School Psychology Quarterly
2014, Vol. 29, No. 4, 517–535

1045-3830/14/$12.00 DOI: 10.1037/spq0000050

Measuring Engagement in Fourth to Twelfth Grade Classrooms: The Classroom Engagement Inventory

Ze Wang, Christi Bergin, and David A. Bergin
University of Missouri

Research on factors that may promote engagement is hampered by the absence of a measure of classroom-level engagement. Literature has suggested that engagement may have 3 dimensions—affective, behavioral, and cognitive. No existing engagement scales measure all 3 dimensions at the classroom level. The Classroom Engagement Inventory (CEI) was developed to fill this gap. In Study 1, exploratory and confirmatory factor analyses were conducted on data from 3,481 students from the 4th to 12th grade. The results suggested a 4-factor model of the CEI. Using these results, in Study 2 several items were revised and data were collected 1 year later from 4th to 12th grade students in the same school district as Study 1. Analyses were conducted on data from 3,560 students after data cleaning. A series of potential models was tested. The final results suggest a 5-factor 24-item CEI: (1) Affective Engagement, (2) Behavioral Engagement–Compliance, (3) Behavioral Engagement–Effortful Class Participation, (4) Cognitive Engagement, and (5) Disengagement. Results advance understanding of the construct of classroom engagement. The CEI fills a significant gap in measurement of engagement. The CEI is classroom level, measures multiple dimensions of engagement, uses self-report, is relatively short, and can be readily administered in classrooms from the 4th to 12th grade.

Keywords: engagement, factor analysis, measurement invariance, scale development

Supplemental materials: http://dx.doi.org/10.1037/spq0000050.supp

Many school reform efforts seek to increase student engagement because educators presume it is a key variable that drives the outcomes prioritized by the Federal Department of Education: increased achievement, enrollment in challenging courses, and reduced dropout rates. School reform efforts also seek to increase engagement because it is amenable to intervention (Finn & Zimmer, 2012).

Despite its role in predicting important outcomes and its potential malleability, there is a dearth of well-articulated, comprehensive measures of classroom-level engagement (Fredricks & McColskey, 2012; Fredricks et al., 2011). The purpose of this article is to fill this gap by discussing the development of the Classroom Engagement Inventory (CEI) in two sequential studies. The aim is to provide schools with a relatively short, but valid instrument, that they can use to inexpensively measure classroom engagement. We begin by defining engagement and then discuss issues related to its measurement.

Classroom Engagement

Classroom engagement refers to a student's active involvement in classroom learning activities (Reeve, Jang, Carrell, Jeon, & Barch, 2004; Skinner, Kindermann, & Furrer, 2009). This includes "attention, interest, investment, and ef-

This article was published Online First April 7, 2014.

Ze Wang, Department of Educational, School, and Counseling Psychology, University of Missouri; Christi Bergin, Assessment Resource Center, University of Missouri; David A. Bergin, Department of Educational, School, and Counseling Psychology, University of Missouri.

This research was partially funded by Grant URC–11–060 from the Research Council of the University of Missouri.

We thank Renée Jamroz for her assistance in identifying items in the preliminary instrument development, and Sara Finney for her support.

Correspondence concerning this article should be addressed to Ze Wang, Department of Educational, School, and Counseling Psychology, University of Missouri, 16 Hill Hall, Columbia, MO 65211. E-mail: wangze@missouri.edu

fort students expend in the work of learning" (Marks, 2000, p. 155) within the classroom. Fredricks, Blumenfeld, and Paris (2004) described the three dimensions of engagement—affective (emotional), cognitive, and behavioral. Research by Archambault and colleagues (Archambault, Janosz, Fallu, & Pagani, 2009) supported these dimensions. At the classroom level, *affective* engagement refers to positive emotions during class, such as interest, enjoyment, and enthusiasm (Fredricks et al., 2004; Skinner et al., 2009). *Behavioral* engagement refers to observable behavior such as time-on-task, overt attention, classroom participation, question asking, and choice of challenging tasks. *Cognitive* engagement refers to mental effort, such as meaningful processing, strategy use, concentration, and metacognition. There is a need for research that explores the multiple dimensions of classroom engagement simultaneously. As Fredricks et al. (2004) stated, "the vast majority of studies test the impact of a single type of engagement" (p. 83).

An important consideration in the measurement of engagement is the level at which the construct is measured. A seemingly identical construct can bear different meanings at different levels of generality. Engagement at the school level, for instance, is typically measured as bonding to school, participation in extracurricular activities, attendance, and number of referrals and suspensions. However, school engagement does not directly reflect classroom engagement. The same student may be highly engaged in one class, but not in others, and classrooms vary in average student engagement (Darr, 2012). Yet, many measures conflate classroom-level and school-level engagement (Fredricks et al., 2004). For example, Archambault et al. (2009) asked students about liking school, having fun at school, being interested in French class, and putting effort into French class. The first two are about school-level engagement and the latter two are about classroom-level engagement. Lawson and Lawson (2013) recommended that classroom engagement be distinguished from school engagement.

A commonly used measure of engagement is the Student Engagement Instrument (SEI), which has good evidence of reliability and validity (Appleton, Christenson, Kim, & Reschly, 2006; Betts, Appleton, Reschly, Christenson, & Huebner, 2010; Carter, Reschly, Lovelace, Appleton, & Thompson, 2012). However, the SEI is designed to measure school-level engagement. Sample items include, "Overall, adults at my school treat students fairly" and "What I'm learning in my classes will be important in my future." The SEI does not claim to measure behavioral engagement, and includes scales that go beyond engagement such as Future Aspirations and Goals and Family Support for Learning (Betts et al., 2010). In contrast, the CEI measures psychological, behavioral, and affective engagement at the classroom level.

School-level versus classroom-level measures of engagement may serve different functions and be appropriate in different contexts. Engagement should be measured at the same specificity level as the intervention and as other key variables. The measure should be classroom-level if the purpose is to (1) evaluate the effectiveness of an intervention enacted at the classroom level, (2) provide feedback to teachers regarding student perceptions of their classroom, (3) investigate what teachers can do in the classroom to improve engagement, or (4) investigate the link between engagement and learning in specific classes. For example, if increasing engagement in classrooms is the desired outcome of an intervention aimed at helping teachers use positive discipline or develop warm teacher–student relationships, then a classroom-level measure would be most appropriate. Measuring engagement at an inappropriate level can obscure results and make an intervention appear less successful.

A measure of classroom engagement will be most useful if it is invariant across groups such as lower and higher income, male and female, and younger and older students. If measures lack invariance, it may not be appropriate to compare results across groups. In contrast, invariance across age allows comparison of engagement at different ages cross-sectional and longitudinally. Invariance across subject matter allows comparison of engagement in different subjects. Other measures of engagement have proven invariant across gender and age (Archambault et al., 2009; Betts et al., 2010; Martin, 2009). In the present studies we replicate these results with classroom-level engagement, and also examine invariance across income groups and subject matter.

Preliminary Instrument Development

To develop the CEI, we reviewed articles purporting to measure engagement in school settings. The articles were found in computerized databases (Education Full Text, ERIC, and PsycINFO) and reference lists from other articles. Next, we extracted nonoverlapping items that closely aligned with the definition of engagement given earlier. Items that purported to measure engagement, but in fact measured other constructs, such as peer or teacher relationships or perceived relevance of school, were not extracted. This resulted in 35 items. Items were adapted to be classroom-level by including "In THIS class" at the beginning.

We searched for items that seemed to measure each of the three major dimensions. Behavioral engagement items were taken from the behavioral engagement and disengagement scales of Skinner et al. (2009) that were adapted from Wellborn (1992), such as, "In this class, I listen very carefully." Cognitive engagement items were adapted from Linnenbrink (2005) and Stipek and Gralinski (1996), which were adapted from Meece, Blumenfeld, and Hoyle (1988), such as, "I do my work without thinking too hard." Items also were adapted from a scale developed by the Motivation Research Group (Maehr & Pintrich, 2001; Pintrich & Maehr, 2002), such as, "In this class, when I do work, I ask myself questions to help me understand what to do." Finally, some items were taken from Archambault et al. (2009), such as, "I do not want to stop working at the end of class."

Few existing measures included items addressing affective engagement in classrooms. Research has shown that the Positive and Negative Affect Schedule Extended form or PANAS–X (Watson & Clark, 1994) is a valid and reliable measure; it is widely used in emotions research. Therefore, we took five affective items (Angry, Happy, Excited, Interested, Proud) from the PANAS–X. We generated five additional items, selecting only emotions that previous work has found youth experience in classrooms (embarrassed, smile or laugh, bored, fun, amused; C. Bergin & Bergin, 2009; Pekrun, Elliot, & Maier, 2009). This resulted in seven positive and three negative affect items. Other researchers have measured affective engagement with similar short scales, but relatively restricted sets of emotions, such as: "Class is fun, I enjoy learning, I feel good, I feel interested" (Skinner et al., 2009).

On the CEI, respondents indicated their level of agreement with each item using a 5-point Likert scale of frequency (25 items) using the following scale ranging from 1 (*never*), 2 (*hardly ever*), 3 (*monthly*), 4 (*weekly*), and 5 (*each day of class*) or a 7-point Likert scale (10 items) using the following scale ranging from 1 (*not at all true*) to 7 (*very true*). These different scales did not pose a problem for analysis because a latent response variable formulation was used to obtain polychoric correlations that were then analyzed (B. Muthén, 1984). More details are in the Results section of Study 1. All items were coded so that higher scores indicate higher levels of engagement. Six reverse-coded items were placed throughout the scale to reduce response bias.

After compiling the items, some were rewritten to remove difficult language and to shorten item length. We then asked eight students, primarily in fourth grade, to "think aloud" as they completed the survey, telling us how they understood the individual items. We also asked them to circle words they did not understand. If they did not circle a word that we were concerned they might not understand, we asked them to define it for us in their own words. This feedback resulted in minor modifications. In addition, in Study 1 we provided students with an "I don't know what this means" option on every item. Items with more than 1% of the students selecting this option were revised for Study 2. This process resulted in rewording or clarifying just a few items. For example, the word *amused* was not understood by some students in Study 1, so we changed the item from "I feel amused" to "I feel amused (smile, laugh, have fun)" for Study 2. A list of the resulting 35 items and their original sources is available from the authors. Additional constructs were measured and analyzed in Study 2 as validity variables.

Study 1

Method

Participants. Participants were 3,925 students from the fourth to 12th grade in a school district in a medium-sized city in Missouri (estimated Metropolitan Statistical Area of

174,327 people; U.S. Census Bureau, 2012) in the United States. Of the 3,222 students whose demographic data were available from the school district: 84.5% were White, 2.0% were Asian/Pacific Islander, 5.9% were Black, 5.5% were Hispanic, and 2.2% were American Indian/Alaska Native; 49.0% qualified for free or reduced lunch; and 51.6% were female. The survey included special education, remedial, and advanced students who were attending classes that varied in subject from gym, to remedial math, to AP literature.

Procedure. Data were gathered in April 2010. Teachers in all regular classrooms in 18 buildings (13 elementary, 3 middle schools, 1 high school, 1 career center) across the district administered the survey. Surveys were administered at the same time across an entire building to ensure that no student would be asked to take the CEI more than once and to ensure that a wide variety of subject areas was covered. Students were instructed to respond for the specific class (academic subject area) they were in at that moment. Students who were in study halls were asked to respond for the class they would attend next. The class identity was verified from a teacher survey. Students were told that the survey was voluntary; they could refuse to complete it. To improve quality of data, teachers read standardized instructions to students informing them of the importance of their honesty and assuring them that the results would be confidential. Demographic data were taken from school records. The project met institutional review board requirements at the university.

Surveys were completed on scannable sheets. Before scanning, all completed surveys were visually inspected for suspect patterns (e.g., making a perfect "X" shape across all the items on the page), vulgar statements, and nonresponse. Suspicious surveys were coded as missing and not analyzed. There were 444 (11%) surveys coded as missing due to suspect patterns or incompletion resulting in 3,481 completed surveys.

Results

Factor structure of the CEI. Histograms were used to examine the distribution of item responses. The factor structure of the CEI was then explored using Mplus (L. K. Muthén & Muthén, 2012). Polychoric correlations (Olsson, 1979) were used to examine associations between ordered categorical items. A full information maximum likelihood estimation (FIML) method is used in the calculations of polychoric correlations in Mplus. This method assumes missing at random and makes use of all available data. The lowest covariance coverage was 0.818, meaning that all variables and pairs of variables had data for at least 81.8% of the sample; the missing information is small. Factor analyses were based on those polychoric matrices. A range of factor models was determined based on multiple criteria (i.e., model fit indexes, modification indexes, residual correlations, factor loadings, number of indicators per factor). Some researchers have suggested conducting exploratory factor analysis (EFA) and then confirmatory factor analysis (CFA) with the same data to tease out methodological explanations for a lack of fit in CFA before substantive explanations are used (e.g., van Prooijen & van der Kloot, 2001). However, more researchers have recommended conducting EFA at the initial stage of scale development and then conducting CFA with a new sample for validity and generalization (e.g., Worthington & Whittaker, 2006). We did both. In Study 1, we performed EFA and CFA with the same sample. In Study 2, we tested the final CFA model (which was consistent with the EFA model) from Study 1 with new data. We also cross-validated the final CFA model from Study 2 with the data in Study 1. Details are in the Study 2 section.

Separate EFAs were conducted with different numbers of factors. Geomin rotation, an oblique rotation method, was used for EFA. Following EFA, we conducted CFAs. During this process, items with a pattern coefficient less than .50 or cross-loaded on multiple factors were removed, which resulted in removal of 14 items. For CFA, we used the recommended robust mean- and variance-adjusted weighted least squares (WLSMV) estimator (Beauducel & Herzberg, 2006; Finney & DiStefano, 2006). Multiple indexes were examined to assess model fit.

Whole sample EFA and CFA suggested that the CEI consisted of four factors. Examination of the items revealed that the four factors measured three different dimensions of engagement—affective, behavioral, and cognitive. In addition, a Disengagement factor emerged,

Table 1
Standardized Factor Loadings (Pattern Coefficients) in the Four-Factor Confirmatory Factor Analysis Model for the Classroom Engagement Inventory in Study 1

Item	Affective engagement (ω = .91)	Behavioral engagement (ω = .84)	Cognitive engagement (ω = .89)	Disengagement (ω = .86)
1. I feel excited (W&C; B&B)	.85			
2. I feel interested (W&C; B&B)	.85			
3. I feel happy (W&C; B&B)	.80[a]			
4. I have fun (W&C; B&B)	.80			
5. I feel proud (W&C; B&B)	.79			
6. I get really involved in class activities (S)		.82		
7. I actively participate in class discussions (P&V), (S)		.78		
8. I form new questions in my mind as I join in class activities (P&V)		.71		
9. I compare things I am learning with things I already knew (S&G), (D&C)		.64[a]		
10. I work with other students and we learn from each other (P&V)		.63		
11. If I make a mistake, I try to figure out where I went wrong (L)			.80	
12. I go back over things I don't understand (S&G)			.78	
13. I think deeply when I take quizzes in this class (P&V)			.75[a]	
14. I ask myself some questions as I go along to make sure the work make sense to me (S&G)			.76	
15. I search for information from different places and think about how to put it together (P&V)			.73	
16. If I'm not sure about things, I check my book or use other materials like charts (S&G)			.67	
17. I judge the quality of my ideas or work during class activities (P&V)			.65	
18. I try to figure out the hard parts on my own (P&V)			.56	
19. I am "zoned out"; not really thinking or doing class work (P&V)				.87[a]
20. I let my mind wander (S)				.82
21. I just pretend like I'm working (S)				.75

Note. All items begin with the stem "In this class." The eight items for Cognitive Engagement are on a 7-point scale. The other items are on a 5-point scale. Reliability coefficients calculated as McDonald's omega (McDonald, 1999). Sources from which items were adapted were: W&C; B&B = Watson and Clark (1994) and Bergin and Bergin (2009); S = Skinner et al. (2009); P&V = Painter and Valentine (1996); S&G = Stipek and Gralinski (1996); D&C = DeBacker and Crowson (2006); L = Linnenbrink (2005).
[a] Items had unstandardized loadings fixed to unity.

which included both behavioral and cognitive components. The standardized factor loadings (i.e., pattern coefficients) from CFA as well as the reliability coefficient (i.e., internal consistency) calculated as McDonald's omega (McDonald, 1999) are provided in Table 1. This four-factor model fit the data well, root mean square error of approximation (RMSEA) = .059, 90% CI [.057, .061], comparative fit index (CFI) = .965, Tucker–Lewis index (TLI) = .960 (see Table 2). However, correlations between the four factors ranged from .48 (between Behavioral Engagement and Disengagement) to .74 (between Behavioral Engagement and Cognitive Engagement). The moderate to high correlations suggest a possible second-order gen-

Table 2
Model Fit Statistics and Testing Measurement Invariance (Equal Factor Loadings and Thresholds) by Groups in Study 1

Model	No. of parameters	χ^2	*df*	RMSEA	90% CI		CFI	TLI
Whole sample (N = 3,477)	127	2,409.4*	183	.059	.057	.061	.965	.960
Age-level groups (n = 3,477), $\Delta\chi^2$ = 637.5, Δdf = 184, $p < .001$								
Configural invariance	381	2,875.0*	549	.060	.058	.063	.961	.955
Measurement invariance	197	3,213.4*	733	.054	.052	.056	.959	.964
Class subject groups (n = 2,488), $\Delta\chi^2$ = 528.1, Δdf = 276, $p < .001$								
Configural invariance	508	2,368.9*	732	.060	.057	.063	.966	.961
Measurement invariance	232	2,626.2*	1,008	.051	.048	.053	.966	.972
Gender groups (n = 2,906), $\Delta\chi^2$ = 282.3, Δdf = 92, $p < .001$								
Configural invariance	254	2,419.0*	366	.062	.060	.064	.963	.957
Measurement invariance	162	2,483.5*	458	.055	.053	.057	.963	.966
Free/reduced lunch status groups (n = 2,906), $\Delta\chi^2$ = 121.4, Δdf = 92, p = .022								
Configural invariance	254	2,392.3*	366	.062	.059	.064	.963	.957
Measurement invariance	162	2,250.8*	458	.052	.050	.054	.967	.970

Note. Sample sizes differed when different grouping variables were used due to missing values on grouping variables. The chi-square difference test is based on the DIFFTEST in Mplus for testing the difference between the configural invariance model and the measurement invariance model using the mean- and variance-adjusted weighted least squares estimator. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis index.
* $p < .001$.

eral construct of classroom engagement. When the four factors were allowed to load on a general Classroom Engagement factor, a second-order CFA model (i.e., the four factors load on a second-order latent factor) had good model fit (RMSEA = .066, 90% CI [.064, .068], CFI = .956, TLI = .950). However, the robust chi-square difference test (Asparouhov & Muthén, 2006) was statistically significant, $\Delta\chi^2(2) = 207.3$, $p < .001$, suggesting that for the classroom engagement construct, a hierarchical factor model was not as good as a first-order multidimensional model. (Note that the robust chi-square difference test is susceptible to sample size, just as the absolute chi-square tests are).

Measurement invariance and engagement across groups. To test whether the measure could validly be used with varied groups, we investigated measurement invariance across age level (three levels: elementary was 4th–5th grade, middle was 6th–8th grade, and high school was 9th–12th grade), class subject (math, language arts, science, and social studies), gender, and free/reduced lunch status (as a proxy for income) using multiple group CFAs. Again, polychoric correlations were used to describe associations between ordered categorical items. The existence of measurement invariance for factors measured by ordinal (i.e., ordered categorical) indicators involves equality constraints on factor loadings and item thresholds simultaneously across groups because constraints on thresholds and factor loadings are "in tandem because the item probability curve is influenced by both parameters" (L. K. Muthén & Muthén, 2012, p. 485). The constrained model (i.e., measurement invariance model) was compared with the unconstrained model when thresholds and factor loadings were free across groups (i.e., configural invariance model). We used the default delta parameterization in Mplus.

The four-factor CFA model fit the data well in the whole sample and within each group we examined. Table 2 shows model fit statistics for the whole sample and for the configural invariance and measurement invariance models for each of the four grouping variables (age level, class subject, gender, and free/reduced lunch status). All models had good model fit. Measurement invariance could be assumed across the groups based on the results. It should be noted that recommendations of changes in model fit indexes have been proposed for testing measurement invariance (Chen, 2007; Cheung & Rensvold, 2002). These recommendations include a decrease in CFI by at least .01, supplemented by an increase in RMSEA by at least .015 for testing factor loading invariance or for testing intercept or residual invariance. By these standards, measurement invariance still existed in this study and in Study 2 described later. However, these recommendations are based on simulation studies that used the maximum likelihood estimation. We used the WLSMV estimator instead in the current research. For the WLSMV estimator, Mplus provides a robust chi-square difference test between nested models (Asparouhov & Muthén, 2006), and these test results are in Table 2. However, this robust chi-square difference is susceptible to sample size, just as the absolute chi-square tests are. For more information on and an example of how different estimation methods affect measurement invariance testing and substantive research hypothesis testing, see Wang et al. (2014).

In summary, factor analysis in Study 1 supported the hypothesis that engagement has three distinctly measurable dimensions—affective, behavioral, and cognitive—and that this dimensionality was the same across age level, gender, free/reduced lunch status, and class subject. Results also suggest a Disengagement factor and a possible second-order unidimensional construct. From the original 35 items, 21 were identified as making significant, distinct contributions to the four factors. In Study 2, we made improvements to and modified the CEI.

Study 2

Method

Participants. After similar data cleaning steps as in Study 1, participants were 3,560, fourth to 12th grade students in the same school district and same 18 buildings as Study 1. Study 2 took place in spring 2011, the year following Study 1, with students responding regarding a different class. We were able to obtain demographic information of 2,524 students from the school district: 82.3% were White, 2.0% were Asian/Pacific Islander, 6.4% were Black, 6.1% were Hispanic, 2.4% were American Indian/

Alaska Native, and 0.8% reported multiple races; 57.5% qualified for free or reduced lunch; and 51.7% were female. The project met institutional review board requirements at the university.

Procedure. We modified the CEI before administering it again. Three items from the 2010 survey were combined into one item due to high polychoric correlations and overlapping meaning. The three old items were: (1) In this class, I feel amused; (2) In this class, I smile or laugh; and (3) In this class, I have fun. The new item was: "In this class, I feel amused (smile, laugh, have fun)." The deletions left two open affect items, which we filled with the following: (1) In this class, I feel anxious; and (2) In this class, I feel frustrated. This resulted in a balanced set of affect items, five positive and five negative. Another major change in the 2011 survey was that the order of the items was scrambled. In the 2010 survey, the affective, behavioral, and cognitive engagement items were placed together in the survey. Thus, the factor structure found in Study 1 may have been influenced by item position within the survey. The revised 2011 survey consisted of 35 items (25 items on a 5-point scale and 10 items on a 7-point scale). We included items that were eliminated in Study 1 to provide a second test of factor structure with scrambled items. Additional constructs were measured and analyzed as validity variables for the CEI.

Results

In all analyses, item responses were treated as ordered-categorical (ordinal) variables under which were continuous latent response variables (B. Muthén, 1984), as in Study 1. The lowest covariance coverage was 0.927, suggesting there was not much missing information in the dataset.

Step 1: Test the four-factor CFA model derived from Study 1. Data analysis started with the four-factor CFA model with 21 items, derived in Study 1. The initial four-factor CFA model from Study 1 fit the 2011 data satisfactorily (see Model 1 in Table 3). However, one item, "In this class, I compare things I am learning with things I already knew," had large residual correlations (a residual correlation is the difference between an observed correlation and the corresponding model implied correlation between two variables) with three other items and had a relatively low loading (.44) on its factor (Behavioral Engagement). With this item removed, the new model (Model 2) still had good fit with the data. These results were consistent with those in Study 1, supporting the hypothesis that classroom engagement has affective, behavioral, and cognitive dimensions.

Step 2: Justify deletion of items and consider adding items and/or factors. Because we suspected that items that were not included in the four-factor CFA might form new factors

Table 3
Model Fit Indexes in Study 2

Model	No. of parameters	χ^2	df	RMSEA	90% CI		CFI	TLI
1	127	3,053.3*	183	.066	.064	.069	.956	.950
2	122	2,396.8*	164	.062	.060	.064	.965	.959
3	151	5,366.2*	265	.074	.072	.075	.933	.925
4	147	10,417.8*	269	.103	.101	.105	.868	.853
5	155	4,693.2*	265	.069	.067	.070	.942	.934
6	141	2,849.8*	220	.058	.056	.060	.964	.959
7	146	3,214.9*	242	.059	.057	.061	.961	.956
7a	141	4,531.3*	247	.070	.068	.072	.944	.938
7b	146	2,834.0*	242	.056	.054	.057	.965	.960

Note. Model 1: four-factor CFA model obtained in Study 1. Model 2: Model 1 without item "I compare things I am learning with things I already knew." Model 3: Model 2 plus a "Negative Feelings" factor. Model 4: Model 2 with "Negative Feelings" items loading on "Affective Engagement." Model 5: Model 2 plus a "Compliance" factor with five items. Model 6: Model 2 plus a "Compliance" factor with three items. Model 7: Model 6 plus one additional item, "In this class, I do not want to stop working at the end of class" on the "Behavioral Engagement-Effortful Class Participation" factor. A five-factor CFA model. Model 7a: Model 7 with a second-order factor. Model 7b: Model 7 with Study 1 data. CFA = confirmatory factor analysis.
* $p < .001$.

and/or contribute to the four factors, we conducted additional analyses. Those additional analyses included EFA for the 14 items that were left out of the four-factor CFA model; CFA with more potential contributing items; EFA with the 21 items in the four-factor CFA model and potential contributing items. The purpose was twofold: (1) to justify the deletion of items from the final model and shorten the CEI so that it would be more economical for school districts to administer, and (2) to test the stability of the factor structure of the CEI that was found in Study 1. As in Study 1, multiple criteria were used in our decision making (i.e., model fit indexes, modification indexes, residual correlations, factor loadings, number of indicators per factor).

EFAs were conducted with the 14 items that were not included in the initial four-factor CFA model, to find out if there were factors that went undetected. Different numbers of factors were extracted in those EFAs. Examining those EFA results suggested two potential additional factors. One was a Negative Feelings factor with four items and the other was a Compliance factor with five items. To accommodate the potential Negative Feelings factor, we tested two models (Models 3 and 4 in Table 3) and, based on multiple criteria (model fit, residual correlations, and factor loadings) decided not to add this factor or to allow additional indicators of Affective Engagement. To accommodate the potential Compliance factor, we tested Models 5 and 6 (see Table 3). Model 6 has satisfactory model fit.

An additional EFA was conducted on the 15 items comprising three factors (out of five factors in Model 6): Disengagement, Behavioral Engagement–Effortful Class Participation, and Behavioral Engagement–Compliance. This was done because in Study 1 these items were presented in a block, which may have affected the factor structure. In the revised survey they were interwoven with other items. The EFA results suggested adding an additional item to the "Effortful Class Participation" factor. We then used CFA to test a final five-factor model (Model 7) that has satisfactory model fit with 24 items (RMSEA = .059, 90% CI [.057, .061], CFI = .961, TLI = .956). See Table 4 for factor loadings for this model. More details on model revisions are available from the authors.

Comparing Model 7 with Model 1 from Study 1, we found that three factors were exactly the same with the same sets of items. Those factors were Affective Engagement, Disengagement, and Cognitive Engagement. Another factor, Behavioral Engagement–Effortful Class Participation, was similar, with four of the five items overlapping. One item, "In this class, I compare things I am learning with things I already knew," was in Model 1 but not in Model 7. Another item, "In this class, I do not want to stop working at the end of class," was in Model 7 but not in Model 1. Model 7 also had an additional factor that was not in Model 1, which we have labeled Behavioral Engagement–Compliance.

The correlations among the five factors from Model 7 ranged from .56 (between Affective Engagement and Disengagement) to .90 (between Affective Engagement and Behavioral Engagement–Effortful Class Participation). We suspected that there may be a second-order factor. The second-order CFA model (Model 7a) came close to fitting satisfactorily (RMSEA = .070, 90% CI [.068, .072], CFI = .944, TLI = .938), but did not quite meet the criteria. The robust chi-square difference test between Models 7 and 7a was statistically significant, $\Delta\chi^2(5) = 583.0$, $p < .001$.

We also tried to fit Model 7 with data in Study 1. We find it interesting that the model fit was satisfactory (Model 7b; RMSEA = .056, 90% CI [.054, .057], CFI = .965, TLI = .950), suggesting that although four factors were relatively easy to detect for the construct of classroom engagement, a five-factor model was also a plausible one. The final model (Model 7) was a five-factor CFA. Its standardized factor loadings (i.e., pattern coefficients) as well as the reliability coefficient calculated as McDonald's omega (McDonald, 1999) and factor correlations are in Table 4.

Step 3: Test measurement invariance and compare latent means of groups. Measurement invariance across groups was then investigated using multiple group CFAs as in Study 1. Results suggested that measurement invariance existed between age level (elementary, middle, and high school), free/reduced lunch status, gender, and core class subject (math, science, language arts, and social studies; see Table 5). Results of likelihood ratio (LR) chi-square difference tests for each of the

Table 4
Factor Loadings (Pattern Coefficients) of Five-Factor Confirmatory Factor Analysis Model (Model 7) in Study 2

Item	Affective engagement (ω = .90; ICC = .16)	Behavioral engagement–compliance (ω = .82; ICC = .10)	Behavioral engagement–effortful class participation (ω = .82; ICC = .20)	Cognitive engagement (ω = .88; ICC = .15)	Disengagement (ω = .82; ICC = .18)
1. I feel interested (W&C; B&B)	.87				
2. I feel proud (W&C; B&B)	.85				
3. I feel excited (W&C; B&B)	.84				
4. I feel happy (W&C; B&B)	.78[a]				
5. I feel amused (smile, laugh, have fun) (W&C; B&B)	.68				
6. I listen very carefully (S)		.87[a]			
7. I pay attention to the things I am supposed to remember (S&G)		.82			
8. I complete my assignments (M&P; P&M)		.61			
9. I get really involved in class activities (S)			.82		
10. I form new questions in my mind as I join in class activities (P&V)			.69[a]		
11. I do not want to stop working at the end of class (A)			.68		
12. I actively participate in class discussions (P&V), (S)			.67		
13. I work with other students and we learn from each other (P&V)			.58		
14. I go back over things I don't understand (S&G)				.79	
15. If I make a mistake, I try to figure out where I went wrong (L)				.79	
16. I ask myself some questions as I go along to make sure the work makes sense to me (S&G)				.76	
17. I think deeply when I take quizzes in this class (P&V)				.70[a]	

Table 4 (*continued*)

Item	Affective engagement (ω = .90; ICC = .16)	Behavioral engagement–compliance (ω = .82; ICC = .10)	Behavioral engagement–effortful class participation (ω = .82; ICC = .20)	Cognitive engagement (ω = .88; ICC = .15)	Disengagement (ω = .82; ICC = .18)
18. I search for information from different places and think about how to put it together (P&V)				.68	
19. If I'm not sure about things, I check my book or use other materials like charts (S&G)				.64	
20. I try to figure out the hard parts on my own (P&V)				.59	
21. I judge the quality of my ideas or work during class activities (P&V)				.58	
22. I am "zoned out," not really thinking or doing class work (P&V)					.82[a]
23. I let my mind wander (S)					.74
24. I just pretend like I'm working (S)					.77
Factor correlation	Affective	Behavioral-compliance	Behavioral-effortful class participation	Cognitive	
Behavioral-compliance	.68				
Behavioral-effortful class participation	.90	.75			
Cognitive	.66	.77	.80		
Disengagement	.56	.73	.57	.60	

Note. All items begin with the stem "In this class." The eight items for Cognitive Engagement are on a 7-point scale. The other items are on a 5-point scale. Reliability coefficient calculated as McDonald's omega (McDonald, 1999). Intraclass correlation in two-level factor analysis. Within level is student level and between level is classroom level. Sources from which items were adapted were: W&C; B&B = Watson and Clark (1994) and Bergin and Bergin (2009); S = Skinner et al. (2009); P&V = Painter and Valentine (1996); S&G = Stipek and Gralinski (1996); L = Linnenbrink (2005); A = Archambault et al. (2009); M&P; P&M = Maehr and Pintrich (2001) and Pintrich and Maehr (2002).
[a] Items had unstandardized loadings fixed to unity.

Table 5
Model Fit Indexes and Testing Measurement Invariance (Equal Factor Loadings and Thresholds) by Groups in Study 2

Model	No. of parameters	χ^2	df	RMSEA	90% CI		CFI	TLI
Age-level groups (n = 3,548), $\Delta\chi^2$ = 677.8, Δdf = 204, $p < .001$								
Configural invariance	438	3,762.6*	726	.059	.058	.061	.958	.952
Measurement invariance	234	3,988.9*	930	.053	.051	.054	.957	.962
Class subject groups (n = 2,764), $\Delta\chi^2$ = 558.0, Δdf = 306, $p < .001$								
Configural invariance	584	3,090.4*	968	.056	.054	.059	.966	.962
Measurement invariance	278	3,258.9*	1274	.047	.045	.050	.969	.973
Gender groups (n = 2,520), $\Delta\chi^2$ = 191.9, Δdf = 102, $p < .001$								
Configural invariance	292	2,469.3*	484	.057	.055	.059	.963	.957
Measurement invariance	190	2,366.7*	586	.049	.047	.051	.967	.969
Free/reduced lunch status groups (n = 2,520), $\Delta\chi^2$ = 180.0, Δdf = 102, $p < .001$								
Configural invariance	292	2,388.3*	484	.056	.054	.058	.964	.959
Measurement invariance	190	2,267.8*	586	.048	.046	.050	.968	.970

Note. Sample sizes differed when different grouping variables were used due to missing values on grouping variables. The chi-square difference test is based on the DIFFTEST in Mplus for testing the difference between the configural invariance model and the measurement invariance model using the mean- and variance-adjusted weighted least squares estimator. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis index.
* $p < .001$.

24 items by each grouping variable, following recommendations by Kim and Yoon (2011) and Oort (1998), are available as supplemental material online.

Because measurement invariance existed across groups, structured means modeling was conducted to compare groups on their latent factor means. Effect sizes of group differences were calculated as Cohen's d (Cohen, 1988). The confidence intervals of effect sizes were calculated using the asymptotic estimates of standard errors as described in Nakagawa and Cuthill (2007). Effect sizes and their confidence intervals are in Table 6. Students in middle and high schools had lower classroom engagement and higher disengagement than students in elementary schools. Free/reduced lunch recipients, compared to nonrecipients, had lower Cognitive Engagement (Cohen's $d = -0.09$), and lower Behavioral Engagement, including both Compliance (Cohen's $d = -0.22$) and Effortful Class Participation (Cohen's $d = -0.12$). Male students, compared to female students, had lower Affective Engagement (Cohen's $d = 0.13$) and lower Behavioral Engagement, including both Compliance (Cohen's $d = 0.32$) and Effortful Class Participation (Cohen's $d = 0.17$), and higher Disengagement (Cohen's $d = -0.22$). In terms of class subjects, using math classes for comparison, students in language arts classes had lower Compliance (Cohen's $d = -0.15$); students in science classes had higher Affective Engagement (Cohen's $d = 0.20$) and Effortful Class Participation (Cohen's $d = 0.18$) and lower Compliance (Cohen's $d = -0.16$); there were no statistically significant differences between students in social studies classes and students in math classes in classroom engagement.

Step 4: Test validity of the CEI. Validity of the CEI was examined by correlating factor scores from the five-factor CFA model with variables that research and theory predict would be linked to classroom engagement. Items measuring engagement-fostering teacher behaviors were taken from the Instructional Practices Inventory (Painter & Valentine, 1996) and adapted for student report. The items measure teacher behaviors that elicit student engagement, such as asking for deeper thinking (Marks, 2000). Four items assessed whole-class activities (e.g., "In this class, we have class discussions that make me think deeply") and three items assessed teachers' questioning prac-

Table 6
Effect Sizes (Cohen's d) of Latent Mean Differences Between Groups in Study 2

Groups compared	Affective engagement	Behavioral engagement–compliance	Behavioral engagement–effortful class participation	Cognitive engagement	Disengagement[a]
Middle school/elementary school	−0.64*** [−0.72, −0.55]	−0.45*** [−0.53, −0.37]	−0.68*** [−0.76, −0.60]	−0.60*** [−0.69, −0.52]	0.70*** [0.62, 0.79]
High school/elementary school	−0.66*** [−0.75, −0.58]	−0.69*** [−0.78, −0.61]	−0.92*** [−1.01,−0.83]	−0.82*** [−0.91, −0.73]	0.94*** [0.85, 1.03]
Free/reduced lunch/no free/reduced lunch	−0.04 [−0.12, 0.04]	−0.22*** [−0.29, −0.14]	−0.12** [−0.20, −0.04]	−0.09* [−0.16, −0.01]	0.06 [−0.02, 0.14]
Female/male	0.13** [0.06, 0.21]	0.32*** [0.24, 0.40]	0.17*** [0.09, 0.25]	−0.01 [−0.09, 0.07]	−0.22*** [−0.30, −0.15]
Language arts/math	0.09 [−0.01, 0.19]	−0.15* [−0.25, −0.05]	0.09 [0.00, 0.19]	−0.05 [−0.15, 0.04]	0.04 [−0.06, 0.13]
Science/math	0.20** [0.09, 0.31]	−0.16* [−0.28, −0.05]	0.18** [0.06, 0.29]	−0.01 [−0.12, 0.11]	0.06 [−0.05, 0.17]
Social studies/math	0.10 [−0.01, 0.22]	−0.09 [−0.21, 0.02]	0.02 [−0.09, 0.13]	−0.05 [−0.16, 0.07]	0.08 [−0.03, 0.19]

Note. Confidence intervals (95%) of each Cohen's *d* are given in brackets. A negative value means that the group before the minus (–) sign in the first column of the same row had a lower latent factor mean than the group after the minus sign.
[a] The sign of latent mean differences in Disengagement was reversed so that a higher value meant more disengagement.
* $p < .05$. ** $p < .01$. *** $p < .001$.

tice (e.g., "My teacher asks me questions that have more than one right answer"). Items measuring academic self-efficacy, or belief in one's ability to accomplish schoolwork (e.g., "In this class, I'm very sure I can master the skills taught"), were taken from the Patterns of Adaptive Learning Survey (PALS, Midgley et al., 2000). Academic self-efficacy predicts interest and achievement and should be related to engagement (Bandura, 1997). Items measuring student achievement goals also were taken from the PALS. Three items measuring each of three types of goals—mastery, performance-avoidance, and performance-approach—were modified to be classroom level (e.g., for mastery, "In this class, one of my goals is to master a lot of new ideas and skills"; for performance-avoidance, "In this class, it's important to me that I not do worse than others"; for performance-approach, "In this class one of my goals is to show others that I'm good at my work"). Previous research has suggested that mastery goals should be most strongly correlated with engagement, followed by performance-approach goals, and performance-avoidance goals (Linnenbrink, 2005; Wolters, 2004). School-prompted interest measures students' propensity to learn more about school topics, but not as school assignments, and was measured with 10 items (e.g., "Because of this class, I get so interested in this subject that I try to find out more about it on my own by looking the subject up on the Internet") patterned after Bergin and colleagues' work (e.g., D. A. Bergin, 1992; Ciani, Ferguson, Bergin, & Hilpert, 2010). School-prompted interest should be positively related to engagement because the more students are engaged in their classroom activities; the more likely they would be to want to learn more about the topics outside of school. Finally, self-reported grades in the targeted class were obtained; while self-reported grades should be used with caution, they "generally predict outcomes to a similar extent as actual grades" (Kuncel, Credé, & Thomas, 2005, p. 76).

Factor scores for Disengagement were reversed around its grand mean such that a higher value indicated more disengagement and less engagement. Table 7 shows the relationship between each of the validity constructs and the five engagement factors. Separate correlations by age level were computed. Results were similar for elementary, middle, and high school

Table 7
Correlations Between Engagement Factors and Validity Variables in Study 2

Validity variable	Affective engagement	Behavioral engagement–compliance	Behavioral engagement–effortful class participation	Cognitive engagement	Disengagement
Teacher behavior—whole class activities (.77)	.50	.49	.57	.57	−.34
Teacher behavior—questioning practices (.62)	.37	.35	.44	.41	−.17
Self-efficacy (.86)	.55	.63	.61	.67	−.49
Mastery goals (.82)	.60	.70	.68	.77	−.59
Performance-approach goals (.76)	.39	.40	.44	.49	−.31
Performance-avoid goals (.56)	.25	.31	.31	.39	−.22
School-prompted interest (.86)	.57	.49	.61	.55	−.39
Self-reported grades	.34	.39	.35	.30	−.37

Note. Reliabilities (Cronbach's alpha coefficients) are in parentheses. All correlations are significant ($p < .001$).

students, so only the aggregated total across age level is reported in Table 7. All correlations were in the expected direction, supporting validity of the CEI. Correlations between the Disengagement factor and the validity variables were negative and those between the other CEI factors and the validity variables were positive. Regarding the magnitude of the correlations, academic self-efficacy (*r*s = .49–.67) and engagement-fostering teacher behaviors predicted CEI scores, with student reports of whole class activities more strongly predicting engagement (*r*s = .34–.57) than student reports of teacher questioning behavior (*r*s = .17–.44). Students' mastery goals correlated more strongly with engagement (*r*s = .59–.77) than performance-approach goals (*r*s = .31–.49), which, in turn, correlated more strongly than performance-avoidance goals (*r*s = .22–.39). Furthermore, engagement predicted school-prompted interest (*r*s = .39–.61) and self-reported grades (*r*s = .30–.39).

Discussion

In this article we discussed the development of the CEI and investigated its factor structure and validity in two studies. Methods appropriate for ordered categorical indicators were used. There has been an increasing emphasis on considering the ordered categorical nature of response options in factor analysis and structural equation modeling (e.g., Flora & Curran, 2004; Wang & Su, 2013). This article is an application of these advanced methods.

A significant contribution of these two studies is a newly available and much-needed self-report tool for measuring classroom engagement across a wide age range. Observational measures can be useful, but are too time consuming and expensive for schools to use regularly. Four key findings promote understanding of the construct of classroom engagement. First, results of both Study 1 and Study 2 indicate that engagement has three distinct dimensions—affective, behavioral, and cognitive—as well as a distinct disengagement dimension, as measured by the CEI. However, in Study 2, the behavioral dimension consisted of two factors. We called one, "Compliance" because it reflects obedience to classroom norms. We called the second, "Effortful Class Participation." This distinction is important. Fredricks et al. (2004) pointed out that "some studies separate cooperative participation, or adhering to classroom rules, from autonomy participation, or self-directed academic behaviors" (p. 62). They argued that researchers should "differentiate students who become invested in learning from those who do what they are supposed to do . . . but do not become invested in learning" (p. 85). Study 2 contributes to the literature by empirically confirming that, in this large, multiage sample, compliance is distinct from other dimensions of engagement.

These results expand previous research on the multidimensionality of engagement by focusing on multiple dimensions at the classroom level (Glanville & Wildhagen, 2007; Reschly, Huebner, Appleton, & Antaramian, 2008; Skinner et

al., 2009). One can imagine a child who tries hard and pays attention (i.e., is behaviorally engaged) due to external pressure, but is not affectively engaged (experiencing enjoyment or interest) in the learning activities. Indeed, Skinner et al. (2009) found that affective and behavioral dimensions of engagement were distinct but correlated (r = .53 fall; .60 spring) using self-report with items similar to the CEI. They did not measure cognitive engagement. Reeve and Tseng (2011) also found that affective and behavioral engagement were distinct but correlated (r = .30) in a sample of Taiwanese high school students. They did not use a classroom-level measure, but focused on the school-level instead. By comparison, in Study 1 Affective Engagement correlated .73 with Behavioral Engagement and in Study 2 Affective Engagement correlated .90 with Effortful Class Participation and .68 with Compliance. The larger correlations found in the present studies may be partly because items in the CEI were treated as ordinal, and polychoric correlations were used for factor analysis. When observed variables do not have similar distribution shapes, traditional factor analysis based on Pearson correlations would likely underestimate their relationship (Goodwin & Leech, 2006).

Fredricks et al. (2004) expressed concern that cognitive and behavioral engagement may not be distinct factors. However, the present results indicate that the CEI distinctly measures these two dimensions of engagement. That is, a student could be behaviorally engaged without being cognitively engaged. For example, a student may be "paying attention" during an algebra lesson on an algorithm for solving a particular type of problem, but not deeply processing what it means. Alternatively, a student might be deeply processing math in class, but not paying attention to the lesson.

The CEI provides a tool to raise psychologists, teachers, and administrators' awareness of the varied types of engagement, and a tool for measuring all types efficiently. However, we do not want to overemphasize the dimensionality of engagement. We also found that the engagement construct was possibly hierarchical in Study 1, and the first-order factors in Study 2 correlated moderately to moderately high. Thus, one might consider classroom-level engagement a single global construct. These results confirm those of Archambault et al. (2009) who used a CFA in a study of high school students and found that affective and behavioral components were separate, but that a case for a global construct of engagement could also be made. Note that their items were at the classroom and school levels.

The second key finding was that factor structure was invariant across groups. Other researchers have stated that "one question that has yet to be fully evaluated is the extent to which the structure of student engagement differs across maturational levels and with respect to gender" (Betts et al., 2010, p. 85). The present studies directly address this issue. In both studies we found invariant factor structure by age level, gender, free/reduced lunch status, and class subject. Other researchers have found that engagement dimensionality was invariant when measured at the school level, rather than classroom level. Glanville and Wildhagen (2007), using the National Educational Longitudinal Study of 1988 (NELS88) data, found invariance across ethnicity. Betts et al. (2010) found invariant factor structure on the SEI from the sixth to 12th grade. Their results are not directly comparable because the SEI measures a variety of motivational constructs that go beyond the definition of engagement presented in this article, including social support, extrinsic motivation, autonomy support, and utility. The finding that the CEI factor structure is invariant across demographic groups and core class subjects is important because it suggests that the CEI could be used with a variety of students and contexts. It can be used for longitudinal research, as well as evaluation of interventions, from late elementary school through high school.

The third key finding pertained to group differences in engagement. Marks (2000) found that high socioeconomic status (SES) students were more engaged than low SES students, although the effect was small. In contrast, Reschly et al. (2008) found no significant correlations between the free/reduced lunch status and any of their affect, engagement, or coping scales. The present results fit more closely with Marks's sample; we found that nonrecipients of free/reduced lunch had higher cognitive and behavioral engagement. In the present study we found that girls had greater affective and behavioral engagement than boys, and lower disengagement. Marks (2000) also found that girls were moderately more engaged than boys. Like-

wise, Skinner et al. (2009) reported that girls tended to be motivationally "better off" than boys, showing consistently higher levels of behavioral and affective engagement, as well as lower levels of behavioral and emotional disaffection.

The fourth, and final, key finding is that the CEI has internal consistency and evidence of validity. Internal consistency, calculated as McDonald's omega (McDonald, 1999), of each of the five engagement factors ranged from .82 to .90. Evidence for the CEI's validity is provided by the pattern of results, which is consistent with theory and published research. Teacher behaviors that elicit engagement—both whole class activities and questioning practices—predicted the CEI factors, as suggested by previous research (Painter & Valentine, 1996). Two student motivational variables—self-efficacy and achievement goal orientation—predicted CEI factors. Lau and Roeser (2002) observed a correlation of .47 between self-efficacy and what they called classroom engagement (doing homework, use of self-regulated learning strategies, paying attention, doing class activities). We found that all dimensions of the CEI were correlated with academic self-efficacy (rs = .49–.67), with Cognitive Engagement showing the strongest correlation. We also found that mastery goal orientation more strongly correlated with engagement than performance-approach goals, which is congruent with the literature on achievement goal orientation (DeBacker & Crowson, 2006; Linnenbrink, 2005). For example, mastery goals have been demonstrated to correlate positively with indicators of engagement such as deep processing strategies, metacognitive strategies, self-regulation, preference for challenging tasks, effort, interest, and persistence (e.g., Harackiewicz, Durik, Barron, Linnenbrink-Garcia, & Tauer, 2008).

Furthermore, we found that CEI scores predicted outcomes that should be linked to engagement—school-prompted interest and self-reported grades. School-prompted interest is a stringent test of engagement. It is a low-frequency behavior in which students pursue out-of-school learning because of interest sparked in school and is a potential outcome of strong engagement (D. A. Bergin, 1992; Ciani et al., 2010). Effortful Class Participation was correlated more strongly with school-prompted interest (r = .61) than Compliance (r = .49). Both school-prompted interest and Effortful Class Participation reflect active student learning. In contrast, the other behavioral engagement factor—Compliance—may reflect obedience in the classroom, rather than engagement with the content, and mere obedience to classroom rules should be somewhat less likely to lead to effortful out-of-school learning.

Engagement is considered an important variable because it is supposed to lead to greater learning. Indeed, the CEI factors were correlated with self-reported grades (rs = .30–.39). The dimension most strongly correlated with classroom grades was Compliance. This result is logical given that grades are often based on homework completion and obedience to the teachers' agenda. Likewise, Voelkl (1997) observed correlations of .42 and .49 between standardized test scores and behavioral engagement (paying attention, doing more than just the assigned work) among fourth and seventh graders in Chicago, IL. Fredricks and colleagues (2011) reported that "ten of the student self-report measures of engagement reported positive correlations with some measure of student achievement" (p. 20), even though several of these measures addressed school-level bonding.

The study has several limitations. First, the data are cross-sectional and do not demonstrate longitudinal change across time. Second, Studies 1 and 2 were carried out in the same school district with predominantly White, rural, lower income students; it would be useful to administer the CEI in other settings, including urban, multiethnic, and higher income schools. Third, the correlations among variables could be inflated because all variables are measured by self-report simultaneously. In addition, the CEI measures frequency, but not intensity or degree of engagement in the class. We chose frequency because we felt this was the most useful to educators, but for some purposes a measure of intensity may be more appropriate.

The studies in this article provide a foundation for future research on the construct of classroom engagement. In both studies, we asked students to report their engagement for the specific class that they were currently attending. Variation attributable to class was therefore confounded with variation attributable to students. One way to address this issue would be to ask students to report their engagement in mul-

tiple classes and apply generalizability theory studies (Brennan, 2001; Shavelson & Webb, 1991) to separate variation due to students, due to classes, and due to their interaction. Using data for the final CEI model in Study 2, we calculated the intraclass correlations (ICCs) of the five factors using a two-level analysis in Mplus. The within level was the student level and the between level was classrooms. The ICCs ranged from .10 to .20 (see Table 4), a range that is slightly lower than but in close proximity to that for mathematics achievement based on nationally representative data with schools being the between level units (Hedges & Hedberg, 2007, Table 2 on p. 68). Future research also could use both the CEI and the SEI in the same population to investigate the relationship between classroom-level and school-level engagement. Furthermore, researchers could test the links between the scores on the CEI and teacher practices, bullying, or course subject. Finally, future research should focus on acquiring a large, normative database to aid in interpreting the results of specific uses of the CEI.

In summary, the CEI is a relatively brief, self-report measure of classroom-level engagement that has demonstrated good psychometric properties and validity. It is a practical tool for measuring all dimensions of engagement—cognitive, affective, and behavioral—in classrooms from the fourth to 12th grade. Its structure is invariant for elementary, middle, and high school students, for boys and girls, for low-income and higher income students, and for various class subjects, which suggests it can be used in a variety of contexts. A next step will be to apply this measure in other samples and to collect additional validity evidence with relevant external measures.

References

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology, 44,* 427–445. doi:10.1016/j.jsp.2006.04.002

Archambault, I., Janosz, M., Fallu, J.-S., & Pagani, L. S. (2009). Student engagement and its relationship with early high school dropout. *Journal of Adolescence, 32,* 651–670. doi:10.1016/j.adolescence.2008.06.007

Asparouhov, T., & Muthén, B. (2006). *Robust chi square difference testing with mean and variance adjusted test statistics* (Mplus Web Notes No. 10). Retrieved from http://statmodel2.com/download/webnotes/webnote10.pdf

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

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13,* 186–203. doi:10.1207/s15328007sem1302_2

Bergin, C., & Bergin, D. A. (2009, May). *Positive emotions in high school classrooms predict positive outcomes*. Paper presented as a poster at the annual convention of the Association for Psychological Science, San Francisco, CA.

Bergin, D. A. (1992). Leisure activity, motivation, and academic achievement in high school students. *Journal of Leisure Research, 24,* 225–239.

Betts, J., Appleton, J. J., Reschly, A. L., Christenson, S. L., & Huebner, E. S. (2010). A study of the factorial invariance of the Student Engagement Instrument (SEI): Results from middle and high school students. *School Psychology Quarterly, 25,* 84–93. doi:10.1037/a0020259

Brennan, R. L. (2001). *Generalizability theory*. New York, NY: Springer. doi:10.1007/978-1-4757-3456-0

Carter, C. P., Reschly, A. L., Lovelace, M. D., Appleton, J. J., & Thompson, D. (2012). Measuring student engagement among elementary students: Pilot of the Student Engagement Instrument—Elementary version. *School Psychology Quarterly, 27,* 61–73. doi:10.1037/a0029229

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14,* 464–504. doi:10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255. doi:10.1207/S15328007SEM0902_5

Ciani, K. D., Ferguson, Y., Bergin, D. A., & Hilpert, J. (2010). Motivational influences on school-prompted interest. *Educational Psychology, 30,* 377–393. doi:10.1080/01443411003660232

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Darr, C. W. (2012). Measuring student engagement: The development of a scale for formative use. In S. L. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 707–723). New York, NY: Springer. doi:10.1007/978-1-4614-2018-7_34

DeBacker, T. K., & Crowson, H. M. (2006). Influences on cognitive engagement: Epistemological

beliefs and need for closure. *British Journal of Educational Psychology, 76,* 535–551. doi:10.1348/000709905X53138

Finn, J. D., & Zimmer, K. S. (2012). Student engagement: What is it? Why does it matter? In S. L. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 97–131). New York, NY: Springer. doi:10.1007/978-1-4614-2018-7_5

Finney, S., & DiStefano, J. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–313). Greenwich, CT: Information Age Publishing.

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. doi:10.1037/1082-989X.9.4.466

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. doi:10.3102/00346543074001059

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763–782). New York, NY: Springer. doi:10.1007/978-1-4614-2018-7_37

Fredricks, J. A., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). *Measuring student engagement in upper elementary through high school: A description of 21 instruments* (Issues & Answers Rep. REL 2011–No. 098). Washington, DC: U.S. Department of Education, IES.

Glanville, J. L., & Wildhagen, T. (2007). The measurement of school engagement: Assessing dimensionality and measurement invariance across race and ethnicity. *Educational and Psychological Measurement, 67,* 1019–1041. doi:10.1177/0013164406299126

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of *r*. *Journal of Experimental Education, 74,* 251–266. doi:10.3200/JEXE.74.3.249-266

Harackiewicz, J. M., Durik, A. M., Barron, K. E., Linnenbrink-Garcia, L., & Tauer, J. M. (2008). The role of achievement goals in the development of interest: Reciprocal relations between achievement goals, interest, and performance. *Journal of Educational Psychology, 100,* 105–122. doi:10.1037/0022-0663.100.1.105

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29,* 60–87. doi:10.3102/0162373707299706

Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling, 18,* 212–228. doi:10.1080/10705511.2011.557337

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75,* 63–82. doi:10.3102/00346543075001063

Lau, S., & Roeser, R. (2002). Cognitive abilities and motivational processes in high school students' situational engagement and achievement in science. *Educational Assessment, 8,* 139–162. doi:10.1207/S15326977EA0802_04

Lawson, M. A., & Lawson, H. A. (2013). New conceptual frameworks for student engagement research, policy, and practice. *Review of Educational Research, 83,* 432–479. doi:10.3102/0034654313480891

Linnenbrink, E. A. (2005). The dilemma of performance-approach goals: The use of multiple goal contexts to promote students' motivation and learning. *Journal of Educational Psychology, 97,* 197–213. doi:10.1037/0022-0663.97.2.197

Maehr, M. L., & Pintrich, P. R. (2001). *Motivation in collaborative learning groups*. Ann Arbor: University of Michigan, Combined Program in Education and Psychology.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37,* 153–184. doi:10.3102/00028312037001153

Martin, A. J. (2009). Motivation and engagement across the academic life span: A developmental construct validity study of elementary school, high school, and university/college students. *Educational and Psychological Measurement, 69,* 794–824. doi:10.1177/0013164409332214

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80,* 514–523. doi:10.1037/0022-0663.80.4.514

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., . . . Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales*. Ann Arbor: University of Michigan.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and

continuous latent variable indicators. *Psychometrika, 49,* 115–132. doi:10.1007/BF02294210

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews, 82,* 591–605. doi:10.1111/j.1469-185X.2007.00027.x

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44,* 443–460. doi:10.1007/BF02296207

Oort, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 5,* 107–124. doi:10.1080/10705519809540095

Painter, B., & Valentine, J. (1996). *Instructional practices inventory*. Columbia: University of Missouri, Center for School Improvement.

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 101,* 115–135. doi:10.1037/a0013383

Pintrich, P. R., & Maehr, M. L. (2002). *Motivation for learning: The variable role of concepts of self and sociocultural identities*. Ann Arbor: University of Michigan, Combined Program in Education and Psychology.

Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion, 28,* 147–169. doi:10.1023/B:MOEM.0000032312.95499.6f

Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology, 36,* 257–267. doi:10.1016/j.cedpsych.2011.05.002

Reschly, A. L., Huebner, E. S., Appleton, J. J., & Antaramian, S. (2008). Engagement as flourishing: The contribution of positive emotions and coping to adolescents' engagement at school and with learning. *Psychology in the Schools, 45,* 419–431. doi:10.1002/pits.20306

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69,* 493–525. doi:10.1177/0013164408323233

Stipek, D., & Gralinski, J. H. (1996). Children's beliefs about intelligence and school performance. *Journal of Educational Psychology, 88,* 397–407. doi:10.1037/0022-0663.88.3.397

U.S. Census Bureau. (2012). *Annual estimates of the population of metropolitan and micropolitan statistical areas: April 1, 2010 to July 1, 2012*. Retrieved from http://www.census.gov/popest/data/metro/totals/2012/index.html

van Prooijen, J.-W., & van der Kloot, W. A. (2001). Confirmatory analysis of exploratively obtained factor structures. *Educational and Psychological Measurement, 61,* 777–792. doi:10.1177/00131640121971518

Voelkl, K. E. (1997). Identification with school. *American Journal of Education, 105,* 294–318. doi:10.1086/444158

Wang, Z., Rohrer, D., Chuang, C.-C., Fujiki, M., Herman, K., & Reinke, W. (2014). Five methods to score the teacher observation of classroom adaptation checklist and to examine group differences. *The Journal of Experimental Education*. Advance online publication. doi:10.1080/00220973.2013.876230

Wang, Z., & Su, I. (2013). Longitudinal factor structure of general self-concept and locus of control among high school students. *Journal of Psychoeducational Assessment, 31,* 554–565. doi:10.1177/0734282913481651

Watson, D., & Clark, L. A. (1994). *Manual for the Positive and Negative Affect Schedule—Expanded form*. Iowa City: University of Iowa.

Wellborn, L. G. (1992). *Engaged and disaffected action: The conceptualization and measurement of motivation in the academic domain* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Order No. 9218563)

Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology, 96,* 236–250. doi:10.1037/0022-0663.96.2.236

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34,* 806–838. doi:10.1177/0011000006288127

Received June 20, 2013
Revision received November 16, 2013
Accepted November 20, 2013 ■