ASSESSING MATHEMATICS ENGAGEMENT IN THE MOMENT

James A. Middleton
Arizona State University
jimbo@asu.edu

Adi Wiezel
Elon University
awiezel1@asu.edu

Mandy Jansen
University of Delaware
jansen@udel.edu

Engagement can be described as students' tendency to work productively, think deeply, enjoy and value their learning, and to support each other in the process of learning (ZDM article). Each of these dimensions can be indexed by a variety of psychologically validated constructs such as interest and enjoyment, self-regulation, effort, emotional valence and object. The purpose of this paper is to describe a method of assessment that takes under 5 minutes to administer as students are learning mathematics concepts or shortly thereafter in the context of classroom observation and analysis of practices that may support productive engagement. A description of an online survey, its development and psychometric properties assessed in a study of over 1,000 secondary mathematics students in the US is presented, and implications for research on task-level engagement in mathematics classrooms is discussed.

Keywords: Engagement, Emotions, Affect, Assessment

Introduction

The engagement in, and motivation to continue, mathematics has consistently been shown to decrease as students move through compulsory education (Collie et al., 2019). Such engagement is also complex, psychologically, and socially, involving the coordination of memories of prior math experiences, with the social and cognitive characteristics of mathematics tasks and practices interpreted through affective and motivational responses (Fredricks, 2011; Middleton, Jansen, & Goldin, 2017). Moreover, the supportive features of classrooms, in the form of social relationships among learners and their teachers and peers has been shown to be related to different forms of engagement, sometimes enhancing and sometimes diminishing one feature in favor of another (Reindl et al., 2015; Strati et al., 2017). It is generally assumed that aspects of long-term engagement are impacted by the patterns of engagement students experience over time in mathematics tasks and activities. A classic example is the development of personal interest in mathematics from situational interest (Hidi & Reninger, 2016). These aspects of engagement become particularly salient in times of social transition, for example, in the transition from middle to high school mathematics, where one's peer group, level of mathematical rigor and class norms may change substantially (Middleton, Mangu & Lee, 2017).

The larger body of literature on engagement characterizes it as four related dimensions involving affective, cognitive, behavioral, and social facets (e.g., Wang, et al., 2016; Rimm-Kaufman, et al., 2015). Affective engagement consists of the immediate emotional responses to aspects of a learning environment (Fredricks, Blumenfeld, & Paris, 2004), as well as the meta-affective re-evaluations of those responses (e.g., Goldin, 2014; Goldin, 2002). Cognitive engagement can be thought of as students expending effort on coordinating prior information with current information (Middleton, Jansen, & Goldin, 2017). Behavioral engagement is the productive behavior that students engage in in a math environment (Rimm-Kaufman, Bardoody, Larsen, Curby, & Abry, 2015). Finally, social engagement involves the nature of interpersonal relationships and interactions in a math classroom (Wang, Fredricks, Ye, Hofkens, and Lin, 2015). This includes the quality of the relationships and interactions with peers as well as the teacher. We focus on the first two years of high school mathematics, in which each of these facets potentially may change as new environmental demands are placed on incoming students.

One of the difficulties in studying engagement concerns its measurement. While long-term engagement, including students' beliefs about their self-efficacy and personal interest in mathematics have historically been assessed using survey methods (Middleton, et al., 2023), standard surveys have not shown significant success when applied to engagement as experienced in the moment of learning because of the time they take, and the multidimensionality of engagement constructs. It just takes too many standard surveys, each of which takes up lots of time, to probe students' beliefs, emotions, and behaviors in the moment with much fidelity. One exception to this is Experience Sampling Methods (ESM). These typically utilize a (very) short- term survey (taking 2 to 5 minutes only) to uncover students' immediate responses to their experience (Larson & Csikszentmihaly, 1987; Shernoff; Csikszentmihaly, Schneider, and Shernoff, 2003; Shernoff, 2013). In a typical ESM, participants are signaled at random or after a specific pre-determined event (such as a particular lesson or task) to complete a series of closed or open-ended items about their experiences (Shernoff, 2013; Shiffman, Stone, and Hufford, 2008).

An advantage of using ESMs relative to retrospective assessments is that they are better able to capture in-the-moment impressions of events (Shiffman, Stone, and Hufford, 2008). This is important because prior research suggests that while in-the-moment experiences do color and direct more long-term tendencies, people's in-the-moment and after-the-fact impressions of an experience *can* diverge (Shiffman, Stone, and Hufford, 2008). Moreover, ESMs have also been implemented successfully in school settings to capture aspects of students' engagement such as level of concentration (behavioral) and interest and enjoyment (affective) (Shernoff, 2013).

In the remainder of this paper, we describe the development, administration, and psychometric analysis of an instrument designed to assess mathematics students' engagement as closely to moments of learning as possible. Following this analysis, we discuss a broader approach to studying engagement in classroom settings.

Method

Here, we describe the development of an instrument designed to capture engagement in secondary mathematics in the moment. Two primary criteria guided our work: 1) Because we wanted the instrument to be administered immediately following a key activity in a mathematics class, it needed to be short (less than 5 minutes to administer) so as not to unduly interfere with students' learning; and 2) the instrument needed to be multidimensional, meaningfully capturing important features of each of the 4 dimensions of engagement.

Item Development

To ensure broad conceptual coverage, a team of five content experts assembled to define and write items based on the literature across a variety of content areas, including cognitive engagement, behavioral engagement, affective engagement, social engagement, perceived

instrumentality, and mathematics self-efficacy, the latter being two important concepts related to engagement that can also be captured in the moment. These items received several rounds of iterations, including conceptual pairing with a long-term survey designed as part of the same project. Chosen items were reviewed by a focus group of five 9th graders in a high school in the Southwest US and further revised to ensure students' endorsement and understanding of the language used.

Thirty-two 5-point Likert items and one checklist of emotions were retained for pilot testing. Because the instrument was designed to be an in-the-moment assessment, and time limits were a concern, three versions were initially developed for pilot administration. Ten items (including the emotion checklist) that were determined by the team to reflect the most important items for each of the core concepts were selected to form a "core" that would be used across each of the three versions. The remaining 23 items were split across versions A, B, and C. This enabled us to eliminate poorly worded items, note discrepancies in students' interpretations before creating a single instrument with the most consistent items.

ESM Instrument Creation

Following the pilot, a comparison of common items across the three versions, revealed little temporal variation and nearly no structural variation in exploratory factor analyses performed on each of the three versions. We then performed a further comparison of the instrument to another instrument measuring longer-term engagement patterns (Zhang, et al., 2018), yielding a final, single version which included one item about what participants were doing in the moment, 16 multiple-choice items covering cognitive engagement, behavioral engagement, social engagement, perceived instrumentality, and mathematics self-efficacy, and a checklist of 16 emotions which participants would choose, indicating the object of the emotions (e.g., feeling *frustrated* (checked emotion) at the math *activity*, their *peers*, *themselves*, and/or their *teacher* (the objects of the emotion). The survey concludes with a space in which participants can provide any additional comments on their experience. The final instrument was implemented and online using Qualtrics ®.

Sample

The data were collected from first-year high school mathematics classrooms from fourteen teachers across the Mid-Atlantic and the Southwest US in Fall 2018. 450 students had complete responses on all 16 emotion checklist items for each of the four emotion objects as well as the 16 Likert items. 45.8% of these students were from the Southwest, and 54.2% were from the Mid-Atlantic. Student demographics for the schools in the Southwest were: 85-94% low income, 2- 5% White, 1-15% Black, 74-96% Latinx, and 0-5% Asian, Native American, or Multi-Racial; student demographics for the schools in the Mid-Atlantic were: 9-30% low income, 24-57% White, 27-46% Black, 7-24% Latinx, and 0-5% Asian, Native American, or Multi-Racial. Of the sample, 48% of students identified as male, 49% identified as female, and 1% identified as neither or both.

Analysis Strategy

We collected data from students after a focal classroom activity during two semesters, Fall 2018 and Spring 2019. We conducted exploratory factor analyses on the data collected from consented students in Fall 2018 and conducted confirmatory factor analyses on the data collected from consented students in Spring 2019, to examine model fit and configural measurement invariance over time.

Because our items varied in response styles, from 5-point Likert items to binary responses on the checklist of emotions, we assessed the psychometrics for the Likert items and binary emotion checklist items separately. Data was analyzed using polychoric correlations as this limits potential attenuation which may occur if items with relatively few response options are treated as continuous (Byrne, 2005). Moreover, to help clarify the final sample size used in each analysis (as polychoric correlations can make it difficult to determine the final sample size with pairwise deletion), we used listwise deletion to only analyze cases that had complete data on all items, in both our Fall 2018 and Spring 2019 samples.

Results

Likert Items

The optimal number of factors for the 16 Likert items were assessed using a polychoric parallel analysis, which compares a scree plot generated from the actual to one generated using random simulations of the data. The analysis was conducted using the statistical software package R version 3.3.1 and suggested that a maximum of five factors represent the data better than randomness (see Figure 1).

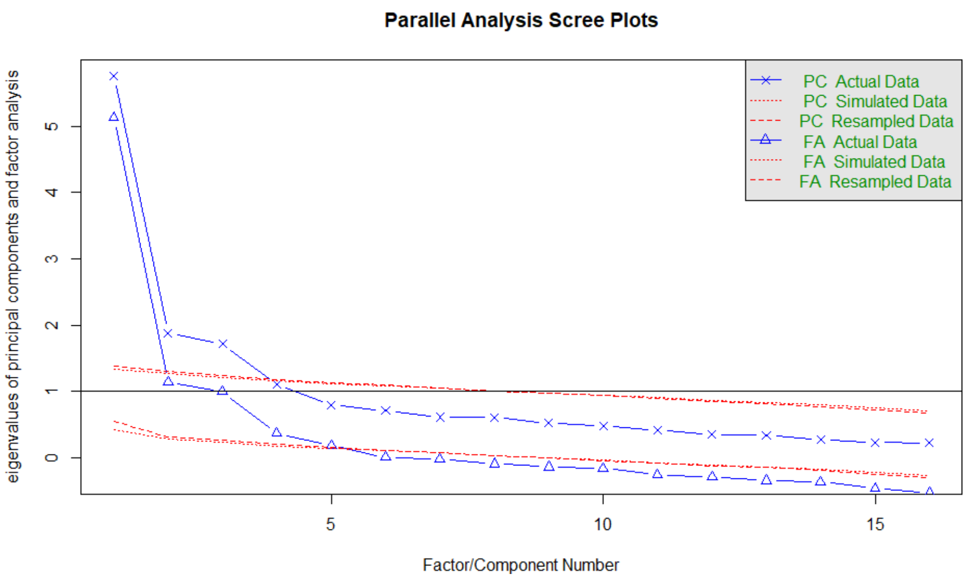

Figure 1. Parallel analysis for Likert items.

Based on a visual inspection of the scree plot from the actual data, we decided to estimate two, three, four, and five factor solutions, using an unweighted least squares (ULS) EFA with Oblimin rotation to allow the factors to correlate. Based on interpretability, we felt the 5-factor solution fit best.

Table 1. Inter-factor correlations.

	Task-Efficacy	Effort	Social Engagement	Instrumentality	Situational Interest
Task-Efficacy	1				
Effort	0.039	1			
Social Engagement	**0.369**	**0.448**	1		
Instrumentality	0.050	**0.272**	**0.388**	1	
Situational Interest	**0.318**	**0.288**	**0.578**	**0.529**	1

Note. Bold text indicates correlations significant at the $p < .05$ level.

Emotion Checklist

The optimal number of factors for the 16 emotion checklist items (across 4 possible objects each: the math activity, the classroom, the teacher, and the self) were assessed using a polychoric parallel analysis. The analysis suggested that a maximum of fifteen factors represent the data better than randomness (see triangles displayed in *Figure 2*).

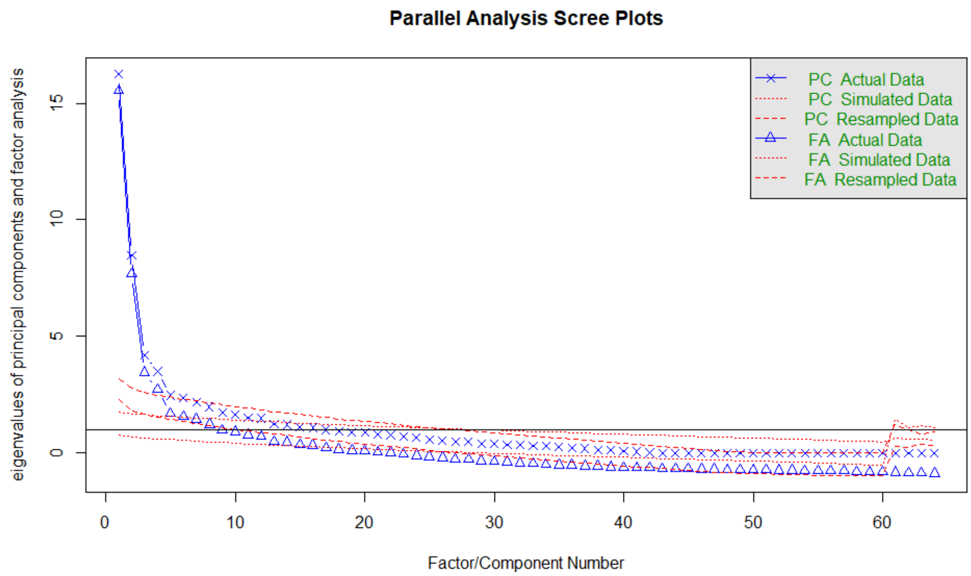

Figure 2. Polychoric Parallel analysis for Emotion Checklist items.

Based on a visual inspection of the scree plot from the actual data, we decided to estimate 4, 5, and 6 factor solutions, each of which was estimated in Mplus version 8 using a mean- and variance-adjusted weighted least squares (WSMLV) EFA with Geomin rotation to address the ordinal nature of the data. Based on interpretability, we felt the six-factor solution best modeled the data. Due to the sheer number of items, to aid interpretation, we eliminated nine items that cross-loaded one more than one factor at a loading level of 0.4 or higher, as well as one item that did not load on any factor at the 0.4 level. This solution generated six factors corresponding to negative emotions about teachers/classmates, positive emotions about teachers/classmates, negative emotions about the self, positive emotions about the self, negative emotions about the math, and positive emotions about the math.

Table 2. Inter-factor correlations of Emotion Checklist Items.

	Negative Emotions about Teachers/ Classmates	Positive Emotions about Teachers/ Classmates	Negative Emotions about the Self	Positive Emotions about the Math	Positive Emotions about the Self	Negative Emotions about the Math
Negative Emotions about Teachers/Classmates	1					
Negative Emotions about Teachers/Classmates	0.24	1				
Negative Emotions about the Self	**0.311**	0.155	1			

Positive Emotions about the Math	-0.004	**0.293**	-0.082	1		
Positive Emotions about the Self	0.208	**0.289**	0.053	**0.322**	1	
Negative Emotions about the Math	**0.293**	0.137	0.245	0.095	**0.204**	1

Note. Bold text indicates correlations significant at the $p < .05$ level.

Confirmatory Psychometric Properties

We next aimed to examined whether these models for the Likert items and the emotion checklist showed measurement invariance by running the model extracted in the Fall, 2018 on data from Spring 2019 using categorical confirmatory factor analyses (CFA) with factor loadings as specified in each of the exploratory analyses for Likert items and the emotion checklist.

Sample. The data were collected from first-year high school mathematics classrooms from fifteen teachers across Southwest and Mid-Atlantic US in Spring 2019. 690 students had complete responses on all 16 emotion checklist items for each of the four emotion objects as well as the 16 Likert items. 45.1% of these students were from the Southwest, and 54.9% were from the Mid-Atlantic.

Likert Items. We examined the fit of the factor structure of the 16 items using the Spring 2019 dataset. Specifically, we used a categorical CFA with polychoric correlations and WLSMV estimation using Mplus version 8 (Muthén & Muthén, 2017). Factors were identified by setting the variance of each latent variable equal to unity, thus standardizing our CFAs. Fits are indicated below.

Table 3. CFA fit statistics for 15 item Likert model

Confirmatory Model	X^2	df	RMSEA	RMSEA 90% CI	CFI	TLI	WRMR
5 Factor Model	462.16	80	0.083	[0.076, 0.091]	0.963	0.952	1.341

In general, this revised model fit the data adequately, per some sources, on the basis of RMSEA, as the value was <0.1 (Browne & Cudeck, 1993). CFI and TLI were also now very good by standard cutoffs (Tucker & Lewis, 1973; Bentler, 1990). Interitem correlations show that these facets of engagement are interrelated, yet distinct—each contributing variability to the final factor structure.

Table 4. CFA inter-factor correlations for the Likert items

	Task-Efficacy	Effort	Social Engagement	Instrumentality	Situational Interest
Task-Efficacy	1				
Effort	**0.270**	1			
Social Engagement	**0.623**	**0.520**	1		

Instrumentality	**0.390**	**0.324**	**0.588**	1	
Situational Interest	**0.586**	**0.397**	**0.682**	**0.754**	1

Note. Bold text indicates correlations significant at the $p < .05$ level.

Reliability as assessed by Cronbach's Alpha for each of the Likert Subscales was good to very good: *Efficacy* ($\alpha = 0.796$), *Effort* ($\alpha = 0.765$), *Social Engagement* ($\alpha = 0.755$), *Instrumentality* ($\alpha = 0.798$), and *Situational Interest* ($\alpha = 0.857$).

Emotion Items

We then examined the fit of the factor structure of the emotion items. Specifically, we used a Categorical CFA with polychoric correlations and WLSMV estimation using Mplus version 8 (Muthén & Muthén, 2017). Fit statistics are indicated below.

Table 5. CFA fit statistics for the emotion items

Confirmatory Model	X^2	df	RMSEA	RMSEA 90% CI	CFI		TLI	WRMR
6 Factor Model	1778.69	1310	0.023	[0.020, 0.025]	0.883		0.877	1.194

In general, this revised model again fit data well based on the RMSEA (Steiger & Lind, 1980, CFI and TLI were also a little higher, although still a bit low by standard cutoffs (Tucker & Lewis, 1973; Bentler, 1990). Recent work using Monte-Carlo simulations of SEM analyses suggests that for smaller sample sizes (less than 500 records), the fit statistics, CFI and TLI, are negatively biased, while RMSEA tends to be positively biased (Shi et al., 2018). As the number of free parameters increases, the relative bias in these estimates becomes more pronounced. The TLI, in particular is affected by number of parameters relative to sample size. Given a $\frac{\chi^2}{df}$ ratio of 1.36, well under the recommended ratio of 3, and our excellent RMSEA, we judge this model to show relatively good fit despite lower estimated values of CLI and TLI (Hu & Bentler, 1999).

Table 6. CFA inter-factor correlations for Emotion subscales, using 54 items.

	Negative Emotions about Teachers/ Classmates	Positive Emotions about Teachers/ Classmates	Negative Emotions about the Self	Positive Emotions about the Math Activity	Positive Emotions about the Self	Negative Emotions about the Math Activity
Negative Emotions about Teachers/Classmates	1					
Negative Emotions about Teachers/Classmates	**0.324**	1				
Negative Emotions about the Self	**0.527**	**0.270**	1			

Positive Emotions about the Math Activity	0.180	**0.433**	0.094	1		
Positive Emotions about the Self	**0.329**	**0.536**	**0.329**	**0.446**	1	
Negative Emotions about the Math Activity	**0.446**	0.115	**0.606**	**0.161**	0.090	1

Reliability as assessed by Cronbach's Alpha for each of the Emotion subscales was good to very good *Negative Emotions about Teacher/Classmates* (α = 0.655), *Positive Emotions about Teacher/Classmates* (α = 0.817), *Negative Emotions about the Self* (α = 0.807): *Positive Emotions about Self* (α = 0.807), *Negative Emotions about the Math* (α=0.700), and *Positive Emotions about the Math Activity* is comprised of (the sum of) eight items (α = 0.749).

Discussion

Using experience sampling methods to create and pilot an assessment of mathematical engagement, we were able to meet our two most important research goals: 1) to design an instrument practical that could be administered in under 5 minutes, targeting tasks that have the potential to be interesting and engaging for students, and 2) to maintain multidimensionality across each of the 4 dimensions of engagement: Affective (emotion checklist, interest), behavioral (Effort), cognitive (Efficacy, Instrumentality), and social (Social Engagement) dimensions. It has long been known that Interest (personal and situational), Efficacy (task- and subject-specific Self-Efficacy), Effort and feelings of Instrumentality are highly related, and that each is implicated in the interpretation of one's experience and the decisions one makes when engaging in challenging content like mathematics (Wiezel et al., 2019). Research subsequent to the development of this instrument has shown that these dimensions of task-level engagement are related to teacher and peer support and are predictive of longer-term motivation in mathematics as well as achievement (Middleton et al., 2023).

In particular, one innovation we have achieved in this effort is the ability to assess the impact of emotion/object interactions. Previously, emotions have been assessed using similar checklists, but object, which is critical to the interpretation of experience and the learner's response to emotional information, has not been studied in mathematics classrooms. The emotion checklist, because it pairs basic emotions with objects that are seen as causes of the emotion (e.g., being frustrated with the mathematics, or proud of oneself) allows for hypotheses about the ways in which emotional experiences are interpreted in academic tasks, and their differential impact on task-level motivation and behavior (see Middleton, et al., 2023). Our research has shown that differential patterns among the emotional objects show that students look to different cues in interpreting their experiences, wherein positive math emotions appear to increase interest and efficacy beliefs about the tasks, while negative emotions are associated with decreased interest and efficacy. Negative emotions about the mathematics appeared to be negatively associated with social engagement and feelings of instrumentality. Study of the impact of task-level emotions on engagement is in its infancy, but the work we are doing with this instrument is proving to be fruitful in this regard.

Such a short assessment of such a complex set of behaviors and attitudes has many limitations including lack of comprehensive coverage of the constructs that make up engagement, reliability of self-reports, and others. In addition, the transient nature of in-the- moment attitudes and emotions precludes using such an instrument as a diagnostic tool. Instead, we recommend a mixed-methods approach, wherein the actions that give rise to students' interpretations of their experiences are captured, and conjectures can be tested *over time*, such that the overall pattern of students' engagement may be recorded and coupled with potential causal factors.

The full instrument and administration guidelines may be obtained by writing the corresponding author.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1661180. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238–246.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.), Testing structural equation models (pp. 136-162). Newbury Park, CA: Sage.

Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. Journal of Personality Assessment, 85(1), 17–32. doi: 10.1207/s15327752jpa8501_02

Csikszentmihaly, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. Journal of Nervous & Mental Disease, 175, 525-536.

Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59-109.

Goldin, G. A. (2002). Affect, meta-affect, and mathematical belief structures. In G. Leder, E. Pehkonen, & G. Torner (Eds.), Beliefs: A hidden variable in mathematics education? (pp. 59–72). Dordrecht, The Netherlands: Kluwer.

Goldin, G. A. (2014). Perspectives on emotion in mathematical engagement, learning, and problem solving. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), Handbook of emotions in education (pp. 391–414). New York, NY: Taylor & Francis.

Middleton, J. A., Jansen, A., & Goldin, G. A. (2017). The complexities of mathematical engagement: Motivation, affect, and social interactions. In J. Cai (Ed.), Compendium for Research in Mathematics Education (pp. 87- 119). Reston, VA: National Council of Teachers of Mathematics.

Middleton, J. A., Mangu, D., & Lee, A. (2019). A longitudinal study of mathematics and science motivation patterns for stem-intending high schoolers in the us. In M. Hannula, G. C. Leder, F. Morselli, M. Vollstedt, and Q. Zhang (Eds), Affect And Mathematics Education. New York: Springer.

Middleton, J. A., Wiezel, A., Jansen, A., & Smith, E. P. (2023). Tracing mathematics engagement in the first year of high school: relationships between prior experience, observed support, and task-level emotion and motivation. ZDM–Mathematics Education, 55(2), 427-445.

Muthén, L. K., & Muthén, B. O. (2017). Mplus. Los Angeles: Muthén & Muthén.

Rimm-Kaufman, S., Baroody, A., Larsen, R., Curby, T., & Abry, T. (2015). To what extent do teacher–student interaction quality and student gender contribute to fifth graders' engagement in mathematics learning? Journal of Educational Psychology, 107(1), 170-185.

Shernoff, D. J. (2013). Measuring student engagement in high school classrooms and what we have learned. Optimal Learning Environments to Promote Student Engagement. New York, NY: Springer New York.

Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. School Psychology Quarterly, 18(2), 158-176.

Shiffman, S., Stone, A.A., & Hufford, M.R. (2008). Ecological momentary assessment. Annual Review of Clinical Psychology, 4, 1-32.

Steiger, J. H., & Lind, J. C. (1980, May). Statistically-based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1–10.

Wang, M.-T., 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.

Wiezel, A., Middleton, J. A., Zhang, Z. V., Grimm, K., & Jansen, A. (2019). Interest and Emotion Predictors of Motivation in Secondary Mathematics Classrooms. In Proceedings of the 43rd annual meeting of the International Group for the Psychology of Mathematics Education. Pretoria, South Africa: University of Pretoria.