (Baars et al., 2017; Efklides, 2011). The MAR model highlights the reciprocal relationship between motivation and affect as part of the SRL process. The results support the MAR model for specific pairs of motivations and affects (Table 4) in terms of mediation effects as evidence of regulation. The most salient mediation effect occurs for the pair of interest and engaged concentration. The MAR model is partially supported for self-concept and is completely not supported for value. Stable effects are found for the same motivations and affects over time. All these are addressed in detail as follows.

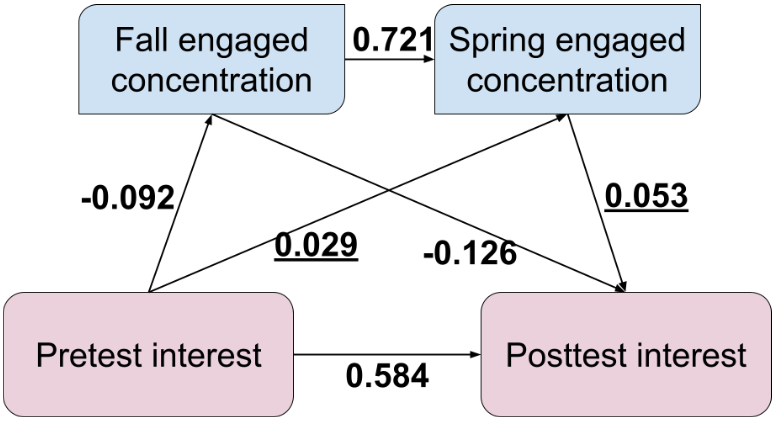
Few but Unique Mediating Effects of Affect on Motivation

The unique contribution of this study is the mediating effects of affect on motivation using longitudinal data and SEM predictive methodology. The results reveal that the full mediating model (Model 2 in Figure 2, where the affects in both phases mediate the effect that early motivations predict later ones) fits empirical data properly for diverse types of motivation and affect. (cf. Model 1 fails to fit empirical data for some types of motivation-affect pairs).

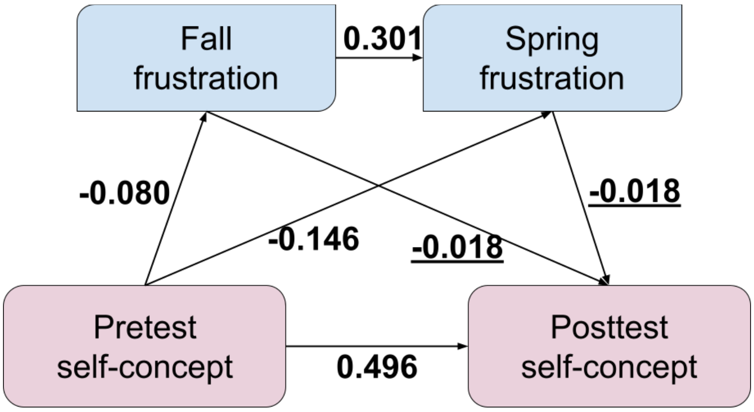
The significance and direction of the mediating effect from affect are different among the three motivation measures. Interest’s mediating effects from affect are most salient, unique but explainable (e.g., engaged concentration in Figure 3-A). For the self-concept, the most significant mediating effects are from a negative-valance affective construct in a negative direction, an unsurprising result (e.g., frustration, Figure 3-B). The value has no mediating effect.

Figure 3

Example Path Coefficients



Example A. The Interest-Engagement Model



Example B. The Self-Concept and Frustration Model

Note. The underlined path coefficients are not significant at $p < 0.05$.

Interest

The most salient mediating effect is engaged concentration, mediating interest from the pre-test to the post-test. However, this mediating effect is unique in its patterns. The path from pre-test interest to engaged concentration in the fall is negative, and the path from engaged concentration in the fall to post- test interest is also negative. This overall mediating effect is significant in positivity. A closer look at engaged concentration in the spring finds that the paths leading to and leaving from it are both positive, though non-significant. This result suggests that interest's negativity in engaged concentration only occurs in the early stage. In a later stage of online learning, interest will gradually lead to engaged concentration. To add another mediating effect, interest leads to confusion only in a later stage (spring).

The interest is a motivation construct relating to science career choices (Cerinsek et al., 2013; Maltese & Tai, 2010). In order to explain the unique finding, we need to delve into the essence of 'interest.' The interest includes both subject-matter interest and situational interest (Høgheim & Reber, 2015), with situational interest being related to engaging in tasks but not to achievement (Zhu et al., 2009). Even though this study uses self-report 'subject-matter' (mathematics) interest items (Table 1), the unique pattern of mediating effects from disengagement and confusion suggests interest is sensitive to learning format (or situation) changes (e.g., between traditional and online learning).

For educational practices, teachers need to be aware that students interested in mathematics may experience initially less engaged concentration but lately (with confusion) in online mathematical problem- solving. Confusion is a sign of good or improved learning (Richey et al., 2019). The reason for this may be that the differences between traditional paper-and-pencil and computer-supported learning formats matter for students who are interested in mathematics. More support in learning format changes is needed for students interested in the traditional format of learning mathematics (e.g., manipulatives, drawing, and paper-and-pencil). An online system design simulating traditional mathematics teaching or learning environment (e.g., virtual reality, diverse hands-on activities, and creative arts) may also be a solution. These speculations are worth future research.

Self-Concept

Self-concept in mathematics is consistent with affect detected during online mathematical learning. This is evident in the self-concept's multiple negative paths with the negative-valence affect. The most salient affect is frustration. Students initially perceiving high self-concept in pre-test are less frustrated in both fall and spring. This is also true for boredom and for initial confusion, although the mediating effect is not salient (without reduced path parameters from pre-test to post-test self-concept).

The results concur with most scholars' emphasis on self-concept (confidence) for persistence and action taken (Burton, 2004). Another reason for the results may be that self-concept is more highly related to achievements than all other motivation measures (Chiu, 2018; Marsh & Hau, 2004). Achievement may be one of the major sources of motivation and affect (Marsh & Hau, 2004). Future research needs to validate this speculation further by incorporating achievement.

For education practice, the self-concept can prevent students from the negative affect of online learning lasting for a long time (e.g., over two semesters in this study). Cultivating the self-concept appears to benefit students in both traditional and online learning formats.

Value

Unlike the other two motivational constructs (the self-concept and the interest), the affective states did not significantly mediate the relationship for value. A direct reason for the finding is that value is not mediated by online affect. Another reason may be that the items of value focus on 'extrinsic motivation' (Cerasoli et al., 2014). Extrinsic motivation relates learning mathematics to the outer world (i.e., to help life, learn other subjects, enter college, and get jobs; Table 1), not inner value (or interest) (e.g., I enjoy working on math problems).

The research has indicated that extrinsic value or motivation is related to social norms and authorities in students' lives (Chiu, 2017). The affect experiences detected through online learning may be more likely to mediate or reflect students' inner-oriented motivations (e.g., self-concept and interest) than their extrinsic value. This speculation still needs to be validated, perhaps by incorporating measures of related social norms or authorities.

Stable Longitudinal Development of Motivation and Affect

For each specific measure, both motivation and affect are relatively stable. That is, pre-test motivation predicts later motivation; the affect in fall predicts that in spring. Specifically, the degrees of stability for motivation over time are self-concept, interest, and value, and those for affect are engaged concentration, boredom, confusion, and frustration, in descending order. The findings are consistent with past research findings that early motivation (e.g., self-concept and interest) predicts later one (e.g., Marsh et al., 2005). Past research on affect also provides partial evidence. More empirical studies need to validate these findings, especially for affect and for the differential degrees of stability over time between different constructs.

For educational practices, the findings suggest a stable, long-term quality of motivation and affect. This calls for a need for earlier interventions to promote desirable motivation (e.g., confidence and interest) and affect (e.g., engaged concentration).

Conclusion and Implications

The MAR Model is evidenced by mediating effects of the affect on motivation. To date, very few studies focus on relationships between the self-report motivation and the detected affect. This study posits the MAR model to address this issue. The results based on empirical datasets find that earlier self-concept can be mediated by fewer frustrations in online learning and lead to later self-concept. Interest triggers early disengagement and late confusion during online learning. Value is not mediated by affect.

SEM is used for longitudinal experimental design. When data is collected longitudinally, it is suitable to use predictive algorithms. SEM with mediating models offers opportunities to investigate more complex relationships when common measures are collected across time.

The findings of this study offer implications for pedagogical and online learning system designs. Both offline pedagogical designs and online learning system designs need to consider the gaps between different formats of learning. Students with a high interest in learning mathematics in real classrooms may fail to adjust themselves easily to online mathematical problem-solving and experience disengagement in the early stages (one semester) and confusion in the later stages (after one semester). Future research may investigate the effectiveness of different types of support for students in traditional and online learning formats.

Limitations and Suggestions for Future Research

Parsimony is a desirable characteristic of theories. This study finds a full-path model (Model 2) fits empirical data more properly than a concise model (Model 1). Even though a full-path model may reflect the fact, a more complicated model (including measures such as computer-assisted learning preference, learning styles, gender, and school) with reduced paths may be a solution to build a parsimonious model by future research.

"The affect" is a complex term with different meanings in different academic fields. This study uses a specific methodology in order to investigate relationships between related constructs in two major fields (educational psychology and technology). Unifying related constructs in different fields is a challenging task because even for the same field, affective constructs are not equivocal and orthogonal. For example, engaged concentration has a high correlation with boredom (Table 2). The commonality and differentiation between different types of affect and motivation assessed or detected using different tools over time remain an interesting issue to address in future research. Perhaps the first step is to develop the paper to synthesise past related studies by systematic review or meta-analysis, simultaneously suggesting methods to solve this complex issue.

This study examines the mediating role of detected affect (occurring during online learning) only. Potential mediators exist. For example, the achievement may mediate motivation (or affect) changes given the stable relationships between motivation and achievement. The MAR model may benefit from adding the achievement for its further elaborated, thorough the model.

Acknowledgments

The paper was part of the outcome when the author served as a 2020 Fulbright scholar at the University of Pennsylvania Graduate School of Education, USA, working for facilitating dialogues between educational and data sciences for STEM education. The author wishes to express her gratitude to Dr. Ryan Baker and his lab members (including Dr. Jaclyn Ocumpaugh and Dr. Shamya Karumbaiah) for valuable insights, data provision, and initial analysis of affective computing for the paper. The author, however, is fully responsible for the content of the paper. The author does not perceive conflicts of interest for this study.

Data declaration

The initial datasets are not publicly accessible, but available to researchers upon request to Penn Centre for Learning Analytics (https://www.upenn.edu/learninganalytics/geniemail.html). The data were from the original researchers in fully identified form, which fits under the category of secondary analysis of identified data and is exempt from ethics committee review, category 4-ii, in the USA. Secondary analysis of identified data is also exempt from ethics committee review in Taiwan (the author's country).

References

Baars, M., Wijnia, L., & Paas, F. (2017). The association between motivation, affect, and self-regulated learning when solving problems. *Frontiers in Psychology, 8,* Article 1346. https://doi.org/10.3389/fpsyg.2017.01346

Baker, R. S. J. d., & Rossi, L. M. (2013). Assessing the disengaged behavior of learners. In R. Sottilare, A. Graesser, X. Hu, & H. Holden (Eds.), *Design recommendations for intelligent tutoring systems: Volume 1, Learner modeling* (pp. 155–165). US Army Research Laboratory.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B, 57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics, 3,* 220–238. https://doi.org/10.18608/jla.2016.32.11

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Sage.

Burton, L. (2004). "Confidence is everything"-Perspectives of teachers and students on learning mathematics. *Journal of Mathematics Teacher Education, 7,* 357–381. https://doi.org/10.1007/s10857-004-3355-y

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin, 140*, 980–1008. https://doi.org/10.1037/a0035661

Cerinsek, G., Hribar, T., Glodez, N., & Dolinsek, S. (2013). Which are my future career priorities and what influenced my choice of studying science, technology, engineering or mathematics? Some insights on educational choice-case of Slovenia. *International Journal of Science Education, 35*, 2999–3025. https://doi.org/10.1080/09500693.2012.681813

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Chiu, M.-S. (2017). High school student rationales for studying advanced science: Analysis of their psychological and cultural capitals. *Journal of Advances in Education Research, 2*, 171–182. https://doi.org/10.22606/jaer.2017.23005

Chiu, M.-S. (2018). Effects of early numeracy activities on mathematics achievement and affect: Parental value and child gender conditions and socioeconomic status mediation. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(12), Article em1634. https://doi.org/10.29333/ejmste/97191

Chiu, M.-S. (2020). Exploring models for increasing the effects of school information and communication technology use on learning outcomes through outside-school use and socioeconomic status mediation: The ecological techno-process. *Educational Technology Research and Development, 68,* 413–436. https://doi.org/10.1007/s11423-019-09707-x

Chiu, M.-S. (2022). Transcend socioeconomic status constraints to mathematics and science achievement by collaborative problem-solving: The female people-smartness hypothesis. *Frontiers in Psychology, 13*, Article 944329. https://doi.org/10.3389/fpsyg.2022.944329

D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology, 105,* 1082–1099. https://doi.org/10.1037/a0032674

D'Mello, S., Dieterle, E., & Duckworth, A. (2017). Advanced, analytic, automated (AAA) measurement of engagement during learning. *Educational Psychologist, 52,* 104–123. https://doi.org/10.1080/00461520.2017.1281747

D'mello, S. & Graesser, A. (2007). Mind and body: Dialogue and posture for affect detection in learning environments. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Frontiers in artificial intelligence and applications: Vol. 158*, *Artificial intelligence in education* (pp. 161–168). IOS Press.

D'Mello, S. & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22,* 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

D'Mello, S. & Mills, C. (2014). Emotions while writing about emotional and non-emotional topics. *Motivation and Emotion, 38*, 140–156. https://doi.org/10.1007/s11031-013-9358-1

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46*, 6–25. https://doi.org/10.1080/00461520.2011.538645

Gillet, N., Vallerand, R. J., Lafreniere, M. A. K., & Bureau, J. S. (2013). The mediating role of positive and negative affect in the situational motivation-performance relationship. *Motivation and Emotion, 37*, 465–479. https://doi.org/10.1007/s11031-012-9314-5

Gomez-Chacon, I. M. (2000). Affective influences in the knowledge of mathematics. *Educational Studies in Mathematics, 43*, 149–168. ttps://doi.org/10.1023/A:1017518812079

Hair, J. F., Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective (7th ed.)*. Pearson Education.

Hannula, M. S. (2002). Attitude towards mathematics: emotions, expectations and values. *Educational Studies in Mathematics, 49*, 25–46. https://doi.org/10.1023/A:1016048823497

Heyder, A., Kessels, U., & Steinmayr, R. (2017). Explaining academic-track boys' underachievement in language grades: Not a lack of aptitude but students' motivational beliefs and parents' perceptions?. *British Journal of Educational Psychology, 87*, 205–223. https://doi.org/10.1111/bjep.12145

Hietajärvi, L., Lonka, K., Hakkarainen, K., Alho, K., & Salmela-Aro, K. (2020). Are schools alienating digitally engaged students? Longitudinal relations between digital engagement and school engagement. *Frontline Learning Research, 8*(1), 33–55. https://doi.org/10.14786/flr.v8i1.437

Høgheim, S. & Reber, R. (2015). Supporting interest of middle school students in mathematics through context personalization and example choice. *Contemporary Educational Psychology, 42*, 17–25. https://doi.org/10.1016/j.cedpsych.2015.03.006

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods, 8,* 115–128. https://doi.org/10.1037/1082-989X.8.2.115

Maltese, A. V. & Tai, R. H. (2010). Eyeballs in the fridge: Sources of early interest in science. *International Journal of Science Education, 32*, 669–685. https://doi.org/10.1080/09500690902792385

Marsh, H. W. & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96,* 56–67. https://doi.org/10.1037/0022- 0663.96.1.56

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76,* 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

McLeod, D. B. (1994). Research on affect and mathematics learning in the JRME: 1970 to the present. *Journal for Research in Mathematics Education, 25,* 637–647. https://doi.org/10.2307/749576

Morin, A. L., Miller, S. J., Smith, J. R., & Johnson, K. E. (2017). Reliability and validity of the child and adolescent behavior assessment (CABA): A brief structured scale. *Child Psychiatry & Human Development, 48,* 200–213. https://doi.org/10.1007/s10578-016-0632-9

Ocumpaugh, J., Baker, R.S.J.d., Gaudino, S., Labrum, M.J., & Dezendorf, T. (2013) *Field observations of engagement in Reasoning Mind.* Proceedings of the 16th International Conference on Artificial Intelligence and Education (pp. 624–627). https://doi.org/10.1007/978-3-642-39112-5_74

Ocumpaugh, J., Baker, R. S., & Rodrigo, M. M. T. (2015). *Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 technical and training manual.* Teachers College, Columbia University, Ateneo Laboratory for the Learning Sciences.

Ocumpaugh, J., San Pedro, M.O., Lai, H-y., Baker, R.S., & Borgen, F. (2016). Middle school engagement with mathematics software and later interest and self-efficacy for STEM Careers*. Journal of Science Education and Technology, 25*, 877–887. https://doi.org/10.1007/s10956-016-9637-1

Pachman, M., Arguel, A., Lockyer, L., Kennedy, G., & Lodge, J. (2016). Eye tracking and early detection of confusion in digital learning environments: Proof of concept. *Australasian Journal of Educational Technology, 32*(6), 58–71. https://doi.org/10.14742/ajet.3060

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*, 667–686. https://doi.org/10.1037/0022- 0663.95.4.667

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behaviour Research Methods, 40*, 879–891. https://doi.org/10.3758/BRM.40.3.879

Raykov, T., Lee, C. L., Marcoulides, G. A., & Chang, C. (2013). A commentary on the relationship between model fit and saturated path models in structural equation modeling applications. *Educational and Psychological Measurement, 73*, 1054–1068. https://doi.org/10.1177/0013164413487905

Richey, J. E., Andres-Bray, J. M. L., Mogessie, M., Scruggs, R., Andres, J. M., Star, J. R. … McLaren, B. M. (2019). More confusion and frustration, better learning: The impact of erroneous examples. *Computers & Education, 139*, 173–190. https://doi.org/10.1016/j.compedu.2019.05.012

Ryan, K. E., & Ryan, A. M. (2005). Psychological processes underlying stereotype threat and standardized math test performance. *Educational Psychologist, 40,* 53–63. https://doi.org/10.1207/s15326985ep4001_4

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422–445. https://doi.org/10.1037/1082- 989X.7.4.422

Slater, S., Ocumpaugh, J., Baker, R., LIb, J., & Labrum, M. (2018). Identifying changes in math identity through adaptive learning systems use. In *Proceedings of the 26th International Conference on Computers in Education* (pp. 71–76). Association for Computing Machinery. https://learninganalytics.upenn.edu/ryanbaker/i_paper_93.pdf

van Halem, N., van Klaveren, C., Drachsler, H., Schmitz, M., & Cornelisz, I. (2020). Tracking patterns in self- regulated learning using students' self-reports and online trace data. *Frontline Learning Research, 8*(3), 140–163. https://doi.org/10.14786/flr.v8i3.497

Wang, Y., & Baker, R. (2018). Grit and intention: Why do learners complete MOOCs?. *The International Review of Research in Open and Distributed Learning, 19*(3), 20–42. https://doi.org/10.19173/irrodl.v19i3.3393

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*, 221–229. https://doi.org/10.1016/j.cedpsych.2009.05.002