Learning and Instruction 43 (2016) 16–26

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

The Math and Science Engagement Scales: Scale development, validation, and psychometric properties

Ming-Te Wang [a,*], Jennifer A. Fredricks [b,1], Feifei Ye [a], Tara L. Hofkens [a,1], Jacqueline Schall Linn [a]

[a] University of Pittsburgh, USA
[b] Connecticut College, USA

ARTICLE INFO

Article history:
Received 24 June 2015
Received in revised form
12 January 2016
Accepted 15 January 2016
Available online 5 February 2016

Keywords:
Student engagement
Math and science engagement
Multidimensionality
Multi-informant
Bifactor model

ABSTRACT

There is an urgent need to develop appropriate instruments to measure student engagement in math and science for the fields of research and practice. The present study developed and validated student- and teacher-report survey measures of student engagement in math and science. The measures are built around a multidimensional perspective of engagement by using a bifactor modeling approach. The sample was recruited from an ethnically and socioeconomically diverse middle and high school student population in the United States. The findings confirmed that student engagement is comprised of multiple related yet distinct dimensions, with evidence to support a bifactor structural model. There was also empirical evidence supporting measurement invariance and predictive validity. The results demonstrate the soundness of the psychometric properties of the Math and Science Engagement Scales.

Active engagement in math and science classes is a key contributing factor to adolescents' academic success and selection of college majors and careers in science, technology, engineering, and mathematics (STEM) (Maltese & Tai, 2010; Wang & Degol, 2014b). Research shows a decline in math and science engagement during the secondary school years, especially among low-income and minority youths (Martin, Way, Bobis, & Anderson, 2015). In order to increase student engagement in math and science and identify students who have the highest risk for opting out of the STEM pipeline, we need to conceptualize and measure *"student engagement"* appropriately. Unfortunately, research in this area has been hindered by inconsistencies in both the definition and measurement of the student engagement construct (Greene, 2015; Sinatra, Heddy, & Lombardi, 2015). Despite these variations, there is growing consensus that engagement is a multidimensional construct that includes behavioral, emotional, and cognitive components (Fredricks, Blumenfeld, & Paris, 2004; Wang, Willett, & Eccles, 2011). However, current self-report measures do not capitalize on what a multidimensional conceptualization of engagement can offer. In particular, there are only a handful of self-report student engagement measures that include multidimensional indicators, especially in math and science domains (see Kong, Wong, & Lam, 2003, for one exception). Moreover, the extent of psychometric support for these measures is very limited (Appleton, Christenson, & Furlong, 2008; Fredricks & McColskey, 2012; Greene, 2015).

Developing appropriate instruments to measure math and science engagement is urgently needed for both research and practice. The limited number of validated self-report measures that take a multidimensional perspective has made it difficult to examine predictors and consequences of each type of engagement, and investigate how these dimensions develop and interact over time. This impedes our ability to identify those students most at risk for disengaging from math and science classes and to design more targeted and nuanced interventions for enhancing student engagement in math and science learning. The present study addresses these gaps in the literature by using a bifactor modeling approach to test the psychometric properties of two newly developed student- and teacher-report survey measures focusing on math and science domains. The measures were initially developed through a mixed methods research design using an ethnically and socioeconomically diverse middle and high school student sample (see Fredricks et al., 2016; this issue for more information).

* Corresponding author. 230 South Bouquet Street, Pittsburgh, 15213, PA, USA.
E-mail address: mtwang@pitt.edu (M.-T. Wang).
[1] The second and third authors made equal intellectual contributions to the manuscript.

http://dx.doi.org/10.1016/j.learninstruc.2016.01.008
0959-4752/

1. Multifaceted nature of student engagement

This study builds upon self-system motivation theory, which assumes that engagement results from an interaction of the individual with the context and is responsive to variations in contextual characteristics (Connell, 1990). The experiential quality of the learning activity provides adolescents with information about themselves as being competent to succeed, as being related to others in these settings, and as being autonomous learners (Eccles, Wigfield, & Scheifele, 1997). This information cumulates to influence adolescents' engagement across various educational activities, as well as future educational and career aspirations. Over time, these reciprocal, cyclical processes shape the educational achievement and choices linked to these aspirations.

Drawing on the self-system motivation theoretical framework, engagement refers to the observable and unobservable qualities of student interactions with learning activities (Deci & Ryan, 2000). In this study, we included four dimensions of engagement: behavioral, emotional, cognitive, and social engagement. These four components of student engagement are dynamically embedded within the individual and operate at multiple levels—the school level, the subject area/specific classroom setting level, and the moment-to-moment activity level (Wang & Degol, 2014b). Given our interest in understanding the relationship between student engagement and STEM outcomes, we focused on engagement in math and science classroom settings.

The most prevalent conceptualization in the literature suggests that engagement consists of three distinct, yet interrelated components: behavioral, emotional, and cognitive engagement (Fredricks et al., 2004). *Behavioral engagement* is defined in terms of involvement in academic and class-based activities, presence of positive conduct, and absence of disruptive behavior (Fredricks et al., 2004). Previous survey studies have measured behavioral engagement with items about attention, participation, concentration, homework completion, and adherence to classroom rules (Fredricks & McColskey, 2012). *Emotional engagement* is conceptualized as the presence of positive emotional reactions to teachers, peers, and classroom activities, as well as valuing learning and having interest in the learning content (Finn, 1989; Voelkl, 1997). Emotional engagement has been measured with items about students' emotional reactions such as interest, enjoyment, and the perceived value of learning (Fredricks & McColskey, 2012). *Cognitive engagement* is defined in terms of self-regulated learning, using deep learning strategies, and exerting the necessary cognitive strategies for the comprehension of complex ideas (Zimmerman, 1990). Cognitive engagement has been measured with items about the use of shallow and deep learning strategies to learn and understand material, self-regulation, and persistence (Greene, 2015).

In addition to the three components of engagement most often included in prior studies, we added a social engagement dimension to reflect findings from our qualitative interviews with students about the meaning of engagement (see Fredricks et al., 2016; this issue). In these interviews, adolescents viewed engagement in social domains as an integral part of their learning in math and science classrooms. *Social engagement* includes the quality of social interactions with peers and adults, as well as the willingness to invest in the formation and maintenance of relationships while learning.

Previous research has shown that student engagement is a strong predictor of academic performance and choice (Hughes, Luo, Kwok, & Loyd, 2008). Students with higher behavioral and emotional engagement tend to attain higher grades and aspire for higher education (Wang & Holcombe, 2010). The use of self-regulatory and metacognitive strategies is associated with academic achievement (Pintrich & DeGroot, 1990). Students who enjoy, value, and feel competent in their social interactions are more likely to enlist the support of others for academic tasks. Students who want to form positive relationships with their peers are also more likely to have high academic achievement (Kiefer & Ryan, 2011; Wang & Eccles, 2013). Moreover, youths' interests in and beliefs about the importance of math and science are associated with intentions to enroll in elective STEM courses and career aspirations within STEM-related fields (Wang, 2012; Watt et al., 2012).

2. Measurement of student engagement

In a recent review of survey measures of engagement, Fredricks and McColskey (2012) identified only 3 out of 14 self-report survey measures that had scales assessing multiple dimensions of engagement. Items used to measure different dimensions of engagement were used inconsistently across behavioral, emotional, and cognitive dimensions, and the choice of items often did not match the theoretical conceptualizations of these constructs. For example, some measures included effort as an indicator of behavioral engagement to reflect compliance with required work in school, while others included effort as an indicator of cognitive engagement to describe the degree of psychological investment in learning. The wide variation in both the measurement and operationalization of engagement has made it challenging to compare findings across studies and draw conclusions about both the precursors and outcomes of engagement (Fredricks et al., 2004).

The majority of the survey measures (9 out of 14) focused on general engagement in school rather than engagement in specific subject areas. They excluded self-report measures of engagement in math or science that incorporate the multidimensional concept identified in the review. An extensive body of research suggests that motivational constructs can be domain specific, especially constructs that are situation- and subject-relevant (Guthrie & Wigfield, 2000). Some preliminary research also supports the domain specificity of student engagement, though more research is necessary to determine how this construct differs across subject areas (Martin, 2008). For example, Sinatra et al. (2015) contends that epistemic cognition, involvement in math and science practices, topical emotions, and attitudes are domain-specific aspects of science engagement that are important to consider.

Although researchers have conceptualized student engagement as a multidimensional construct, many studies have failed to examine the unique contributions of each dimension of engagement, as well as the general construct of engagement. Therefore, it is unclear if we can separate the unique contributions of the individual dimensions from the effects of the general construct. The uncertainty of distinguishing between the general construct and the individual dimensions makes it difficult to test both simultaneously (Chen, Jing, Hayes, & Lee, 2013). The bifactor model approach has recently been proposed to test the psychometric properties of the psychological constructs that are comprised of multiple related yet distinct dimensions (Chen et al., 2013). A bifactor model will allow us to examine if there is a global engagement factor that accounts for the commonality shared by the four dimensions. Additionally, it allows the investigation of whether there are multiple distinct factors that account for the unique contribution of the specific engagement dimension above and beyond the global engagement factor (Aguado et al., 2015). The bifactor model also enables us to test the association of an outcome variable with the global factor, and the unique contributions of the specific factors that are distinct from the global factor (Chen et al., 2013).

Furthermore, few valid teacher report measures were identified in the review. Collecting information from teachers can provide an alternative and important perspective on student engagement. Having information on student engagement from multiple informants will facilitate the assessment of predictive validity, and will help identify the unique contributions of each source that may explain the variance in student achievement outcomes. Finally, the diversity of participants needs to be expanded. The use of more ethnically and socioeconomically diverse samples is important for understanding whether some dimensions of engagement are more important than others for enhancing math and science engagement among minority youths and students from low-income families. This information is critical for designing interventions aimed at increasing the representation of minority and low-income youths in STEM courses and careers.

To address these limitations, the present study aims to develop and validate student- and teacher-report survey measures of student engagement in math and science that are built around a multidimensional perspective of engagement (i.e., behavioral, emotional, cognitive, and social engagement). We expand upon existing research by examining the goodness-of-fit for a bifactor model of student math and science engagement. The specific validation activities include confirmatory factor structure and dimensionality analyses, measurement invariance test by grade, gender, and SES levels, and tests of predictive validity. The sample was recruited from secondary schools with a socioeconomically and ethnically diverse student population in the United States.

3. Methods

3.1. Sample and procedure

Participants included middle school and high school students and their respective math and science teachers. These students and teachers were recruited from six public school districts in Western Pennsylvania. The student sample included 3883 6th through 12th graders (17.5% 6th grade 18.8% 7th grade, 19.4% 8th grade, 12.9% 9th grade, 10.9% 10th grade, 11.3% 11th grade, and 9.2% 12th grade). The student sample was 52.1% female, 66.1% European American, 23.8% African American, 7.2% multiracial, and 2.9% Asian American. Approximately 38.2% of the student sample qualified for free or reduced price lunch.

The teacher sample included 65 middle school teachers and 65 high school teachers. The teacher sample was 59.1% female and 96.7% European American. Fifty percent of the teachers taught math, 46.9% taught science, and 3.1% taught both math and science. Sixty-one percent of the teachers had their master's degree, 35.8% had their bachelors' degree, and 3.7% had their PhD degree. On average, they had 12.2 years of teaching experience, ranging from a new teacher to over 35 years of experience.

At each school, we first described the study to math and science teachers and obtained their consent accordingly. Students who agreed to participate in the study were provided with a computer-based survey and completed the math and science engagement survey during their regular instruction time in school. In addition to student engagement information, the survey also asked students to report on their future career aspirations related to STEM using scales commonly implemented in national surveys (e.g., NELS and Add Health). After students completed the survey, we randomly selected five students who participated in the study to be rated by their teachers on engagement in math or science classes through completion of a computer-based survey. In total, teachers reported on the engagement of approximately 300 students in math and science. The demographic characteristics of the 300 students were not significantly different from the 3883 students who completed the student report survey. Responses were confidential to the researchers and identification codes were used rather than names. Student demographic data was collected from school records and teacher demographic data was collected from teacher self-report surveys. Student math and science course grades in the fall and spring were also gathered from school records.

We used a sequential mixed-methods process in designing our engagement scale (see Fredricks et al., 2016; for more information on methodology). We first reviewed relevant literature to assemble conceptualizations of student engagement and existing instruments from which potential items might be borrowed or adapted for use in math and science. Second, we conducted open-ended interviews with secondary school math and science teachers and student focus groups and interviews to learn how they conceptualized “engagement” in math and science class. The interview data elicited a comprehensive list of indicators for the construct of math and science engagement. Third, we developed an initial list of 45 student-report items and 32 teacher-report items to reflect the multidimensionality of student engagement. Fourth, we subjected our items to an expert validation procedure in order to ensure that the items still corresponded to the construct of student engagement. Finally, we used a cognitive pretesting procedure with secondary school students and teachers to ensure that the items were comprehended and interpreted as we intended (see Karabenick et al., 2007 for a description of cognitive pretesting). This process resulted in 33 student-report and 20 teacher-report items. The teacher survey did not include as many questions about emotional and cognitive engagement as the student survey since emotional and cognitive indicators are internal and harder to evaluate by teachers. The teacher-report items were worded from the teachers' perspective, but were otherwise comparable to the student-report items.

3.2. Data analytic strategy

The factor structures of student and teacher items were evaluated separately with confirmatory factor analysis (CFA). We evaluated whether the math and science engagement could be modeled respectively with a bifactor model that contains a global latent factor of engagement and four specific uncorrelated dimensions (behavioral, cognitive, emotional, and social). The bifactor model was compared to and hypothesized to fit better than a second-order CFA model which measures the overall engagement factor as a second-order latent factor of four first-order factors.

The student-report engagement scale has 33 items (17 positively worded and 16 negatively worded) that are all ordinal on a 5-point Likert scale. The teacher-report engagement scale has 20 items (14 positively worded and 6 negatively worded) on a 5-point Likert scale. Approaches for modeling method effect due to negatively worded items include using one method factor for positively worded or negatively worded items, or both positive and negative method factors (Wang, Chen, & Jin, 2014). For the student-report items, we used a method factor approach to account for the relationship among positively worded items. This approach provided the best model fit and interpretability of factors. The method factor was not modeled for the teacher items as there were fewer negatively worded items. Additionally, neither the positive nor the negative method factor improved the model fit. The positive method factor is uncorrelated with the general or the four specific engagement factors. In summary, we compared the bifactor model and the second-order CFA for the teacher-report items, while we compared the bifactor model and the second-order CFA, both with a positive method factor, for the student-report items (see Fig. 1 for a comparison of models for the student-report items). To identify the bifactor models, the variances of the latent factors were set to 1, and all factors were uncorrelated with each other. Items with

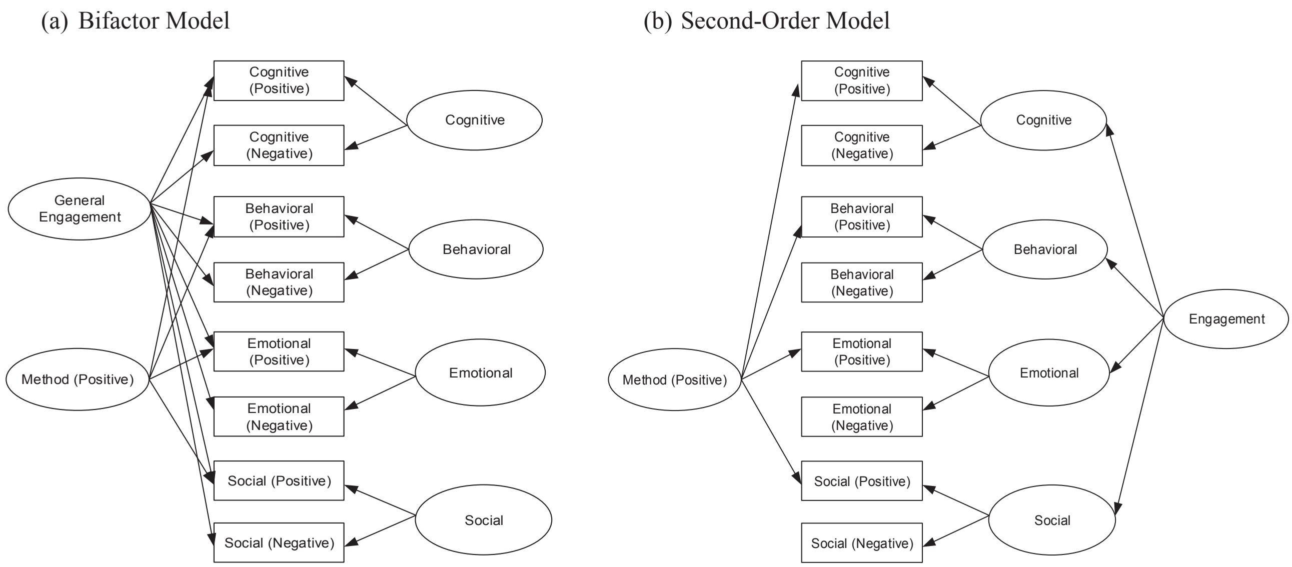

Fig. 1. Models tested for the student-report engagement scale: bifactor model with positive method factor and second-order factor model with positive method factor. Individual items were grouped for illustrative purposes only.

standardized factor loading values ≥.3 were considered to be strong indicators for the general and specific factors (Reise, Scheinese, Widaman, & Haviland, 2013).

We chose the means and variance adjusted weighted least squares estimation method for ordinal variables (WLSMV in *Mplus*; B. Muthén, du Toit, & Spisic, 1997), as WLSMV could provide accurate test statistics, parameter estimates, and standard errors (Flora & Curran, 2004). Since the sampling unit is the classroom instead of the student, the model chi-square statistic and standard error of model estimates were adjusted to account for the dependence of students within the same math or science classroom using sampling weight (Asparouhov, 2006).

In evaluating goodness of fit of the hypothesized models, we used a set of four fit indexes that focused on different aspects of model fit: chi-square statistic, Comparative Fit Index (CFI), Tucker Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA). A model that adequately fits the data will have: CFI ≥ .95, TLI ≥ .95, and RMSEA ≤ .06 (Hu & Bentler, 1999).

We tested measurement invariance by gender (male vs. female), race (European Americans vs. African Americans), grade (middle vs. high school), and SES levels (regular vs. free/reduced lunch). Multi-group CFA was conducted to compare models with and without constraints on item loadings (metric invariance). For identification, one factor loading was constrained to be 1 for each factor while the factor variance was freely estimated. The χ^2 difference test statistic has been shown to be highly sensitive to even trivial differences between the actual and modeled covariance matrices if sample size is large (Tomarken & Waller, 2003). Therefore, we examined the difference in the value of the CFI and RMSEA to assess model fit; a difference larger than .01 in the CFI and a difference larger than .015 in the RMSEA indicates a meaningful difference in model fit for testing measurement invariance (Chen, 2007; Cheung & Rensvold, 2002).

To examine predictive validity, we conducted multiple regression analyses in structural equation modeling (SEM) to investigate the extent to which the global engagement factor and the four dimensions of engagement could predict subsequent math and science course grades and STEM major aspirations differentially (e.g., how likely are you to pursue a college major in STEM fields?).

4. Results

4.1. Descriptive statistics

Item descriptive statistics are presented in Table 1 for the student- and teacher-report math and science engagement items.

4.2. Dimensionality test with confirmatory factor analysis (CFA)

4.2.1. Student report survey

Fit indices for the bifactor model and the second-order factor model are presented in Table 2. As hypothesized, the bifactor model with the positive method factor provided a better fit to the data than the second-order factor model with the positive method factor (math: ΔCFI = .020; ΔRMSEA = .008; science: ΔCFI = .019; ΔRMSEA = .007). The standardized factor loadings for the bifactor model are displayed in Table 3. For the general engagement factor, all loading values reached statistical significance (math: M = .52, range = .15–.75; science: *M* = .50, range = .14–.77) and were all above .30 except three social engagement items (items 27, 29, and 33), supporting a strong general engagement factor for both math and science. Loading values were also varied among items for the different specific factors. The items with the lowest loadings on the general engagement factor were from the social engagement factor (math: range = .15–.50; science: range = .14–.50).

For the specific factors, loadings were, in general, lower than the loadings on the general engagement factor, except the social engagement factor. Specifically, the loadings of the cognitive engagement factor (math: M = .18, range = −.01–.43; science: M = .23, range = .03–.43) were significantly lower than their general engagement factor loadings (math: M = .51, range = .41–.61; science: M = .48, range = .41–.59). This pattern holds for behavioral engagement items in that specific loadings (math: M = .18, range = −.07–.42; science: M = .18, range = .01–.37) were lower than the general loadings (math: M = .62, range = .39–.75; science: M = .60, range = .39–.77). The emotional engagement items had slightly lower specific loadings (math: M = .42, range = .09–.64; science: M = .44, range = .17–.67) when compared to their general loadings (math: M = .58, range = .48–.72; science: M = .55, range = .44–.71). The social

Table 1
Item descriptive statistics for student- and teacher-report math and science engagement scales.

	Student report engagement				Teacher report engagement			
	Math (n = 3883)		Science (N = 3883)		Math (N = 282)		Science (N = 300)	
	M	SD	M	SD	M	SD	M	SD
Cognitive Engagement								
1. I go through the work for science/math class and make sure that it's right.	3.78	.99	3.72	1.00	3.45	1.28	3.37	1.29
2. I think about different ways to solve a problem.	3.67	1.09	3.47	1.11	3.21	1.27	3.22	1.19
3. I try to connect what I am learning to things I have learned before.	3.88	1.08	3.80	1.10				
4. I try to understand my mistakes when I get something wrong.	4.21	.96	4.09	1.02	3.80	1.24	3.51	1.23
5. I would rather be told the answer than have to do the work (rev)	3.59	1.30	3.46	1.29	3.22	1.32	3.23	1.30
6. I don't think that hard when I am doing work for class (rev)	3.70	1.19	3.71	1.15				
7. When work is hard, I only study the easy parts (rev)	3.98	1.11	3.99	1.09				
8. (S) do just enough to get by (rev)/(T) do more than required in class.	3.15	1.25	3.17	1.24	2.58	1.26	2.67	1.35
Behavioral Engagement								
9. I stay focused	3.72	1.09	3.69	1.08	3.51	1.27	3.31	1.34
10. I put effort into learning science/math	4.12	.96	4.12	.94	3.73	1.17	3.65	1.17
11. I keep trying even if something is hard.	3.99	1.00	3.95	1.00	3.40	1.27	3.39	1.33
12. I complete my homework on time	4.10	1.13	4.18	1.09	3.81	1.37	3.60	1.420
13. I talk about science/math outside of class	2.77	1.26	2.89	1.29				
14. (S) don't participate in class (rev)/(T) participate in class.	4.12	1.10	4.16	1.08				
15. I do other things when I am supposed to be paying attention (rev)	3.84	1.11	3.85	1.09	3.51	1.31	3.35	1.26
16. If I don't understand, I give up right away (rev)	4.18	.98	4.21	.97				
Emotional Engagement								
17. I look forward to science/math class.	3.13	1.39	3.33	1.34	3.39	1.10	3.28	1.10
18. I enjoy learning new things about science/math.	3.38	1.31	3.79	1.23	3.53	1.10	3.72	1.11
19. I want to understand what is learned in science/math class.	4.30	1.01	4.35	.97	4.01	1.14	3.87	1.09
20. I feel good when I am in science/math class.	3.37	1.27	3.48	1.19				
21. I often feel frustrated in science/math class (rev)	3.30	1.31	3.48	1.24	3.40	1.14	3.79	1.01
22. I think that science/math class is boring (rev)	3.33	1.40	3.53	1.31				
23. I don't want to be in science/math class (rev)	3.69	1.38	3.82	1.32				
24. I don't care about learning science/math (rev)	4.33	1.09	4.34	1.07				
25. I often feel down when I am in science/math class (rev)	3.98	1.19	4.10	1.11				
26. I get worried when I learn new things about science/math (rev)	3.87	1.22	3.91	1.18				
Social Engagement								
27. I build on others' ideas.	3.14	1.14	3.20	1.14	3.25	1.20	3.26	1.25
28. I try to understand other people's ideas in science/math class.	3.58	1.11	3.62	1.09	3.59	1.10	3.41	1.22
29. I try to work with others who can help me in science/math	3.73	1.18	3.76	1.17	3.67	1.10	3.55	1.12
30. I try to help others who are struggling in science/math	3.50	1.23	3.44	1.23	3.37	1.29	3.33	1.26
31. I don't care about other people's ideas (rev)	4.25	.99	4.29	.96				
32. When working with others, I don't share ideas (rev)	4.09	1.04	4.10	1.04	3.94	1.00	3.91	1.11
33. I don't like working with classmates (rev)	4.09	1.17	4.17	1.11				

Note. (rev) indicates reverse coded items; (S) refers to student item only; (T) refers to teacher item only.

Table 2
Fit Statistics of three models for Student- and Teacher-Report Math and Science Engagement Scales.

	Model	*df*	χ^2	CFI	TLI	RMSEA (90% CI)
Student	Math					
	Bifactor	445	4099.14***	.957	.949	.047 (.046, .049)
	2nd order CFA	474	5810.97***	.937	.930	.055 (.054, .057)
	Science					
	Bifactor	445	3552.99***	.956	.948	.044 (.042, .045)
	2nd order CFA	474	4978.55***	.937	.929	.051 (.050, .052)
Teacher	Math					
	Bifactor	149	282.53***	.988	.985	.057 (.047, .067)
	2nd order CFA	166	353.26***	.983	.981	.064 (.055, .073)
	Science					
	Bifactor	149	248.99***	.995	.993	.051 (.039, .062)
	2nd order CFA	166	383.15***	.989	.987	.071 (.062, .080)

Note. $^{***}p < .001$; Comparative Fit Index (CFI), Tucker Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA).

engagement items had higher specific loadings (math: M = .43, range = .23–.52; science: M = .43, range = .23–.55) than their general loadings (math: M = .35, range = .15–.52; science: M = .35, range = .14–.48). In summary, the general factor accounted for a large part of the variance of cognitive and behavioral engagement items, but not for social and emotional engagement items. All factor loadings on the positive method factor were moderate to strong and statistically significant (math: range = .21–.48; science: range = .28–.50).

4.2.2. Teacher report survey

Similar to the student-report items, the bifactor model provided a better fit to the data than the second-order factor model (see Table 2; math: ΔCFI = .005; ΔRMSEA = .007; science: ΔCFI = .006; ΔRMSEA = .020). The standardized factor loadings for the bifactor model are displayed in Table 3. For the general engagement factor, all loading values were above .30 and statistically significant (math: M = .79, range = .55-.92; science: *M* = .78, range = .35-.96), supporting a strong general engagement factor.

For the specific factors, loadings were, in general, lower than the loadings on the general engagement factor. Specifically, the loadings of the cognitive engagement factor (math: M = .23, range = .12–.43; science: M = .18, range = .00–.36) were significantly lower than their general factor loadings (math: M = .85, range = .80–.89; science: M = .87, range = .80–.91). This pattern holds for the behavioral engagement factor in that specific loadings (math: M = .25, range = .09–.41; science: M = .13, range = −.01–.28) were lower than the general loadings (math: M = .85, range = .73-.92; science: M = .90, range = .79–.96). Similarly, the emotional engagement factor had lower specific

Table 3
Standardized factor loadings of the bifactor measurement model for student- and teacher-report math and science engagement scales (math items/science items).

	Student			Teacher	
	General engagement	Specific engagement	Method (POS)	General engagement	Specific engagement
Cognitive Engagement					
1. I go through the work for science/math class and make sure that it's right.	.59/.51	*.02*/.12	.42/.47	.89/.91	.25/.18
2. I think about different ways to solve a problem.	.43/.42	−.06/.05	.42/.49	.79/.86	.43/.36
3. I try to connect what I am learning to things I have learned before.	.41/.43	*−.01/.03*	.47/.50		
4. I try to understand my mistakes when get something wrong.	.52/.48	.08/.19	.40/.45	.92/.90	.19/.19
5. I would rather be told the answer than have to do the work (rev)	.61/.59	.30/.32		.80/.80	*.12/.18*
6. I don't think that hard when I am doing work for class (rev)	.45/.45	.39/.40			
7. When work is hard I only study the easy parts (rev)	.60/.58	.43/.43			
8. (S) do just enough to get by (rev)/(T) do more than required in class.	.44/.41	.30/.31		.87/.89	*.14/.00*
Behavioral Engagement					
9. I stay focused	.63/.60	.42/.37	.30/.30	.87/.93	.35/.28
10. I put effort into learning science/math	.68/.66	.19/.16	.44/.44	.92/.96	.14/*.06*
11. I keep trying even if something is hard	.66/.63	.08/.11	.38/.40	.91/.95	*.09/−.01*
12. I complete my homework on time	.51/.45	.40/.33	.21/.28	.84/.88	.25/*.10*
13. I talk about science/math outside of class	.39/.39	*.00/.01*	.36/.41		
14. (S) don't participate in class (rev)/(T) participate in class	.64/.64	*.04/.03*			
15. I do other things when I am supposed to be paying attention (rev)	.66/.65	.38/.37		.73/.79	.41/.22
16. If I don't understand, I give up right away (rev)	.75/.77	*−.07/.05*			
Emotional Engagement					
17. I look forward to science/math class.	.49/.44	.64/.67	.44/.45	.80/.68	.40/.33
18. I enjoy learning new things about science/math.	.52/.48	.49/.50	.45/.48	.86/.88	.24/.22
19. I want to understand what is learned in science/math class.	.53/.48	.09/.17	.41/.37	.82/.90	.32/.26
20. I feel good when I am in science/math class.	.57/.50	.52/.55	.39/.39		
21. I often feel frustrated in science/math class (rev)	.51/.50	.34/.36		.61/.55	*−.03/.05*
22. I think that science/math class is boring (rev)	.66/.61	.56/.58		.79/.62	.32/.53
23. I don't want to be in science/math class (rev)	.68/.66	.58/.60			
24. I don't care about learning science/math (rev)	.72/.71	.28/.33			
25. I often feel down when I am in science/math class (rev)	.62/.61	.41/.41			
26. I get worried when I learn new things about science/math (rev)	.48/.46	.27/.27			
Social Engagement					
27. I build on others' ideas.	.20/.22	.39/.37	.44/.43	.74/.77	.45/.32
28. I try to understand other people's ideas in science/math class.	.35/.36	.44/.44	.48/.45	.78/.77	.51/.51
29. I try to work with others who can help me in science/math	.15/.14	.52/.55	.37/.44	.56/.51	.36/.56
30. I try to help others who are struggling in science/math	.52/.48	.23/.23	.38/.40	.80/.79	.29/.37
31. I don't care about other people's ideas (rev)	.50/.50	.50/.47			
32. When working with others, I don't share ideas (rev)	.46/.45	.44/.43		.55/.35	.22/.58
33. I don't like working with classmates (rev)	.26/.28	.50/.51			

Note. All factor loadings were significant at $p < .05$ except those in italic; (rev) indicates reverse coded items; (S) refers to student item only; (T) refers to teacher item only.

loadings (math: M = .25, range = −.03–.40; science: M = .28, range = .05–.53) when compared to the general engagement factor loadings (math: M = .78, range = .61–.86; science: M = .73, range = .55–.90). The social engagement items also had lower specific loadings (math: M = .37, range = .22–.51; science: M = .47, range = .32–.58) than the general loadings (math: M = .69, range = .55–.80; science: M = .64, range = .35–.79). In summary, the general engagement factor accounted for a larger part of variances for teacher-report engagement items in all four domains.

4.3. Reliability

4.3.1. Student report survey

We concluded with a bifactor model with four specific factors for both math and science engagement items. We estimated Cronbach's alpha for the four subscales and the scale of overall engagement (see Table 4). Cronbach's alpha for the overall scale was high. Moderate to high reliability was found for each of the four subscales.

4.3.2. Teacher report survey

We concluded with a bifactor model with four specific factors for both math and science engagement items. Table 4 shows that

Table 4
Cronbach's alpha for the student- and teacher-report math and science engagement scales.

	General engagement	Cognitive engagement	Behavioral engagement	Emotional engagement	Social engagement
Student-report engagement scale					
Math	.93	.75	.82	.89	.74
Science	.92	.76	.81	.89	.73
Teacher-report engagement scale					
Math	.97	.92	.93	.87	.86
Science	.97	.93	.94	.85	.86

Cronbach's alphas for the overall scale and subscales were high.

4.4. Measurement invariance

4.4.1. Student report survey

The bifactor model fit the data well for the entire sample and for each group (see Table 5). The metric invariance held by gender, race, grade, and SES for both math and science engagement items as the constrained model with equal factor loadings all fit better than the unconstrained (baseline) models with larger CFI and TLI, and smaller RMSEA.

4.4.2. Teacher report survey

The bifactor model fit the data well for the entire sample and within each group (see Table 6). Similar to the student-report items, metric invariance held by gender, race, grade and SES for both math and science items as the constrained model with equal factor loadings all fit better with larger CFI and TLI, and smaller RMSEA, except for grade on math engagement items.

4.5. Predictive validity

4.5.1. Student report survey

Structural equation modeling analysis assessing the predictive validity of the engagement scales is presented in Table 7. Students scoring higher on general math and science engagement were more likely to score higher on math and science achievement (math: $\beta = .45$, $p < .001$; science: $\beta = .39$, $p < .001$). Similarly, students with higher general engagement in math and science were also more likely to report intentions to pursue STEM college majors (math: $\beta = .36$, $p < .001$; science: $\beta = .41$, $p < .001$). In addition, each specific engagement factor predicted math and science course grades and STEM major aspirations differentially. For instance, science behavioral engagement factor was the strongest predictor of science course grade, while science emotional engagement factor was the strongest predictor of STEM career aspiration. Overall, the general engagement factor is more predictive of STEM outcomes than specific engagement factors.

4.5.2. Teacher report survey

Students scoring higher on general math and science engagement were more likely to score higher on math and science achievement (math: $\beta = .61$, $p < .001$; science: $\beta = .63$, $p < .001$). Regarding STEM career aspirations, students with higher general engagement in math and science were also more likely to report intentions to pursue STEM college majors (math: $\beta = .16$, $p < .001$; science: $\beta = .22$, $p < .001$). Moreover, each specific engagement factor predicted math and science course grades and STEM major aspirations differentially. Again, overall, the general engagement factor is more predictive of STEM outcomes than specific engagement factors.

4.6. Convergence and divergence between student and teacher reports

Student report and teacher report were moderately correlated on general engagement ($r = .45$). Among the four dimensions of engagement, student report and teacher report were more highly correlated on behavioral and cognitive engagement (math: $rs = .57–.45$; science: $rs = .47–.45$) while they were less correlated on emotional and social engagement (math: $rs = .34–.21$; science: $rs = .39–.28$). Regarding the correlations between student engagement and educational outcomes, teacher report was more strongly correlated with course grades than student report on the general engagement and specific engagement dimensions. Student report was more strongly correlated with career aspirations than teacher report across the four specific dimensions and general engagement (see Table 8).

5. Discussion

Improving student engagement in math and science is critical for academic and professional success in STEM fields. Unfortunately, educators and researchers lack reliable and valid instruments to accurately assess adolescents' engagement in math and science learning activities. This study makes a pivotal contribution through the development of robust and multidimensional student- and teacher-report survey measures for assessing middle and high school students' engagement in math and science. This is the first study in the student engagement literature demonstrating that a bifactor model fit the engagement factorial structure well. Our bifactor model suggests that student math and science engagement are characterized by a global engagement construct, as well as four unique dimensions: behavioral, cognitive, emotional, and social engagement. By accounting for the global engagement and multiple specific dimensions of adolescents' engagement in

Table 5
Test of measurement invariance for the student-report math and science engagement scales.

Math	Model	df	χ2	CFI	TLI	RMSEA (90% CI)	Δχ2	Δdf	ΔCFI	ΔRMSEA
Gender (male vs. female)	1. Baseline Model	890	4523.06***	.960	.953	.047 (.046, .048)				
	2. Metric invariance	967	3233.37***	.975	.973	.036 (.034, .037)	140.63***	77	–	–
Race (European American vs. African American)	1. Baseline Model	890	4031.25***	.963	.957	.044 (.042, .045)				
	2. Metric invariance	967	3256.03***	.973	.971	.036 (.034, .037)	251.94***	77	–	–
Grade (middle vs. high school)	1. Baseline Model	890	4075.74***	.961	.953	.044 (.043, .045)				
	2. Metric invariance	967	3201.71***	.972	.970	.035 (.034, .037)	199.63***	77	–	–
SES (regular vs. free/reduced lunch)	1. Baseline Model	890	3696.75***	.965	.959	.043 (.042, .044)				
	2. Metric invariance	967	3122.70***	.973	.971	.036 (.035, .038)	273.40***	77	–	–
Science										
Gender (male vs. female)	1. Baseline Model	890	4074.10***	.958	.950	.044 (.043, .046)				
	2. Metric invariance	967	2981.21***	.973	.971	.034 (.032, .035)	127.84***	77	–	–
Race (European American vs. African American)	1. Baseline Model	890	3843.07***	.960	.953	.043 (.041, .044)				
	2. Metric invariance	967	3027.18***	.972	.970	.034 (.033, .036)	199.13***	77	–	–
Grade (middle vs. high school)	1. Baseline Model	890	3667.69***	.959	.951	.041 (.040, .043)				
	2. Metric invariance	967	3155.75***	.968	.965	.035 (.034, .037)	271.27***	77	–	–
SES (regular vs. free/reduced lunch)	1. Baseline Model	890	3417.62***	.963	.956	.041 (.040, .043)				
	2. Metric invariance	967	2840.84***	.973	.970	.034 (.033, .035)	213.33***	77	–	–

Note. ***$p < .001$. – indicates that the more constrained model provides better fit than the less constrained model and thus difference test cannot be performed; Comparative Fit Index (CFI), Tucker Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA).

Table 6
Test of measurement invariance for the teacher-report math and science engagement scales.

Math	Model	df	χ2	CFI	TLI	RMSEA (90% CI)	Δχ2	Δdf	ΔCFI	ΔRMSEA
Gender (male vs. female)	1. Baseline Model	298	467.37***	.986	.982	.064 (.053, .075)				
	2. Metric invariance	333	508.09***	.986	.984	.062 (.051, .073)	75.01***	35	–	–
Race (European American vs. African American)	1. Baseline Model	298	475.02***	.984	.980	.066 (.055, .077)				
	2. Metric invariance	333	483.57***	.986	.985	.057 (.046, .048)	57.94**	35	–	–
Grade (middle vs. high school)	1. Baseline Model	298	424.68***	.990	.988	.056 (.043, .067)				
	2. Metric invariance	333	517.48***	.986	.984	.064 (.053, .074)	96.54***	35	.004	.008
SES (regular vs. free/reduced lunch)	1. Baseline Model	298	500.67***	.980	.975	.077 (.065, .088)				
	2. Metric invariance	333	469.31**	.987	.985	.060 (.047, .072)	44.09	35	–	–
Science										
Gender (male vs. female)	1. Baseline Model	298	446.68***	.993	.991	.062 (.050, .073)				
	2. Metric invariance	333	469.48***	.994	.993	.056 (.044, .067)	62.59**	35	–	–
Race (European American vs. African American)	1. Baseline Model	298	427.07***	.994	.992	.058 (.045, .070)				
	2. Metric invariance	333	403.24**	.997	.996	.040(.023, .054)	37.88	35	–	–
Grade (middle vs. high school)	1. Baseline Model	298	414.36***	.994	.993	.055 (.041, .067)				
	2. Metric invariance	333	458.13***	.994	.993	.054 (.041, .065)	68.48**	35	–	–
SES (regular vs. free/reduced lunch)	1. Baseline Model	298	409.74***	.991	.989	.059 (.044, .073)				
	2. Metric invariance	333	441.63***	.991	.990	.055 (.040, .069)	58.74**	35	–	–

Note. *$p < .05$; **$p < .01$; ***$p < .001$. – indicates that the more constrained model provides better fit than the less constrained model and thus difference test cannot be performed; Comparative Fit Index (CFI), Tucker Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA).

Table 7
Standard estimates of latent regression on the general and specific factors of the math and science engagement.

	Course grade				STEM career aspiration			
	Student report		Teacher report		Student report		Teacher report	
	Math	Science	Math	Science	Math	Science	Math	Science
General Engagement	.45***	.39***	.61***	.63***	.36***	.41***	.16*	.22**
Behavioral Engagement	.20***	.27***	.03	.27*	−.14***	−.10*	.27*	.13
Emotional Engagement	.03	−.03	−.04	.05	.08**	.16***	.29**	.16*
Cognitive Engagement	−.07	.05	.05	−.04	−.10**	.03	.07	.06
Social Engagement	−.13***	−.07**	−.34*	−.16	−.01	−.02	−.11	−.11
R^2	.26	.23	.49	.48	.18	.20	.16	.08

Note. *$p < .05$; **$p < .01$; ***$p < .001$.

Table 8
Correlation between course grades, STEM career aspiration, and student and teacher engagement measures.

	Science					Math				
	Student–teacher correlation	Correlation with course grade		Correlation with career aspiration		Student–teacher correlation	Correlation with course grade		Correlation with career aspiration	
		Student	Teacher	Student	Teacher		Student	Teacher	Student	Teacher
General Engagement	.45	.31	.59	.43	.24	.45	.39	.57	.36	.22
Cognitive engagement	.45	.38	.61	.38	.23	.45	.39	.58	.26	.18
Behavioral Engagement	.47	.40	.60	.39	.23	.57	.48	.59	.33	.25
Social Engagement	.28	*.07*	.44	.23	.21	.21	.19	.42	.30	.18
Emotional Engagement	.39	.26	.52	.40	.23	.34	.31	.49	.32	.21

Note. All correlations are significant at $p < .01$ except the italic one. Only students that had both student-report and teacher-report were included.

math and science, we are able to identify the types of engagement that most accurately predict STEM-related outcomes. In addition, this identification enables us to test for additive and interactive effects among various factors. Furthermore, we are able to investigate the association of STEM-related outcomes with the global engagement construct, as well as the unique contribution of the specific dimensions to STEM outcomes that are distinct from the global engagement construct.

5.1. Dimensionality and predictive validity

Consistent with the recent literature, our findings support student engagement as a multidimensional construct (Reschly & Christenson, 2012; Wang & Degol, 2014b). The results of our analysis demonstrate that there are four theoretically distinct dimensions of engagement, and do not support recent suggestions to consider student engagement on a continuum rather than using a dimensional perspective (Sinatra et al., 2015). A multidimensional perspective on student engagement provides a richer characterization of how students act, feel, think, and socialize with others in math and science classrooms, rather than considering each of the dimensions separately. Approaching engagement as a multidimensional construct also allows us to consider the impact of each dimension of engagement separately on math and science outcomes.

A bifactor model that includes behavioral, emotional, cognitive, and social engagement, provides a good fit for the data in both student- and teacher-reports in math and science. The four dimensions of math and science engagement form a general factor of global engagement which captures the common variances shared by the four dimensions of the engagement. On the other hand, each dimension of student engagement represents specific factors, which capture their unique variances, above and beyond the global factor of student engagement. These specific dimensions are also differentially predictive of academic achievement and educational aspirations, independent of the global engagement factor.

These findings support the perspective that behavioral, emotional, cognitive, and social engagement are conceptually related to each other at the global construct level, but also represent distinct and unique constructs. We see contextually dependent relevance in both these models, depending on whether scholars aim to measure engagement along with separate dimensions or more generally. For example, many important questions remain about how each dimension of engagement contributes to academic outcomes and how the dimensions function to shape a student's overall engagement in math and science. The multidimensional scales make it possible to test the relations between each type of engagement and potential outcomes in theoretical models and create different student engagement profiles. Moreover, the multidimensional scales can be used in educational settings or interventions to identify and target specific dimensions of engagement among students with low global engagement. In contrast, a global measure of student engagement may be sufficient for testing policy relevant questions related to the outcomes of STEM engagement.

Another important contribution of this study is the inclusion of a social engagement scale, which reflects social interactions around instructional content and affective reactions to peers. In reform-based math and science classrooms, students have extensive opportunities to work in groups, share ideas, and explain their learning. Our findings contribute to a broader understanding of student engagement by underscoring the importance of the classroom social environment and identifying a mechanism by which the quality of social interactions during math and science classwork can influence students' achievement in these subjects. This mechanism can likewise be applied to motivation in pursuing STEM careers.

5.2. Measurement invariance

This study provides empirical evidence to support measurement invariance by gender, grade level, socioeconomic status, and race. It suggests that the content of most items were perceived and interpreted similarly across these demographic groups. Establishing measurement invariance is vital to developing a highly generalizable metric of engagement that can be used to predict academic outcomes for students from an array of demographic groups (Wang et al., 2011; Widaman & Resie, 1997). There has been a growing concern that certain groups of students (e.g., girls, students from low SES families) may be at greater risk for disengaging from math and science learning (Wang & Degol, 2014a). Establishing measurement invariance to ensure that measures of engagement operate similarly across groups allows researchers to make more appropriate comparisons between groups such as boys and girls and those from different SES status groups.

5.3. Convergence and divergence between student and teacher reports

This study also contributes to the literature by developing both student and teacher self-report measures of student engagement. The use of multiple informants to assess student engagement provides a more comprehensive perspective of this construct and enables us to examine the congruence and divergence between student and teacher perceptions of student engagement across different types of engagement. To our knowledge, this is the first study to examine the correspondence between student and teacher reports of student engagement in math and science domains.

The classroom is a shared learning context in which teachers and students are jointly focused on the behavioral and cognitive dimensions of academic activities. Even in classrooms in which students are not globally engaged, students' and teachers' understanding or perceptions of what it means to be engaged is informed by years of socialization into the linguistic, social , and cultural norms of educational environments (Wang & Degol, 2014b; Wang & Eccles, 2013). Indicators of behavioral and cognitive engagement are also more easily observed in the classroom (Skinner, Kindermann, & Furrer, 2008). Thus, it is not surprising that teachers and students are more congruent in their behavioral and cognitive engagement reports in math and science classes. At the same time, however, classrooms are complex social contexts and teachers may not have access or insight into how individual students feel about class or working with peers (Wang & Eccles, 2012). Students regulate their emotions in a variety of ways, including masking negative emotions from teachers or suppressing positive emotions about the class from peers. Furthermore, emotional experience and expression is highly individual and variable, making it difficult for teachers to accurately assess emotional engagement in each of their students over the course of each school day. Similarly, it can be challenging to assess students' social engagement with peers, which is multiply determined within the broader classroom social climate and unfolds fluidly over the course of academic activity.

Taken together, teachers and students reports on engagement could be additive and complementary. Given that the data from each reporter contributed to explaining variation in student achievement outcomes, combined with the moderate relationship between student and teacher reports of student engagement ($r = .45$ for both math and science), it is likely that neither student nor teacher reports alone present a complete picture of student engagement. The unique experiences and perceptions of students and teachers position each stakeholder to have different insights into student engagement and its contribution to learning outcomes. For example, teachers can observe how students' cognitive effort is manifested, which may play a role in how students are graded. Students have insight into their affective reactions, values, and attitudes toward math and science, which may inform the development of aspirations to pursue STEM in the future. When combined, student and teacher reports may provide a more complete, accurate, and balanced picture of each dimension of students' engagement, which can inform how we think about, assess, and intervene to improve student achievement in math and science.

5.4. Domain specificity

A final contribution of this study is validation of a measure that was developed specifically to assess engagement in math and science domains. In the past, one approach has been to adapt general engagement measures for use in specific classes. This approach is limited because it does not take into account domain specific aspects of engagement and assumes engagement is manifested similarly across different subject areas (Fredricks et al., 2016; Sinatra et al., 2015). Some of the indicators, especially in the behavioral domain (e.g., participation, ask questions), are similar to items in previous general engagement measures (Fredricks & McColskey, 2012). However, other indicators of engagement, such as emotions (e.g., frustration) and social engagement (e.g., build on others' ideas) are either unique to math and science or reflected differently in these domains. Incorporating domain-specific and differentiated measures of student engagement is critical to determine the extent to which student engagement represents a general tendency and the extent to which it is content specific (Sinatra et al., 2015).

5.5. Limitations, future research, and conclusion

It is important to interpret the findings of this study in light of the following limitations.

First, this study relied exclusively on survey methods. The integration of multiple methods, such as interviews, experience sampling methods, and observations to assess engagement could prove valuable in its ability to holistically explore the construct (Sinatra et al., 2015; Wang & Degol, 2014a). Second, teachers rated five students only on engagement, possibly leading to dependence among ratings from the same teacher. Future studies should obtain a larger sample of teachers to rate a greater number of students, allowing a multilevel factorial structure to be conducted, thereby addressing potential clustering effects. Finally, future studies need to determine the test-retest reliability of the engagement scales, whether or not scores are variable over time, and whether the scales are sensitive to changes in the learning environment.

Ultimately, in order to effectively meet schools' increasing demand for student math and science engagement, educators and researchers need a richer conceptualization and measurement of the construct. Our study positively contributes to this objective by providing empirical evidence supporting the psychometric properties of the Math and Science Engagement Scales. We anticipate this measure will be of interest to scholars investigating the contextual predictors and academic consequences of math and science engagement. We also anticipate this measure will be useful for teachers interested in identifying students at risk for math and science disengagement.

Acknowledgment

This study was supported by the National Science Foundation Grant 1503181.

References

Aguado, J., Luciano, J. V., Cebolla, A., Serrano-Blanco, A., Soer, J., & Garcia-Campayo, J. (2015). Bifactor analysis and construct validity of the five facet mindfulness questionnaire (FFMQ) in non-clinical Spanish samples. *Frontiers in Psychology, 6*, 1–14.

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: critical conceptual and methodological issues of the construct. *Psychology in the Schools, 45*, 369–386.

Asparouhov, T. (2006). General multilevel modeling with sampling weights. *Communications in Statistics: Theory and Methods, 35*, 439–460.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 464–504.

Chen, F. F., Jing, Y., Hayes, A., & Lee, J. M. (2013). Two concepts or two approaches? A bifactor analysis of psychological and subjective well-being. *Journal of Happiness Study, 14*, 1033–1068.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indices for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 233–255.

Connell, J. P. (1990). Context, self, and action: a motivational analysis of self-system processes across the life-span. In D. Cicchetti (Ed.), *The self in transition: From infancy to childhood* (pp. 61–97). Chicago: University of Chicago Press.

Deci, E., & Ryan, R. M. (2000). What is the self in self-directed learning? Findings from recent motivational research. In G. Straka (Ed.), *Conceptions of self-directed learning: Theoretical and conceptual considerations*. Munster: Waxmann.

Eccles, J. S., Wigfield, A., & Scheifele, U. (1997). Motivation to succeed. In W. Damon, & N. Eisenberg (Eds.) (5th ed.,*Social, emotional, and personality development: Vol. 3. Handbook of child psychology* (pp. 1017–1095). New York: Wiley.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59*, 117–142.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466–491.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59–109.

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: a comparative analysis of various methods and student self-report instruments. In S. Christenson, A. L. Reschy, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 319–339). New York: Springer.

Fredricks, J. A., Wang, M. T., Schall Linn, J., Hofkens, T. L., Sung, H. C., Parr, A. K., & Allerton, J. J. (2016). Using qualitative methods to develop a survey measure of math and science engagement. *Learning and Instruction*.

Greene, B. A. (2015). Measuring cognitive engagement with self-report scales: reflections over 20 years of research. *Educational Psychologist, 50*, 14–30.

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L.Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (3rd ed., pp. 403–422). New York: Longman.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55.

Hughes, J. N., Luo, W., Kwok, O., & Loyd, L. K. (2008). Teacher-student support, effortful engagement, and achievement: a three-year longitudinal study. *Journal of Educational Psychology, 1*, 1–14.

Karabenick, S. A., Wolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., et al. (2007). Cognitive processing of self-report items in educational research: do they think what we mean? *Educational Psychologist, 42*, 139–151.

Kiefer, S. M., & Ryan, A. M. (2011). What characteristics are associated with social success? Changes in students' perceptions of social success during early adolescence. *Applied Developmental Psychology, 32*, 218–226.

Kong, Q. P., Wong, N. Y., & Lam, C. C. (2003). Student engagement in mathematics: development of instrument and validation of construct. *Mathematics Education Research Journal, 15*, 4–21.

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.

Martin, A. J. (2008). Enhancing student motivation and engagement: the effects of a multidimensional intervention. *Contemporary Educational Psychology, 33*, 239–269.

Martin, A. J., Way, J., Bobis, J., & Anderson, J. (2015). Exploring the ups and downs of math engagement in the middle school years. *Journal of Early Adolescence, 35*, 199–244.

Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes* (Unpublished technical report).

Pintrich, P. R., & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33–40.

Reise, S. P., Scheinese, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: a bifactor perspective. *Educational and Psychological Measurement, 73*, 5–26.

Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: evolution and future directions of the engagement construct. In S. L. Christenson, A. L. Reschly, & A. Wylie (Eds.), *Handbook of research on student engagement* (pp. 1–19). New York: springer.

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologists, 50*, 1–13.

Skinner, E., Kindermann, T., & Furrer, C. (2008). A motivational perspective on engagement and disaffection: conceptualization and assessment of children's emotional in academic activities in the classroom. *Educational and Psychological Measurement, 69*, 493–525.

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology, 112*, 578–598.

Voelkl, K. E. (1997). Identification with school. *American Journal of Education, 105*, 204–319.

Wang, M. T. (2012). Educational and career interests in math: a longitudinal examination of the links between perceived classroom environment, motivational beliefs, and interests. *Developmental Psychology, 48*, 1643–1657.

Wang, W. C., Chen, H. F., & Jin, K. Y. (2014). Item response theory models for wording effects in mixed-format scales. *Educational and Psychological Measurement*. http://dx.doi.org/10.1177/0013164414528209.

Wang, M. T., & Degol, J. (2014a). Motivational pathways to STEM career choices: using expectancy-value perspective to understand individual and gender differences in STEM fields. *Developmental Review, 33*, 304–340.

Wang, M. T., & Degol, J. (2014b). Staying engaged: knowledge and research needs in student engagement. *Child Development Perspectives, 8*, 137–143.

Wang, M. T., & Eccles, J. S. (2012). Social support matters: longitudinal effects of social support on three dimensions of school engagement from middle to high school. *Child Development, 83*, 877–895.

Wang, M. T., & Eccles, J. S. (2013). School context, achievement motivation, and academic engagement: a longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction, 28*, 12–23.

Wang, M. T., & Holcombe, R. (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal, 47*, 633–642.

Wang, M. T., Willett, J. B., & Eccles, J. S. (2011). The assessment of school engagement: examining dimensionality and measurement invariance across gender and race/ethnicity. *Journal of School Psychology, 49*, 465–480.

Watt, H. M. G., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: a comparison of samples from Australia, Canada, and the United States. *Developmental Psychology, 48*, 1594–1611.

Widaman, K. F., & Resie, S. P. (1997). Exploring the measurement invariance of psychological instruments: applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: an overview. *Educational Psychologist, 21*, 3–17.