Journal of School Psychology 49 (2011) 465–480

Contents lists available at ScienceDirect

Journal of School Psychology

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

The assessment of school engagement: Examining dimensionality and measurement invariance by gender and race/ethnicity

Ming-Te Wang [b,c,*], John B. Willett [a], Jacquelynne S. Eccles [b]

[a] Harvard University, United States
[b] University of Michigan, United States
[c] University of Pittsburgh, United states

ARTICLE INFO

Article history:
Received 10 March 2010
Received in revised form 1 April 2011
Accepted 2 April 2011

Keywords:
School engagement
Multidimensionality
Measurement invariance
Confirmatory factor analysis

ABSTRACT

The construct of school engagement has attracted growing interest as a way to ameliorate the decline in academic achievement and increase in dropout rates. The current study tested the fit of a second-order multidimensional factor model of school engagement, using large-scale representative data on 1103 students in middle school. In order to make valid model comparisons by group, we evaluated the extent to which the measurement structure of this model was invariant by gender and by race/ethnicity (European-American vs. African-American students). Finally, we examined differences in latent factor means by these same groups. From our confirmatory factor analyses, we concluded that school engagement was a multidimensional construct, with evidence to support the hypothesized second-order engagement factor structure with behavioral, emotional, and cognitive dimensions. In this sample, boys and girls did not substantially differ, nor did European-American and African-American students, in terms of the underlying constructs of engagement and the composition of these constructs. Finally, there were substantial differences in behavioral and emotional engagement by gender and by racial/ethnic groups in terms of second-order factor mean differences.

* Corresponding author at: ISR 5110, 426 Thompson St. Ann Arbor 48106, United States. Tel.: +1 734 647 0862.
E-mail address: wangmi@umich.edu (M.-T. Wang).
ACTION EDITOR: Sara Bolt.

0022-4405/$ – see front matter
doi:10.1016/j.jsp.2011.04.001

1. Introduction

Active engagement in school is important for student learning and academic success. Previous studies have found that students who are more engaged in school have better academic performance (Csikszentmihalyi & Schneider, 2000; Newmann, Wehlage, & Lamborn, 1992). Students who attend school regularly, concentrate on learning, and adhere to the rules of the school, generally receive higher grades and perform better on standardized tests (Bandura, Barbaranelli, Caprar, & Pastorelli, 1996; Caraway, Tucker, Reinke, & Hall, 2003; Wang & Holcombe, 2010). In contrast, lack of school engagement can have serious consequences for students, including underachievement, engagement in deviant behaviors, and increased risk of dropping out of school (Finn & Rock, 1997). Interest in school engagement has intensified recently due to the increasing proportion of adolescents who report feeling alienated, particularly as they progress from elementary to middle and then from middle to high school (e.g., Finn, 1989; Finn & Voelkl, 1993; Marks, 2000; McDermott, Mordell, & Stolzfus, 2001). The National Research Council and Institute of Medicine (2004) reported that 30% to 50% of middle school students are disengaged from school. These findings lead researchers and educators to underscore the need to enhance students' school engagement.

1.1. Theoretical framework for school engagement

According to Fredricks, Blumenfeld, and Paris (2004), school engagement is a multidimensional construct that is made up of three components: *behavior*, *emotion*, and *cognition*. Behavioral engagement refers to the actions and practices that students direct toward school and learning; it includes positive conduct (e.g., attending class and completing schoolwork), involvement in learning and academic tasks (e.g., effort and concentration), and participation in extracurricular activities (e.g., athletics). Emotional engagement represents a student's affective reactions and sense of connectedness to school (Finn, 1989; Skinner & Belmont, 1993). Finn (1989) conceptualizes it as students' sense of belonging to and valuing of their school. Cognitive engagement refers to a student's self-regulated and strategic approach to learning in which students use metacognitive strategies to plan, monitor, and evaluate their cognition (Connell & Wellborn, 1991; Zimmerman, 1989).

These engagement components do not operate in isolation, but rather, they are understood as interwoven and dynamic. Fredricks et al. (2004) proposed that patterns of engagement across these three dimensions have long-term effects on students' academic achievement. Emotional identification with school interacts with behavioral engagement and cognitive involvement in school learning. This interaction is of concern as low levels of each dimension have been shown to lead to unsuccessful school outcomes. In addition, a lack of behavioral participation and cognitive engagement is also related to emotional withdrawal from school-related activities, resulting in even less academic success. Over time, behavioral participation, emotional identification, and cognitive engagement exert reciprocal influence. Ultimately, the degree to which students engage in school behaviorally, emotionally, and cognitively influences their academic success, which in turn, may influence changes in all three aspects of school engagement.

Although empirical researchers have developed scales for measuring engagement, until quite recently many have measured school engagement as either a uni-dimensional combination of these various indicators (e.g., Daly, Shin, Thakral, Selders, & Vera, 2009; Perry, Liu, & Pabian, 2010; Simons-Morton & Chen, 2009; You & Sharkey, 2009) or, at most, as two of the three types (e.g., Appleton, Christenson, Kim, & Reschly, 2006; Connell & Wellborn, 1991; Skinner, Furrer, Marchand, & Kindermann, 2008). The various measurement models for school engagement are reflective of differences that exist in how researchers define the construct. For instance, some researchers have developed scales for measuring engagement as a global construct, such as the Research Assessment Package for Schools (Institute for Research and Reform in Education, 1998), the High School Survey of Student Engagement (Yazzie-Mintz, 2007), the Rochester Assessment Package for Schools (Wellborn & Connell, 1987), and the engagement scale of the School Success Profile (Bowen & Richman, 1995). While these global instruments of engagement allow the researcher to reflect a holistic picture of engagement, a major weakness of this type of measurement model is that it may not provide sufficient numbers of items to understand individual scales or new understandings about the multidimensionality of the construct. If school engagement is indeed a multidimensional construct, this practice of measuring it as a single construct limits examining differences among the various types of engagement and understanding their possible antecedents and consequences (Fredricks et al., 2004; Guthrie & Wigfield, 2000; Wang & Holcombe, 2010).

Some researchers adopting a two-component model frequently include a behavioral and an emotional or affective component in their engagement instruments (Finn, 1989; Marks, 2000; Skinner et al., 2008). For instance, Marks (2000) measured engagement with a student-report survey designed to assess students' effort, attentiveness, lack of boredom, and the extent to which they completed assignments. Although there is evidence to suggest the importance of cognitive engagement to school performance, few instruments include separate cognitive components (see Appleton et al., 2006; Jimerson, Campos, & Greif, 2003 for a review). To address this gap, Appleton et al. (2006) designed the Student Engagement Instrument to measure two subtypes of student engagement with school: psychological and cognitive engagement. Martin (2007) developed the Motivation and Engagement Scale to assess behavioral and cognitive dimensions of school engagement. He proposed a 4 second-order factor model, including adaptive and maladaptive behavioral dimensions and adaptive and maladaptive cognitive dimensions. Most recent studies of engagement have proposed a tripartite conceptualization that includes behavioral, emotional, and cognitive components (Fredricks et al., 2004; Jimerson et al., 2003). However, few existing instruments are adequate for addressing this model. In summary, within the varied approaches, studying school engagement globally and studying it in terms of its component parts, there is a lack of consistency in the specific measurement tools that researchers use.

We have sought to develop an instrument that captures all three of aspects of school engagement and to study them simultaneously given that they are likely to be interrelated. Thus our instrument development was guided explicitly by the theoretical framework proposed by Fredricks et al. (2004). Like other recent papers in this field, we have aimed to create a theoretically-grounded instrument that captures multiple aspects of school engagement (see Appleton et al., 2006; Martin, 2007).

1.2. Measurement invariance

Another potential weakness in common practices of measuring engagement is the lack of accounting adequately for the error present in its measurement (e.g., Bowen & Richman, 1995; Connell & Wellborn, 1991; Jimerson et al., 2003; Skinner et al., 2008). Across studies, researchers often apply different measurement strategies and design different indicators to measure engagement. However, because each of these items or indicators represents that student's level of engagement imperfectly, analyses that use such an index are vulnerable to the negative consequences of measurement error (Bollen, 1989). Thus, we use confirmatory factor analysis (CFA) to take measurement error into account by estimating the measurement error variances associated with each indicator directly.

Prior research indicates that students' level of school engagement and academic performance may differ profoundly by gender and race/ethnicity (Brooks-Gunn & Duncan, 1997). Generally, throughout schooling, girls report higher levels of school engagement than boys regardless of what types of engagement are considered (see Johnson, Crosnoe, & Elder, 2001; Martin, 2004, 2007; Zimmerman & Martinez-Pons, 1990). Research on engagement has produced mixed evidence on racial–ethnic differences, despite most studies indicating that African American students are less likely to engage in their schools than European American students (Ainsworth-Darnell & Downey, 1998). For instance, Johnson et al. (2001) found that African-American students reported lower levels of school attachment but were more likely to pay attention and complete homework, whereas Voelkl (1997) found that African American students had higher levels of school identification than European American students. However, most of these studies have not considered the implications of *measurement invariance* when making group comparisons (Martin, 2007). *Measurement invariance* generally refers to the extent to which the content of each item is being perceived and interpreted in the same way across samples (Byrne & Watkins, 2003). In fact, if measures of engagement operate differently across gender and ethnic groups and these variations are not taken into account in the measurement, it is inappropriate to compare levels of engagement or its effects across groups (Glanville & Wildhagen, 2007).

It is important to investigate whether the measurement of engagement functions similarly across groups of students. By doing so, levels of engagement can be compared appropriately from group to group or common scores of engagement can be used to predict school achievement for students with different demographic characteristics (e.g., Byrne & Watkins, 2003; Reise, Widaman, & Pugh, 1993; Wicherts, Dolan,

& Hessen, 2005). To confirm measurement invariance, we must test the equivalence of a measured construct across two or more independent groups to assure that the construct is being assessed similarly in each group (Chen, Sousa, & West, 2005). Meaningful comparisons of statistics – such as means and regression coefficients – can only be made then. Without assuring measurement invariance, group comparisons described by researchers are likely to be substantively misleading and potentially artifactual (e.g., Byrne, 1998; Byrne & Stewart, 2006; Thompson & Green, 2006; Van de Vijver & Leung, 1997).

Testing for measurement invariance includes a series of hierarchical steps, beginning with the establishment of a baseline model in each group, followed by tests for equivalence across groups at several increasingly more restricted levels. These steps involve configural, factor loading, and intercept levels based on techniques developed by Meredith (1993) and Widaman and Resie (1997). The first step (configural invariance) is to test whether each group has the same number of dimensions and patterns of fixed and free parameters (Bollen, 1989). If the fit of this baseline model is acceptable, higher levels of invariance could be examined. The second step is to assess whether the factor loadings for the latent variables are invariant across groups. If this condition holds, one can test whether the intercepts are invariant (Widaman & Resie, 1997).

1.3. Conclusion: specific research questions

The main goal, in this study, was threefold. First, we propose and fit a multidimensional factor model of school engagement. Second, we examine the extent to which this model demonstrates measurement invariance by gender and race/ethnicity. Third, we compare latent variable mean differences across groups if an adequate level of measurement invariance is present. In doing so, we address the following research questions:

1. Does the construct of school engagement display a second-order multidimensional factor structure in which each of three hypothesized second-order factors subsumes two specified first-order factors?
2. Does the multidimensional factor model of school engagement demonstrate measurement invariance separately by gender and race/ethnicity?
3. Are there gender and race/ethnic group differences in school engagement and is the pattern of difference the same across all types of school engagement?

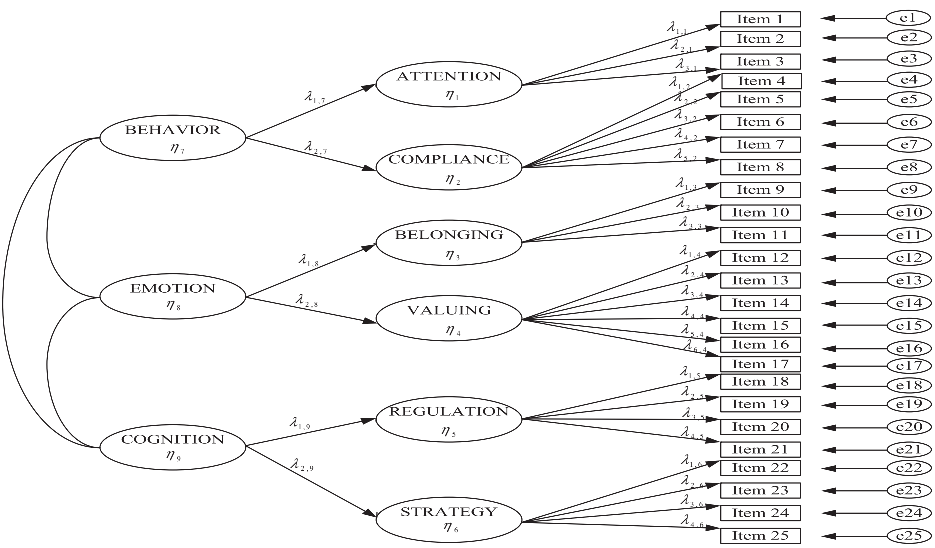

Fig. 1. Factor model of school engagement depicting the second-order factor structure hypothesized to underlie the six first-order factors.

We present our hypothesized measurement models in Fig. 1. We hypothesize that the indicators that we used to measure school engagement will display a first-order factor structure that contains six factors, with each of the six factors composed of multiple items, as indicated. These first-order factors are as follows: (a) attentiveness, (b) school compliance, (c) school belonging, (d) valuing of school education, (e) self-regulated learning, and (f) cognitive strategy use. We further hypothesize that these six first-order factors will display a tri-dimensional second-order factor structure with each of the three second-order dimensions subsuming two first-order factors: (a) the behavioral dimension includes school attentiveness and compliance, (b) the emotional dimension includes school belonging and valuing of school education, and (c) cognitive dimension includes self-regulated learning and cognitive strategy use. In addition, we hypothesize that girls will display greater behavioral, emotional, and cognitive engagement than boys. Given the mix of findings on race/ethnic group differences, we make no specific hypotheses regarding these differences.

2. Method

2.1. Dataset

The data are drawn from the Maryland Adolescent Development in Context Study (MADICS), an ongoing longitudinal study of more than 1000 adolescents, their families, and their teachers. The MADICS was conducted for two main purposes: to examine how social context influences psychological determinants of behavioral choices and to examine various developmental trajectories during adolescence and into adulthood. This dataset contains rich descriptors of the home, school, and peer group that may relate to adolescents' social and psychological development. Many of the questionnaire items were derived from other large-scale longitudinal studies and from previously validated scales.

2.2. Sample

Participants in the sample were part of the MADICS from 23 public schools in a large, ethnically diverse county on the East Coast of the United States. In the current study, we used data from the third wave, collected in the fall when the adolescents were at ninth grade. The sample contained 1103 students. Approximately 56% of participating students were African American, 32% are European American, and 12% were either biracial or other ethnic minorities. Approximately 52% of the students were girls. The African and European American households were characterized by wide and similar ranges of income, occupational, and educational statuses for the duration of the study. The median household income was $41,265 for African Americans and $46,822 for European Americans. Because the corresponding national incomes were $18,676 and $31,231, the county was more affluent than the nation at large. 86% of primary caregivers reported being employed and 55% of them had received college degree (38% for African Americans and 60% for European Americans).

The MADICS subsample was based on parents' willingness to participate and on a stratified sampling procedure designed to get proportional representations of families from each of 23 middle schools in the district. To examine sampling bias, the district provided selected data on all students in the district during the study years ($N = 25{,}627$), including students who joined the school system partway through the year and those from whom we did not receive parental consent. Cook and his colleagues evaluated the differences between our sample and the data for the entire district. They found only two significant differences across a full range of sociodemographic and school achievement data: our sample has a small underrepresentation of boys and students who were receiving subsidized lunches. These differences amounted to less than .1 of a standard deviation. Thus, there was some bias in the sample relative to the county population, but it did not lead to nonrepresented values at the extremes (see Cook, Herman, Phillips, & Setterstein, 2002 for details).

2.3. Measures

Using the student self-report items in the MADICS dataset, we developed scales for adolescents' school engagement in eighth grade (Eccles et al., 1993; Fredricks et al., 2004). Experts in the area of engagement (e.g., Drs. Blumenfeld, Fredricks, and Eccles) were consulted in the development of the school engagement scale. We were in direct contact with these scholars while these scholars were developing their theoretical framework on engagement and took their advice into account in developing the full set of engagement and achievement items in this dataset. Therefore, the item development and selection were drawn from multidimensional engagement theory and reviewed by experts. We hypothesized that there were three second-order factors underlying the six first-order factors that measured adolescents' school engagement: behavioral, emotional, and cognitive engagement. The items were coded so that the higher scores indicate higher school engagement (see Appendix A for an overview of the item questions used in the study).

2.3.1. Behavioral engagement

The construct of behavioral engagement was measured by two sets of factors: items from the Attentiveness subscale and items from the School Compliance subscale. The three-item subscale Attentiveness measures the extent to which students report being distracted in classes and have trouble getting schoolwork done. A sample item is "How often do you have trouble paying attention in classes?" Responses for each item were rated along a 5-point scale, ranging from 1 (*almost never*) to 5 (*almost always*). Some item responses for this indicator were reverse coded, so that higher scores indicated higher levels of attentiveness. The five-item scale School Compliance describes the levels to which students engage in misconduct in school. A sample item is "How often have you skipped classes?" Responses for each item were rated along a 5-point scale, ranging from 1 (*almost never*) to 5 (*almost always*). Item responses for this indicator were reverse coded, so that higher scores indicate higher levels of compliance.

2.3.2. Emotional engagement

Two factors were included to represent the construct of emotional engagement: the School Belonging scale and the Valuing of School Education scale. The three-item scale School Belonging represents the extent to which adolescents feel personally accepted, respected, and supported by adults and peers in school. A sample item is "In general, I feel like a real part in this school." Responses for each item were rated along a 5-point scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The six-item scale Valuing of School Education assesses adolescents' interest and belief in the importance and relevance of the general goals of education and academic achievement espoused by the school. A sample item is "I have to do well in school if I want to be a success in life." Responses for each item were rated along a 5-point scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

2.3.3. Cognitive engagement

We included two factors to represent the construct of cognitive engagement: the Self-Regulated Learning scale and the Cognitive Strategy Use scale. The four-item scale Self-Regulated Learning represents adolescents' perceived ability of self-monitoring and evaluation. A sample item is "How often are you very good at carrying out the plans you make for solving problems?" Responses for each item were rated along a 5-point scale, ranging from 1 (*almost never*) to 5 (almost always). The four-item scale Cognitive Strategy Use measures adolescents' perceived use of strategic approach to learning. A sample item is "How often do you try to relate what you are studying to other things you know about?" Responses for each item were rated along a 5-point scale, ranging from 1 (*almost never*) to 5 (*almost always*).

2.4. Procedures

Seventh-grade students were recruited from 23 schools to participate in this study through a letter that was sent home. Families who were interested in participating in this study were asked to sign and return a consent form. Subsequently, a questionnaire composed of Likert-type rating scales was administered to adolescents by research assistants (Cook et al., 2002). This data collection process took place in the home, and the self-administered questionnaire took approximately 20 min to complete. Participating adolescents were offered $20.

2.5. Data analysis

We examined the hypothesized factor structure of the measurement of school engagement, its invariance by gender and race/ethnicity, and the latent mean differences on the basis of mean and covariance structures, within the framework provided by CFA. The item-level correlations were the foundation of the CFA. The analyses were conducted using the MLR estimator in Mplus 5.10 program (Muthén & Muthén, 2007). Mplus was chosen, in part, because the current dataset includes data from students nested within 23 schools, and Mplus is able to account for the nested nature of our data by fitting a multilevel model with random-effects and produce correctly adjusted standard errors in the model estimations.

The amount of missing data was less than 6% and the data were missing completely at random, as evidenced by non-significant results derived from the generalized least squares combined test of homogeneity of means and covariance matrices representing complete and incomplete data, χ^2 (1,154) = 982.34, $p = .12$ (Bentler, 2005; Little & Rubin, 1987). We dealt with the missing data through full-information maximum likelihood estimation, allowing us to include all available data and identifying the parameter values that have the highest probability of producing the sample data (Baraldi & Enders, 2010).

To address our research questions, we carried out the following steps. First, to assess the dimensionality of school engagement, we examined the fit of a first-order factor model (full measurement model) by permitting all factors to intercorrelate. After establishing the adequacy of the first-order measurement models, we modeled these six factors as reflective indicators of second-order behavioral, emotional, and cognitive dimensions of engagement to assess whether the six first-order factors of engagement could be subsumed into the hypothesized second-order factor model (see Fig. 1).

In evaluating goodness of fit of the hypothesized models with MLR estimator, we used the χ^2 statistic based on the Satorra–Bentler scaled statistic (SBχ^2; Satorra & Bentler, 1988) and other goodness-of-fit indices to assess model fit. The SBχ^2 incorporates a scaling correction for the χ^2 statistic and takes into account the model, the estimation method, and the sample kurtosis values (Hu, Bentler, & Kano, 1992). In addition, we reported the root mean square error of approximation (RMSEA; Steiger, 1990), the standardized root mean square residual (SRMR; Hu & Bentler, 1998), and the comparative fit index (CFI; Bentler, 1990) as supplements to assess model fit. Structural equation modeling (SEM) literature suggests that model fit is excellent when the coefficient for CFI is greater than .95; and model fit is deemed adequate if the coefficient is greater than .90 (Byrne, 2001; Hu & Bentler, 1999). For the RMSEA and SRMR, a coefficient less than .05 indicates an excellent fit and a coefficient under .08 indicates an acceptable fit (Kline, 1998).

Second, a series of nested factor models were fit in our analyses in order to test measurement invariance, following the general procedures suggested by Widaman and Resie (1997). Each pair of models in the testing sequence was nested because a set of parameters were constrained to be equal across groups in the more restricted model. To assess measurement invariance, one of the factor loadings was fixed to 1 and constraints were added sequentially.

Testing for configural invariance (Model 1)

We constrained the number of factors and pattern of fixed and free factor loadings for the first- and second-order factor loadings to be the same across gender and race/ethnicity groups. However we also allowed different estimates for the corresponding parameters in the different groups. This unconstrained multi-group model served as a baseline model against which we evaluated the fits of successively more restrictive models.

Testing for first-order factor loading invariance (Model 2)

To test for factorial invariance at this level, we constrained all first-order factor loadings to be equal across groups.

Testing for second-order factor loading invariance (Model 3)

To test for factorial invariance at this level, we constrained all first- and second-order factor loadings to be equal across groups.

Testing for intercept of observed variables invariance (Model 4)

In testing this form of invariance, we placed equality constraints on the intercepts of observed variables, in addition to the constraints on the factor loadings of the latent variables.

Testing for intercept of first-order latent factorial invariance (Model 5)

In testing for this level of invariance, we constrained first- and second-order loadings and the intercepts of the observed variables and the means of the first-order factors to be equal across groups. When this level of invariance is achieved, it suggests that the units of measurement (factor loadings), as well as the origins of the scales (intercepts) are equivalent across groups, and thus we can compare estimates of the second-order factor means across groups (Widaman & Resie, 1997).

To compare the fit of two nested models, we used a test based on the difference in the SBχ^2 statistic, albeit a correction to the value is needed as this difference is not distributed as χ^2 statistic (Bentler, 2005). If the difference in the SBχ^2 statistic is greater than the critical value, this difference suggests that the constraint imposed in the more restrictive model is not equivalent across groups. On the other hand, if the difference in the SBχ^2 statistic is statistically nonsignificant, it suggests that the specified equality constraints are tenable; the more restricted model should be accepted. However, researchers using this approach have argued that the difference in the SBχ^2 statistic is sensitive to large sample sizes and any nonnormality among the indicators (Fan, Thompson, & Wang, 1999; Tomarken & Waller, 2003); thus, one is recommended to use additional goodness-of-fit indexes to further assess the model fit (e.g., Cheung & Rensvold, 2002; Little, 1997; Marsh, Hey, & Roche, 1997). For this study, we relied on guidelines proposed by Cheung and Rensvold (2002), who suggested that a difference of larger than .01 in the CFI indicates a meaningful difference in model fit for testing measurement invariance. Therefore, both the difference in the SBχ^2 statistic test and difference in the value of the CFI were used to assess model fit. In cases in which the two approaches disagree, we relied more on the changes in CFI due to the large sample size in this study (Chen et al., 2005).

Finally, building on evidence of invariant factor loadings and intercepts at both the first-order and the second-order, we tested the mean differences on the second-order factors by group (Model 6). In this case, we set the factor means for one group, the reference group, to zero whereas we estimated the comparison groups' factor means freely. These freely estimated latent means then represent the difference between the factor means of the two groups. We used a z statistic to test differences between the latent means of the two groups (Aiken, Stein, & Bentler, 1994).

3. Results

3.1. Assessing the measurement model

Confirmatory factor analysis (CFA) of the students' responses to the 23 items of school engagement verified that the hypothesized factors were measured by discrete, single latent variables. In Fig. 2, we present the factor models with all of the parameter estimates displayed. The six factors were allowed to inter-correlate simultaneously to specify the measurement model, which represents a first-order model. The standardized

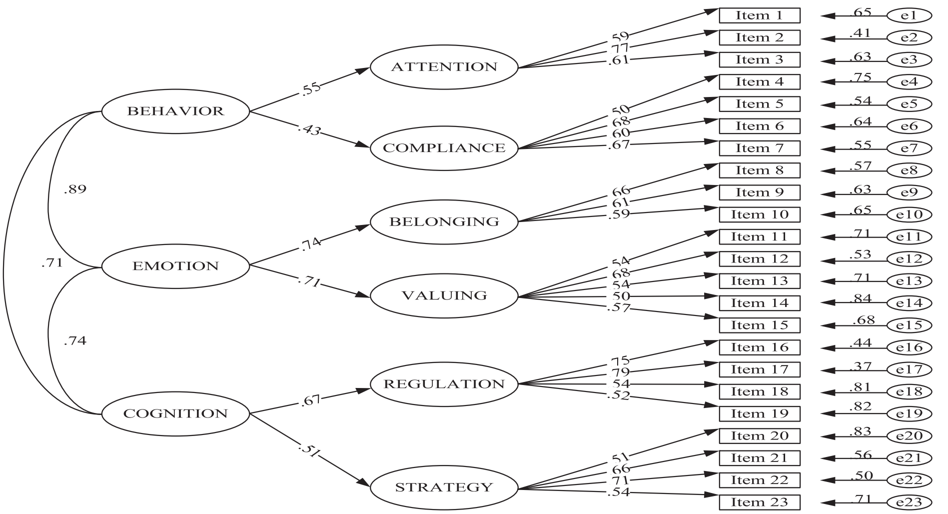

Fig. 2. Final factor model of school engagement depicting a second-order factor structure hypothesized to underlie the six first-order factors. All standardized factor loadings are statistically significant ($p<.001$). Please see complete item question in Appendix A.

loadings ranged from .50 to .79 and were all statistically significant at the .05 level. We found that the six factors provided acceptable fit for these data, SBχ^2 (271, N = 1,103) = 634.44, p < .001, CFI = .914; SRMR = .043; RMSEA = .042. Thus, the proposed six-factorial structure is consistent with the data.

We found expected patterns in the bivariate correlations of the factors. Attentiveness, School Compliance, School Belonging, Valuing of School Education, Self-Regulated Learning, and Cognitive Strategy Use were positively correlated with each other. The correlations among the six factors are moderate to high (.25–.53), suggesting that a second-order model is plausible (Glanville & Wildhagen, 2007).

3.2. Assessing dimensionality and factor structure of school engagement

In Table 1, we present goodness-of-fit statistics for the first- and second-order measurement models. We tested several models, beginning with the first-order model (Model 1a) in which we hypothesized that there were six interdependent factors. To assess whether the six first-order factors of engagement could support a single second-order engagement factor, we first fit a second-order model in which we included only one second-order engagement factor (Model 2a). Unsurprisingly, given our argument that research should distinguish between different dimensions of engagement, the fit of Model 2a was statistically significantly worse than the first-order model in which it was nested, ΔSBχ^2 = 89.57, df = 9, p < .001; and ΔCFI = .019. We further compared the first-order model with a two-factor second-order model (Model 3a), and the fit of Model 3a was still statistically significant worse than the first-order model, ΔSBχ^2 = 70.35, df = 8, p < .001; and ΔCFI = .015. We then estimated a second-order model that distinguished among behavioral, emotional, and cognitive engagement (Model 4a). Differences in fit between this model and the corresponding first-order model were not statistically significant, ΔSBχ^2 = 4.77, df = 6, p = .23, ΔCFI = 0, and the other indexes of fit were no worse than in Model 1a. Therefore, based on theory and the good fit of the model that specifies separate dimensions of engagement, we chose Model 4a as our preferred model, and the remaining reported results were based on this model.

In Fig. 2, we present the estimated standardized factor loadings for Model 4a. All of the first- and second-order factor loadings are non-zero, providing evidence of convergent validity (Anderson & Gerbing, 1988). The composite reliability coefficients for each first-order latent variable are greater than or equal to .70 (Attentiveness = .70; School Compliance = .78; School Belonging = .75; Valuing of School Education = .72; Self-Regulated Learning = .78; and Cognitive Strategy Use = .77; Bagozzi & Yi, 1988). In sum, these results provide empirical evidence for the theoretical distinction proposed by Fredricks et al. (2004) to exist between the behavioral, emotional, and cognitive components of school engagement and thus support their call to measure engagement as a multidimensional, rather than uni- or bi-dimensional concept. In addition, the results support our assumption that behavioral, emotional, and cognitive engagements are higher order factors that explain the covariance among the six first-order factors.

3.3. Assessing measurement invariance by gender

Configural invariance (Model 1b)

The configural model, in which we imposed no equality constraints between groups, served as our baseline model. As shown in Table 2, the goodness of fit statistics suggest an adequate fit for this model,

Table 1
Goodness-of-fit statistics from the fitting of the hypothesized first- and second-order factor models (N = 1,103).

Model	SBχ^2	df	RMSEA	SRMR	CFI	Model Comparison	ΔSBχ^2	Δdf
1a. First-order	634.44*	271	.042	.043	.914	__	__	__
2a. Second-order, one second-order factor	728.21*	280	.046	.047	.895	2 vs.1	89.57*	9
3a. Second-order, two second-order factors [a]	710.49*	279	.045	.045	.899	3 vs.1	70.35*	8
4a. Second-order, three second-order factors	641.12*	277	.042	.044	.914	4 vs.1	4.77	6

Note. SBχ^2 = Satorra–Bentler scaled statistic, CFI = comparative fit index, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual.

[a] We fit a second-order model in which behavioral engagement drives attentiveness, school compliance, school belonging, and valuing of school education, whereas cognitive engagement influences self-regulated learning and cognitive strategy use.

* p < .001.

SBχ^2 = 966.35, df = 499, p < .001; RMSEA = .047; SRMR = .047; and CFI = .904. The results provide evidence for the viability of the second-order factor structure of the school engagement and suggest that the factors are adequately defined and measured for both girls and boys.

Invariance of first-order factor loadings (Model 2b)

When all first-order factor loadings were constrained to be equal across gender, the difference in SBχ^2 between Models 2b and 1b was small, ΔSBχ^2 = 17.39, df = 17, p = .12, and the CFI values differed by no more than .01, the cutoff value suggested by Cheung and Rensvold (2002). These results indicate that the first-order factor loadings are invariant by gender.

Invariance of second-order factor loadings (Model 3b)

When all first- and second-order factor loadings were constrained to be equal by gender, the difference in SBχ^2 statistic between Models 3b and 2b was small and not statistically significant, ΔSBχ^2 = 5.13, df = 3, p = .11, and the CFI value remained the same. These results indicate that the second-factor loadings are invariant across girls and boys.

Invariance of intercepts of observed variables (Model 4b)

In addition to the constraints already imposed on the first- and second-order factor loadings, when the intercepts of the observed variables were constrained to be equal by gender, the difference in SBχ^2 statistic value between Models 4b and 3b was large and statistically significant, ΔSBχ^2 = 59.62, df = 17, p < .001, and d = 0.09. However, because of the concerns about the overly sensitive nature of the test to sample size, we also used the CFI index to assess difference in model fit. It declined less than .01(.928 vs. .925), and this indicates that there are no substantial differences between Models 4b and 3b. Based on this criterion, there appears to be no substantial difference in the intercepts of the observed variables between girls and boys.

Invariance of intercepts of first-order factor (Model 5b)

When first- and second-order factor loadings and the intercepts of the observed variables and first-order factor means were constrained to be equal by gender, the difference in the SBχ^2 statistic between Models 5b and 4b was large and statistically significant, ΔSBχ^2 = 47.62, df = 6, p < .001, d = 0.10; however, because there was no substantial change in CFI (.925 vs. .924), we conclude that the intercepts of the first-order factors are invariant by gender.

3.4. Assessing measurement invariance by race/ethnicity

To test whether the second-order factor structure is equivalent between African Americans and European Americans, we examined a series of nested models, similar to those we conducted for gender

Table 2
Summary of fit statistics for testing the measurement invariance of hypothesized second-order factor model of school engagement by gender (N = 1,103).

Model	SBχ^2	df	RMSEA	SRMR	CFI	Model Comparison	ΔSBχ^2	Δdf
1b. Configural invariance	966.35	499	.047	.051	.929	__	__	__
2b. First-order factor loadings invariant	985.02	516	.047	.053	.928	M2 vs. M1	17.39	17
3b. First- and second-order factor loadings invariant	991.05	519	.047	.053	.928	M3 vs. M2	5.13	3
4b. First- and second-order factorings and intercepts of observed variables invariant	1054.40	536	.048	.054	.925	M4 vs. M3	59.62*	17
5b. First- and second-order factor loadings, and intercepts of observed variables and first-order factors invariant	1106.05	542	.048	.055	.924	M5 vs. M4	47.62*	6
6b.Second-order factor mean differences	1101.32	539	.050	.057	.918	__	__	__

Note. SBχ^2 = Satorra–Bentler scaled statistic, CFI = comparative fit index, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual.
* p < .001.

(see Table 3). Due to a small number of students who were categorized as "other minority" ethnicities, we only focused on the comparison between African American and European American students in this study.

In the analysis of measurement invariance across the African American and European American samples, the baseline model (see Model 1c in Table 3) again fits well and supports the validity of the second-order factor structure of school engagement. Starting with the baseline model, we assessed the invariance of first- and second-order factor loadings (see Models 2c and 3c in Table 3). The models with factor loadings constrained across groups fit well and no extreme or inconsistent loadings were encountered based on the criteria of the $SB\chi^2$ difference test and the change of CFI value, Model 2c: $\Delta SB\chi^2=23.26$, $df=17$, $p=.08$ and Model 3c: $\Delta SB\chi^2=4.13$, $df=3$, $p=.18$. The results suggest that the first- and second-order factor loadings for the African American and European American students are consistent. In the next step, we added intercept parameters of item and first-order factors to the model (see Model 4c and 5c in Table 3). In these models, the $SB\chi^2$ difference tests were significant, Model 4c: $\Delta SB\chi^2=71.52$, $df=17$, $p<.001$, $d=0.11$ and Model 5c: $\Delta SB\chi^2=42.38$, $df=6$, $p<.001$, $d=0.07$. However, when we examined the change in the CFI, we found no changes that were greater than the standard and therefore conclude that Model 5 represents the ultimate level of invariance that could be achieved, suggesting that there is no substantial difference in the intercepts of the observed variables and in the first-order factors between African American and European American students.

3.5. Assessing group differences in the second-order factor means

To estimate the difference between the higher-order factor means by gender, we chose the girls as a reference group and set their second-order latent means to zero. The latent means of the boys then captures the difference in factor means between the two groups. The tests revealed statistically significant mean differences by gender on two of the second-order factors. Specifically, boys had lower scores on (a) behavioral engagement and (b) emotional engagement than girls (behavioral engagement difference $=-.39$, $z=-2.59$, $p=.010$, $d=0.25$ and emotional engagement difference $=-.46$, $z=-4.22$, $p<.001$, $d=0.32$). The gender difference in cognitive engagement was not statistically significant (difference $=-.06$, $z=-.591$, $p=.555$, $d=0.07$).

In terms of latent means differences by race/ethnicity, we selected the African American group as a reference group and estimated the latent mean of the European American group. Results suggest that there were two statistically significant means differences between African American and European American students on the second-order factors. European American students had lower scores on (a) emotional engagement than African American students (difference $=-.23$, $z=-2.09$, $p=.037$, $d=0.20$) whereas European American students had higher scores on (b) behavioral engagement than African American students (.25, $z=2.46$, $p=.013$, $d=0.19$). The difference in cognitive engagement between African American and European American students was not statistically significant (difference $=.09$, $z=.721$, $p=.471$, $d=0.05$). According to the magnitude of these coefficients, it is noteworthy that gender differences are larger than racial/ethnic differences.

Table 3
Summary of fit statistics for testing measurement invariance of hypothesized second-order factor model of school engagement by race/ethnicity ($N=1,103$).

Model	$SB\chi^2$	*df*	RMSEA	SRMR	CFI	Model Comparison	$\Delta SB\chi^2$	Δdf
1c. Configural invariance	934.06	499	.052	.055	.915	__	__	__
2c. First-order factor loadings invariant	959.48	516	.052	.059	.911	M2 vs. M1	23.26	17
3c. First- and second-order factor loadings invariant	966.52	519	.052	.064	.910	M3 vs. M2	4.13	3
4c. First- and second-order factorings and intercepts of observed variables invariant	1045.32	536	.053	.064	.902	M4 vs. M3	71.52*	17
5c. First- and second-order factor loadings, and intercepts of observed variables and first-order factors invariant	1093.20	542	.054	.069	.901	M5 vs. M4	42.38*	6
6c. Second-order factor mean differences	1084.60	539	.054	.068	.900	__	__	__

Note. $SB\chi^2$ = Satorra–Bentler scaled statistic, CFI = comparative fit index, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual.
* $p<.001$.

4. Discussion

In the present study, we tested the multidimensional theoretical conceptualization of school engagement proposed by Fredricks et al. (2004). To do this, we examined the psychometric properties and measurement invariance of a series of questionnaire items that were designed to measure middle-school students' levels of school engagement. Specifically, based on Fredricks et al. (2004), we hypothesized that there was a second-order factor structure for the school engagement, with six first-order factors, and behavioral, emotional, and cognitive engagement as the three second-order factors. The study contributes to literature on the measurement of school engagement in several ways. First, we confirmed the good fit of the six-factor engagement measurement model to our data and found empirical evidence to support the second-order engagement factor structure. Second, there is empirical evidence for measurement invariance of these underlying engagement factors for boys and girls and European American and African American students. This measurement feature is critical for future studies designed to look at group differences in the origins and consequences of school engagement. Finally, in terms of second-order factor mean differences, we found the expected gender differences in behavioral and emotional engagement, providing one indicator of the validity of our measures. We also found differences in these same indicators across our two racial/ethnic groups.

4.1. School engagement as a multidimensional construct

Consistent with the recent literature on school engagement, the findings support Fredricks et al. (2004) theory that school engagement is a multidimensional construct. The hypothesized second-order factor model demonstrates that responses to the measurement of engagement can be grouped into six reliable first-order factors (Attentiveness, School Compliance, Valuing of School Education, School Belonging, Self-Regulated Learning, and Cognitive Strategy Use). Moreover, we found that these six factors can be further grouped into three second-order factors (Behavioral, Emotional, and Cognitive Engagement) that subsume the six first-order factors. Compared to a correlated first-order six-factor model, the second-order factor model is more parsimonious and provides theoretically error-free estimates of each first- and second-order factor. Future research should examine whether the different dimensions of engagement have different precursors and consequences.

4.2. Measurement invariance across gender and race/ethnicity

In this study, we extend the traditional CFA approach by using covariance structure and mean structure to analyze the hypothesized invariance of factor structure, factor loadings, intercepts, and mean differences among subgroups of students. The results provide empirical evidence supporting measurement invariance of school engagement by gender and racial/ethnic groups. That is, for group comparisons by gender and race/ethnicity, the same construct of school engagement appears to be measured and in the same metric. Hence, if any difference in the factor score is found, it is likely that they reflect possible group differences in the amount of school engagement rather than item bias or measurement artifact. In addition, we are more confident that comparing and explaining variation is meaningful regardless of the group membership because cross-group variances are confirmed to be on the same metric. There has been a growing concern that certain groups of students (i.e., African American students and boys) may be at higher risk for low engagement in school. Establishing measurement invariance allows researchers to make more appropriate comparisons between groups such as boys and girls and those from different cultural groups. Establishing measurement invariance also provides support for pooling the data of boys and girls as well as European American and African American students in subsequent data analysis, particularly in relation to the facets of engagement proposed in this study.

Given that both factor loadings and intercepts were invariant, we examined the latent mean differences on the second-order factors of behavioral, emotional, and cognitive engagement. Regarding gender differences, we found the mean scores for girls to be higher than boys on the behavioral as well as emotional engagement but not on the cognitive engagement. These results are consistent with other recent research (e.g., Johnson et al., 2001; Voelkl, 1997) and thus provide predictive validity for our scales. With regard to racial/ethnic differences, the results suggest that African American students report higher scores of emotional engagement but are less likely to be behaviorally engaged when compared to European American students. With the empirical support of measurement invariance, we are more confident in our conclusion that the findings of a discrepancy in behavioral and emotional engagement between African American and European American students may reflect possible racial/ethnic differences rather than construct bias or measurement error between the two groups in our sample. Future research should investigate why such between-group discrepancies exist and examine whether there are gender or racial/ethnic differences in the relationships between school engagement and its causes and consequences.

4.3. Limitations

Several limitations of this study and some caveats need to be noted. First, the present data relies mainly upon self-report information from students to assess school engagement, which raises an important validity concern. Students may be influenced by social demands to answer in a socially desirable way about their own behavior or attitude, thus introducing bias into the results. Future use of multiple sources of information (e.g., teachers and parents) and multiple methodologies (e.g., interviews and observations) can provide a more robust, valid method of identifying school effects related to engagement (Fredricks et al., 2004). For instance, school record or teacher-report data may provide a more accurate source for the behavioral engagement information. Second, these findings are specific to the United States and a largely urban sample on the East Coast. It is important to recognize that adolescents in suburban or rural areas may have different experiences with school engagement. The conclusions cannot be generalized beyond the conditions of the county investigated in this study. Future studies should use samples from different age groups across multiple environments to provide empirical support for the engagement instrument, as well as to potentially support the generalizability of the current findings. Third, different results were obtained depending on the statistical criterion used (SBχ^2 difference test or changes in CFI). There is currently a lack of conventionally accepted criteria for evaluating the fit of invariance tests in multigroup analyses. It is also noteworthy that the CFI suggests that our final model fit was adequate as opposed to excellent. Thus, additional research should be conducted to verify the measurement invariance across groups. Finally, we did not have a sufficiently large sample to examine socioeconomic status and other important socio-demographic and psychological characteristics as well as both race/ethnicity and gender. We hope that future studies will be conducted to look at the association of these many critical characteristics with school engagement and with changes in school engagement across time.

4.4. Implications for practice

Our analyses provide empirical support to school psychologists for assessing the student engagement in their schools as a first step towards designing appropriate school interventions. School psychologists may identify a set of common factors of student engagement on which students vary in degree and which can be targeted with intervention. It could be a useful instrument in the general practice of school psychologists to immediately assess class or school levels of engagement across six salient dimensions. The examination of mean differences helps identify the groups of students who are at greater risk for academic failure and potential for dropping out, and remedies should be sought for them in particular. School interventions might be studied in conjunction with this engagement instrument in the future to help address student disengagement.

The findings suggest that uni-dimensional interpretations for school engagement may not reveal the discrete types of engagement. As research has indicated that school engagement links to students' academic outcomes (e.g., Finn & Rock, 1997), school psychologists may consider aiming intervention at multiple dimensions of school engagement for students in order to enhance their academic achievement as well as to reduce school drop-out rates. Students with patterns of lower scores in specific dimensions or generally depressed scores across all dimensions could be identified for follow-up evaluation, and more comprehensive open-ended clinical consultation models could be invoked. Prevention efforts may be more effective when they involve integration of distinct strategies that address students' behavioral, emotional, and cognitive needs.

4.5. Implications for future research

These findings hold substantive and methodological implications for researchers studying issues relevant to engagement and to school psychologists seeking to enhance educational outcomes that rely in large part on the extent to which their students are behaviorally, emotionally, and cognitively engaged with school. Given the importance of the concept of school engagement discussed in the present study, it is critical that studies of school engagement use the best possible measurement strategies to insure valid inferences. Future research on the conceptualization and assessment of adolescent engagement in school can be enhanced by the conceptual argument and empirical strategies presented in the study. Moreover, the measurement invariance of the factor structure also has implications for school-based intervention programs. Having such invariant instruments may be valuable for school psychologists who might want to assess group differences within their school or might want to determine whether interventions have differential effects on various subgroups within their school.

Acknowledgments

The article is adapted from a doctoral dissertation by Ming-Te Wang submitted to the Graduate School of Education at Harvard University. This research was supported by dissertation grants from the American Psychological Association (Division 15) and Harvard University awarded to Ming-Te Wang. The first author would like to thank Robert Selman and Stephanie Jones, members of his dissertation committee, and Terry Tivnan for their helpful feedback.

Appendix A. Item questions and descriptive statistics for school engagement measures

Factor	Item	Item questions	Mean (*SD*)	α
Behavioral engagement			2.65 (0.79)	.82
Attentiveness	1	*How often do you have trouble paying attention in classes?	3.24 (0.87)	.79
	2	How often do you get schoolwork done on time?		
	3	*How often do you find that it is hard for you to keep your mind on your work in school?		
School compliance	4	*How often have you hit someone for what they said/did?	2.19 (0.63)	.83
	5	*How often have you been involved in a physical fight?		
	6	*How often have you been sent to office?		
	7	*How often have you skipped class?		
Emotional engagement			3.61 (0.77)	.86
School belonging		Please indicate how strongly do you agree or disagree with each of the following statements.	3.30 (0.81)	.82
	8	I feel happy and safe in this school.		
	9	In general, I feel like a real part in this school.		
	10	I would recommend to other kids that they go to my school.		
Valuing of school education		Please indicate how much you agree or disagree with the following statements.	3.78 (0.65)	.85
	11	I have to do well in school if I want to be a success in life.		
	12	*Schooling is not so important for kids like me.		
	13	*I learn more useful things from my friends and relatives than I learn in school.		
	14	Getting good education is the best way to get ahead in life for the kids in my neighborhood.		
	15	I often learn a lot from my school work.		
Cognitive engagement			3.60 (0.74)	.82
Self-regulated learning	16	How often do you try to figure out problems and planning how to solve them?	3.82 (0.65)	.78
	17	How often do you try to carry out the plans you make for solving problems?		
	19	How often do you try to bounce back quickly from bad experiences?		
	19	How often do you try to learn from your mistakes?		

Appendix A (*continued*)

Factor	Item	Item questions	Mean (*SD*)	α
Cognitive strategy use	20	When you are doing homework or school work, how often do you try to decide what you are supposed to learn, rather than just read the material?	3.32 (0.78)	.80
	21	How often do you try to relate what you are studying to other things you know about?		
	22	How often do you try to plan what you have to do for homework before you get started?		
	23	How often do you make sure you get started on it early?		

Note. *Item reverse coded.

References

Aiken, L. S., Stein, J. A., & Bentler, P. M. (1994). Structural equation analyses of clinical subpopulation differences and comparative treatment outcomes: Characterizing the daily lives of drug addicts. *Journal of Consulting and Clinical Psychology*, *62*, 488–499.

Ainsworth-Darnell, J. W., & Downey, D. B. (1998). Assessing the oppositional culture explanation for racial/ethnic differences in school performance. *American Sociological Review*, *63*, 536–553.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*, 411–423.

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.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*, 74–94.

Bandura, A., Barbaranelli, C., Caprar, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development*, *67*, 1206–1222.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*, 5–37.

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

Bentler, P. M. (2005). *EQS 6.1: Structural equations program manual.* Encino, CA: Multivariate Software.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley.

Bowen, G. L., & Richman, J. M. (1995). *The school success profile.* Chapel Hill: University of North Carolina at Chapel Hill.

Brooks-Gunn, J., & Duncan, G. (1997). The effects of poverty on children and youth. *The Future of Children*, *7*, 55–71.

Byrne, B. M. (1998). *Structural Equation Modeling with Lisrel, Prelis, and Simplis: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum.

Byrne, B. M., & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of s second-order structure: A walk through the process. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*, 287–321.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, *34*, 155–175.

Caraway, K., Tucker, C. M., Reinke, W. M., & Hall, C. (2003). Self-efficacy, goal orientation, and fear of failure as predictors of school engagement in high school students. *Psychology in the Schools*, *40*, 417–427.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, *12*, 471–492.

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

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. R. Gunnar, & L. A. Sroufe (Eds.), *Self processes in development: Minnesota symposium on child psychology*, *Vol. 23*. (pp. 43–77) Chicago: University of Chicago Press.

Cook, T. D., Herman, M. R., Phillips, M., & Setterstein, R. A. (2002). Some ways in which neighborhoods, nuclear families, friendship groups, and schools jointly affect changes in early adolescent development. *Child Development*, *73*, 1283–1309.

Csikszentmihalyi, M., & Schneider, B. (2000). *Becoming adult: How teenagers prepare for the world.* New York: Basic Books.

Daly, B. P., Shin, R. Q., Thakral, C., Selders, M., & Vera, E. (2009). School engagement among urban adolescents of color: Does perceptions of social support and neighborhood safety really matter? *Journal of Youth and Adolescence*, *38*, 63–74.

Eccles, J. S., Midgley, C., Buchanan, C. M., Flanagan, C., Mac Iver, D., Reuman, D., et al. (1993). Development during adolescence: The impact of stage/environment fit. *The American Psychologist*, *48*, 90–101.

Fan, X., Thompson, B., & Wang, L. (1999). The effects of sample size, estimation methods, and model specification on SEM fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 56–83.

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

Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. *The Journal of Applied Psychology*, *82*, 221–234.

Finn, J. D., & Voelkl, K. E. (1993). School characteristics related to school engagement. *Journal of Negro Education*, *62*, 249–268.

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.

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.

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research*, *Vol. 3*. (pp. 403–422). Mahwah, NJ: Erlbaum.

Hu, L. -T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453.
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.
Hu, L. -T., Bentler, B. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351–362.
Institute for Research and Reform in Education (IRRE) (1998). *Research Assessment Package For Schools (RAPS) Manual.* Philadelphia, PA: Institute for Research and Reform in Education.
Jimerson, S. R., Campos, E., & Greif, J. L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *The California School Psychologist*, *8*, 7–27.
Johnson, M. K., Crosnoe, R., & Elder, G. (2001). Students' attachment and academic engagement: The role of the race and ethnicity. *Sociology of Education*, *74*, 318–340.
Kline, R. B. (1998). *Principals and practices of structural equation modeling.* New York, NY: Guilford Press.
Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*, 53–76.
Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.
Marks, H. M. (2000). Student engagement in instructional activity: Patterns in elementary, middle and high school years. *American Educational Research Journal*, *37*, 153–184.
Marsh, H. W., Hey, J., & Roche, L. A. (1997). Structural of physical self-concept: Elite athletes and physical education students. *Journal of Educational Psychology*, *89*, 369–380.
Martin, A. J. (2004). School motivation of boys and girls: differences of degree, differences of kind, or both? *Australian Journal of Psychology*, *56*, 133–146.
Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *The British Journal of Educational Psychology*, *77*, 413–440.
McDermott, P. A., Mordell, M., & Stolzfus, J. C. (2001). The organization of student performance in American schools: discipline, motivation, verbal learning, and nonverbal learning. *Journal of Educational Psychology*, *93*, 65–76.
Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.
Muthén, L. K., & Muthén, B. O. (2007). Mplus user's guide (5th ed). [Computer software manual]. Los Angeles: author.
National Research Council and Institute of Medicine (2004). *Engaging schools: Fostering high school students' motivation to learn.* Washington, DC.: The National Academies Press.
Newmann, F., Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagement. In F. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11–39). New York: Teachers College Press.
Perry, J. C., Liu, X., & Pabian, Y. (2010). School engagement as mediator of academic performance among urban youth: The role of career preparation, parental career support, and teacher support. *The Counseling Psychologist*, *38*, 269–295.
Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*, 552–566.
Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *American Statistical Association 1988 proceeding of the business and economics section* (pp. 308–313). Alexandria VA: American Statistical Association.
Simons-Morton, B., & Chen, R. (2009). Peer and parent influences on school engagement among early adolescents. *Youth and Society*, *41*, 3–25.
Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effect of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*, 571–581.
Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology*, *100*, 765–781.
Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.
Thompson, M. S., & Green, S. B. (2006). Evaluating between-group differences in latent variable means. In G. R. Hancock, & R. O. Mueller (Eds.), *A second course in structural equation modeling* (pp. 119–169). New York: Wiley.
Tomarken, A. J., & Waller, N. G. (2003). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology*, *1*, 31–65.
Van de Vijver, F. J. R., & Leung, K. (1997). Methods and data analysis for comparative research. In J. Berry, Y. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology*, *Vol. 1*. (pp. 259–300)Boston: Allyn & Bacon.
Voelkl, K. E. (1997). Identification with school. *American Journal of Education*, *105*, 204–319.
Wang, M. T., & Holcombe, R. (2010). Adolescents' perceptions of classroom environment, school engagement, and academic achievement. *American Educational Research Journal*, *47*, 633–662.
Wellborn, J. G., & Connell, J. P. (1987). *Manual for the Rochester assessment package for schools.* Rochester, NY: University of Rochester.
Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology*, *89*, 696–716.
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.
Yazzie-Mintz, E. (2007). *Voices of students on engagement: A report on the 2006 High School Survey of Student Engagement.* Bloomington: Center for Evaluation & Education Policy, Indiana University Retrieved December 13, 2009, from. http://ceep.indiana.edu/hssse.
You, S., & Sharkey, J. (2009). Testing a developmental-ecological model of student engagement: A multilevel latent growth curve analysis. *Educational Psychology*, *29*, 659–684.
Zimmerman, B. J. (1989). Models of self-regulated learning and academic achievement. In B. J. Zimmerman, & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theory, research and practice* (pp. 1–25). New York: Springer-Verlag.
Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, *82*, 51–59.