View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by ScholarlyCommons@Penn

University of Pennsylvania
ScholarlyCommons

GSE Publications

Graduate School of Education

7-2016

Problems in Classroom Engagement: Validation of an Assessment for District-wide use in the Early Primary Grades

Katherine Barghaus
University of Pennsylvania, barghaus@upenn.edu

John Fantuzzo
University of Pennsylvania, JOHNF@GSE.UPENN.EDU

Whitney LeBoeuf
University of Pennsylvania

Cassandra Henderson
University of Pennsylvania, chende@gse.upenn.edu

Feifei Li
University of Pennsylvania

See next page for additional authors

Follow this and additional works at: http://repository.upenn.edu/gse_pubs


Part of the Early Childhood Education Commons

Recommended Citation

Barghaus, K., Fantuzzo, J., LeBoeuf, W., Henderson, C., Li, F., & McDermott, P. (2016). Problems in Classroom Engagement: Validation of an Assessment for District-wide use in the Early Primary Grades. *Early Education and Development, 28* (2), 154-166. http://dx.doi.org/10.1080/10409289.2016.1197012

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/gse_pubs/420
For more information, please contact repository@pobox.upenn.edu.

Problems in Classroom Engagement: Validation of an Assessment for District-wide use in the Early Primary Grades

Abstract
Research Findings: The aim of this study was to provide an initial investigation into the psychometric properties of the Problems in Classroom Engagement Scale (PCES). The PCES was designed and tested for district-wide use as part of the report card system for a large urban school district. The PCES was administered to all first, second, and third grade students in the district. Factor analytic examination revealed a bifactor structure as the best fit to the data. The bifactor structure reflected a general factor of Problems in Behavioral Engagement and two key group factors: Problems in Social Engagement and Problems in Academic Engagement. These factors were found to be reliable within and across grades and demonstrated convergent and divergent relations with academic and behavioral outcomes. Practice or Policy: Findings provide initial evidence to support the routine use of PCES in a large, urban setting. As such, use of the PCES can help in fostering district-wide attention to students' early behavioral, social, and academic engagement difficulties.

Disciplines
Early Childhood Education | Education

Author(s)
Katherine Barghaus, John Fantuzzo, Whitney LeBoeuf, Cassandra Henderson, Feifei Li, and Paul McDermott

This journal article is available at ScholarlyCommons: http://repository.upenn.edu/gse_pubs/420

Problems in Classroom Engagement: Validation of an Assessment for District-wide use in the Early Primary Grades

Abstract

Research Findings: The aim of this study was to provide an initial investigation into the psychometric properties of the Problems in Classroom Engagement Scale (PCES). The PCES was designed and tested for district-wide use as part of the report card system for a large urban school district. The PCES was administered to all first, second, and third grade students in the district. Factor analytic examination revealed a bifactor structure as the best fit to the data. The bifactor structure reflected a general factor of Problems in Behavioral Engagement and two key group factors: Problems in Social Engagement and Problems in Academic Engagement. These factors were found to be reliable within and across grades and demonstrated convergent and divergent relations with academic and behavioral outcomes. *Practice or Policy:* Findings provide initial evidence to support the routine use of PCES in a large, urban setting. As such, use of the PCES can help in fostering district-wide attention to students' early behavioral, social, and academic engagement difficulties.

Problems in Classroom Engagement: Validation of an Assessment for District-Wide Use in the Early Primary Grades

Findings from international studies of academic achievement indicate that students in the United States perform below many other countries in mathematics, reading, and problem solving compared with students from other countries (Lemke et al., 2004). Concerns about the educational well-being of American children, especially for those in large urban areas, have intensified as we are confronted by persistent, low proficiency rates and educational inequities (Vigdor & Ludwig, 2007). The National Assessment of Educational Progress (NAEP) reports that only 33% of all fourth graders meet reading proficiency standards and that this percentage drops to 15% for students in large urban public school districts (U.S. Department of Education, 2009).

There is a growing body of research indicating that poor academic performance and behavioral outcomes are associated with problems of student engagement in the basic academic and social experiences of schooling (Chang & Romano, 2008; Christenson & Thurlow, 2004; Li-Grining, Votruba-Drzal, Maldonado-Carrefio, & Haas, 2010; Matthews, & Kizzie, 2010; McClelland, Acock, Piccinin, Rhea, & Stallings, 2013). For students in large urban school districts, significant problems with classroom engagement are evident in elementary school (Pianta, Cox, & Snow, 2007). In the largest national survey of teachers, 62% of urban elementary school teachers reported "students coming to school unprepared to learn" (e.g., not having materials organized, not exerting effort, etc.) as a moderate to serious problem in their school compared to only 40% of suburban teachers (U.S. Department of Education, 2009). In addition to impacting readiness to learn, research has found that problems in classroom engagement are associated with negative academic achievement and behavioral outcomes, such as truancy and suspension (Fredricks, Blumenfeld, & Paris, 2004). It is therefore not surprising that the nation's

governors have made it a national priority to better understand classroom engagement to help primary school students become active classroom learners and prevent future disengagement and educational challenges (U.S. Department of Education, 2005).

Recent research on student engagement in the classroom defines it as a multidimensional construct that is linked to success in school (Christenson, Reschly, & Wylie, 2012). Newer models of engagement define it as broadly reflecting cognitive, emotional or affective, and behavioral engagement (Christenson, Reschly, & Wylie, 2012). Cognitive engagement focuses primarily on intra-individual processes related to self-efficacy beliefs and mastery orientation (Finn & Zimmer, 2012). Emotional engagement also focuses on intra-individual experiences, including students' affective reactions to the classroom setting and feelings about the value of schooling (Finn & Zimmer, 2012). Behavioral engagement is operationalized to included academic and social engagement (Finn & Zimmer, 2012). Academic engagement is defined by behaviors that directly influence the learning process, such as task persistence and attentiveness to completing academic activities (Finn & Zimmer, 2012). Social engagement reflects behaviors that moderate the influence of academic engagement on achievement, including following rules and appropriately interacting with teachers and other students (Finn & Zimmer, 2012). Thus, in contrast to cognitive engagement and emotional engagement which focus on internal processes that are not readily observable, behavioral engagement focuses on manifest observable aspects of engagement (Finn & Zimmer, 2012).

Using the developmental ecological model (Bronfenbrenner & Morris, 1998), researchers have identified primary school "as the prime developmental period to cultivate student engagement" (Mahatmya, Lohman, Matjasko, & Farb, 2012, p. 54). The developmental model posits that development occurs through interactions between an individual and his/her

environment (Bronfenbrenner & Morris, 1998). In early primary school (kindergarten through third grade) in particular, children spend more time in formal schooling than they ever have previously. Transitioning from primarily experiencing the home environment to spending significant time in an educational setting makes the classroom one of the most influential contexts during this developmental period (Bronfenbrenner & Morris, 1998). The interactions students have in early primary school establish the developmental trajectory for cultivating skills for engaging in the classroom. Thus, early assessment of engagement skills provides critical information for teachers and parents about how they can strategically adapt their interactions with children to further support the development of these essential skills.

To measure and identify engagement difficulties among early primary school students in large, diverse urban school districts, there is a need for assessments that are: (1) developed for a specific purpose and context, (2) scientifically sound, and (3) practical for routine, large-scale use. First, the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999), notes that an essential initial step in creating scientifically sound measures is a priori delineating the explicit purpose and context in which a measure is intended to be used. Recent research has argued that a misfit between the measurement, purpose, and context can impact the accuracy of findings and, more importantly, decision making (LeBoeuf, Fantuzzo, & Lopez, 2010). Second, measures need to be scientifically grounded by evidence demonstrating that they produce information from which valid and reliable inferences can be made. To do so, these assessments must be based on a valid and reliable response process. Two of the most common processes for measuring classroom engagement are student self- and teacher-reports. Self-reported measures are not appropriate in

early primary school as students are still developing the literacy skills and metacognitive knowledge needed to accurately report on their engagement (Fredricks & McColsky, 2012). In contrast, teacher-reports are an effective measurement approach at this age when used to assess observable phenomenon such as the academic and social aspects of behavioral engagement. Third, measures intended for use in large, diverse urban school districts need to be practical for routine use. Such measures provide teachers and parents with regular information on students' classroom engagement so potential problems and interventions can be identified (National Center on Response to Intervention, 2010). Shorter, scientifically based measures embedded within a report card system would be practical for routine use.

There are several teacher-report rating scales of early primary school students' behavioral engagement. Some of these measures are large, multidimensional teacher-report measures such as the Adjustment Scales for Children and Adolescents (ASCA, McDermott, 1993) and the Behavior Assessment System for Children (BASC-2, Reynolds, & Kamphaus, 2004). With 156 items on the ASCA and 139 items on the BASC, these measures are useful for research purposes but are too costly and time consuming for routine, large-scale use. Shorter teacher-report measures of students' behavioral engagement in primary school have also been developed and include the Rochester Assessment Package for Schools (Wellborn & Connell, 1987) and the Student Behavior Checklist (Fincham, Hokoda, & Sanders, 1989). In addition, shorter teacher-report measures of behavioral engagement have been derived from personality assessments (e.g., Effortful and Conduct Engagement Scales, Hughes, Luo, Kwok, & Loyd, 2008), assessments of student adjustment to school (e.g., Teacher Rating Scale of School Adjustment, Birch & Ladd, 1997), and indicators of student participation (e.g., Student Participation Questionnaire, Finn, Folger, & Cox, 1991). However, many of these measures were originally developed with

children from middle-income families, attending school in small to medium suburban or rural districts, and/or from communities with small minority populations. Currently there are no scientifically based measures of behavioral engagement developed specifically for, and successfully incorporated within, a large, diverse urban district's early primary school report card system.

Current research on classroom engagement reveals a need for shorter, scientifically sound teacher-report measures designed specifically for routine use to measure engagement among early primary school students in large, diverse urban school districts. Engagement theory and research indicates that observable aspects of early primary school children's engagement that can be validly and reliably assessed include the academic and social aspects of behavioral engagement. Optimally, the scientific test of validity should be conducted at scale to demonstrate that the measure meets the need for which it was designed–routine use by early primary school teachers within large urban school districts. Thus, the purpose of this study was to provide an initial validation of a teacher-report measure of early primary school students' behavioral engagement that was specifically designed to be used district-wide in a large, diverse urban area.

The Problems in Classroom Engagement Scale (PCES) is currently used by and was developed in partnership with teachers in the Philadelphia School District, the eighth largest district in the nation. The PCES was developed to be used as part of the school district's report card system from first through third grade to assess engagement problems at three reporting periods during the academic year. As part of the report card system, the PCES has the potential to identify student engagement needs not only for teachers but also for parents by sharing this important information at regular parent-teacher conferences. A review of the literature indicated that currently there are no scientifically based measures of early primary school students'

behavioral engagement that have been developed specifically for and successfully incorporated in a large, diverse urban school district's report card system. The primary aim of this study was to provide an initial evaluation of the psychometric properties of this short teacher-reported measure of behavioral engagement while being used at scale across a school district. Specifically, the following research questions were investigated:

1. What is the dimensionality of the PCES? Does this structure reflect the key subdomains of behavioral engagement (i.e., academic and social engagement)?
2. Are scores on the PCES consistent internally and over time?
3. How do the domains of the PCES relate to concurrent assessments of academic achievement and risk to educational progress?

Method

Participants

The present study examined de-identified administrative data from the School District of Philadelphia including enrollment, report card, and academic achievement records. The study cohort included students from 2003-2006 who were enrolled for at least one year in public schools operated by the School District of Philadelphia during first (N = 14,380), second (N = 13,519), and third (N = 13,097) grades. The first-, second-, and third-grade samples were defined as any student within each of the three grades who had complete data for the individual items. Complete data were available for 96.8% of first graders, 95.1% of second graders, and 94.6% of third graders. Table 1 provides a comparison of the demographic characteristics across the three grade-based samples. Approximately 50% of the students were male and over 60% were African-American.

Measures

Problems in Classroom Engagement Scale (PCES). The PCES is a 15-item checklist designed to identify children's problem behaviors during routine classroom situations in primary school settings. The PCES was originally designed as part of a performance assessment battery that examined children's cognitive skills and abilities, motor skills, and behavioral engagement across early primary grades (Fantuzzo, Rouse, McDermott, Sekino, Childs, & Weiss, 2005). To develop the battery and attend to content and response process validity, committees of teacher leaders across grades wrote items assessing observable and mutable skills represented in the educational literature. The subset of items measuring behavioral engagement in the classroom was corroborated by using other pertinent measures specifically developed for use within a large, diverse urban context including the Learning Behaviors Scale (Stott, McDermott, Green, & Francis, 1988) and the Adjustment Scales for Children and Adolescents (McDermott, 1993). After several years of use, the School District decided to remove the cognitive and motor assessments. A committee of experienced primary-grade teachers and researchers reviewed and revised the remaining behavioral engagement items for use as a standalone measure. The revised version of the PCES contained 15 dichotomous response items on which teachers indicate if a child needs improvement or not (scored 1 or 0, respectively). Ratings were recorded three times per academic year during first, second, and third grades.

TerraNova. Children's reading and mathematics achievement in the spring of third grade was used to evaluate the relations between the PCES scores and external variables. Reading and mathematics achievement was assessed using the Complete Battery Plus version of the *TerraNova, Second Edition* (CTB/McGraw, 1997). The TerraNova is a direct child assessment that was designed to measure concepts, processes, and skills in the content areas of Reading and Language Arts, Mathematics, Science, and Social Studies taught to students from Kindergarten

to Grade 12. TerraNova scores were standardized on the norms that were developed from a representative sample of students across the country. Evidence of validity has been established by following the curriculum guidelines and examining intercorrelations between the subscales (CTB/McGaw Hill, 2003).

The Pennsylvania System of School Assessment (PSSA). Children's achievement in the spring of third grade was also assessed by the state standardized assessment. The PSSA was also used to evaluate the relations between the PCES scores and external variables. The PSSA third grade assessment consists of Reading and Mathematics subtests that are used in accordance with mandated federal reporting under the No Child Left Behind Act (NCLB, U.S. Congress, 2001). Reliability and validity of the PSSA scaled scores has been well established (CTB McGraw-Hill, 2006; Thacker, Dickinson, & Koger, 2004), including internal consistency (*r* range .91 to .93) and validity evidence from factor analysis and differential item functioning.

Suspension. School district administrative records indicated children who have ever experienced in- or out-of-school suspensions during third grade. In this study, for students with at least one suspension during third grade, the suspension variable equaled 1, while those without a suspension history received a 0. Suspension was used to evaluate the relations between the PCES scores and an external measure of behavior.

Truancy. In addition to suspension, a measure of truancy was used to evaluate the relations between the PCES scores and another external measure of behavior. School district administrative records included information on daily attendance for every child. This information was used to create an indicator of truancy in third grade. Per district policy, for students with eight or more unexcused absences during third grade, the truancy variable equaled 1, compared to 0 for children with fewer than eight unexcused absences.

Data Analysis

Exploratory factor analysis. Each of the spring samples from first, second, and third grades were randomly partitioned into an exploratory subsample and a confirmatory subsample of equal size. Velicer's Minimum Average Partialing (Velicer, 1976) was used to estimate the appropriate number of factors to retain. The exploratory subsamples from each grade were used to conduct exploratory factor analysis (EFA) using *Mplus* version 7 (Muthen & Muthen, 2012). For all analyses, robust weighted least squares (WLSMV) estimation using tetrachoric correlations and a promax rotation were employed. Each model was evaluated for its ability to produce dimensions that met the following criteria: (a) satisfy Cattell's (1966) scree test; (b) retain 3 or more items with salient loadings, where loadings ≥ .40 are considered salient; (c) yield reasonable internal consistency (> .70) for unit-weighted salient items; (d) make psychological sense in terms of mutually exclusive assignment of items to factors, maximum number of items retained, and compatibility with other research (Fabrigar, Wegener, MacCallum, & Strahan, 1999); and (e) minimize the root mean square error of approximation (RMSEA) and root mean square residual (RMSR) (Schmitt, 2011) while meeting the other criteria.

Confirmatory factor analysis. The confirmatory subsamples from each grade were used to conduct confirmatory factor analyses (CFA) using *Mplus* version 7 (Muthen & Muthen, 2012). For all analyses, robust weighted least squares (WLSMV) estimation using tetrachoric correlations was employed. The fit of all models to the data was evaluated using the following fit indices: comparative fit index (CFI) ≥ .95, the Tucker-Lewis index (TLI) ≥ .95, and the root mean-square error of approximation (RMSEA), where RMSEA < .05 indicates good fit and RMSEA < .08 suggests acceptable fit (Hooper, Coughlan, & Mullen 2008; Hu & Bentler 1999).

In addition, a bifactor model was estimated to determine if the PCES captures a general behavioral engagement factor in addition to more specific group factors uncovered with EFA (see Figure 1). In this model, items load on their respective group factors (also referred to as specific factors) as well as on a general factor, and all factors are uncorrelated. The bifactor model was evaluated using the same fit indices listed above as well as the number and average loading of items on the general and group factors (Gibbons et al., 2007), and the percent of common variance explained by the general and group factors (Reise, 2012). Finally, to further examine the utility of the bifactor model it was compared to the EFA model and a unidimensional model using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC; Muthen & Muthen, 2012).[1] An improvement in fit over the EFA model provided evidence for the co-existence of a general behavioral engagement domain. In contrast, comparing the bifactor to a unidimensional model provided evidence of the group factors' utility above and beyond the general factor.

Reliability. The reliability of the PCES domains was assessed using both Cronbach's Alpha measure of internal consistency and the test-retest reliability coefficient. Cronbach's Alpha was calculated for each grade in the spring. Test-retest reliability was calculated for each grade over a three-month period (from the winter to the spring) to assess temporal stability.

Relations to external variables. Listwise deletion was used to handle missing data for these analyses given the low amount of missing data (3.2% of first graders, 4.9% of second graders, and 5.4% of third graders) (Allison, 2001). PCES factor scores were estimated using the Maximum A Posteriori method implemented for categorical indicators in *Mplus* version 7 (Muthen & Muthen, 2012). To assess the concurrent validity of the PCES, scores from the spring

[1] AIC and BIC are unavailable with the WLSMV estimator, therefore all models were re-estimated using robust maximum likelihood to obtain information criteria indices to use for model fit comparisons.

of third grade were correlated with other academic and behavioral measures from the spring. Pearson correlations were calculated with the nationally standardized TerraNova and the state standardized PSSA reading and mathematics scaled scores. Both measures were used to ensure that the local PSSA scores related to the PCES in a similar manner as a nationally standardized assessment. Point-biserial correlations were estimated between the PCES scores and the suspension and truancy indicators.

Results

Exploratory Factor Analysis

The results of Minimum Average Partialing indicated that two factors should be extracted from all three exploratory subsamples. For all grades, the scree plots had two inflexion points– one at two factors and one at three factors.[2] Based on this, models with one to three factors were assessed using the comprehensive criteria detailed above. The one-factor solution produced large values of the RMSEA and RMSR (<.095) indicating an unacceptable fit. The three-factor model failed to satisfy Cattell's scree test,[3] and yielded a third factor which retained only two items, one of which was an item that double-loaded onto the first factor. In contrast, the two-factor model met all model evaluation criteria across the three exploratory subsamples.[4] The two-factor promax model satisfied Cattell's scree test, retained at least three items per factor, yielded internally consistent factors (> .70), and maximized the items retained with mutually exclusive assignment to factors while producing values of the RMSEA and RMSR that were close to 0. With respect to item retention, 14 of the 15 items loaded saliently on only one factor and were retained. One item, "Makes appropriate transition between activities," exhibited a salient loading

[2] The eigenvalues ranged from 10.73 to 11.06 for the first factor, 1.43 to 1.78 for the second factor, and .44 to .45 for the third factor.
[3] The eigenvalue for the third factor was less than one.
[4] Item communality estimates for the two-factor model ranged from .60 to .90.

on both factors and was removed from the scale. The meaning of "appropriate transition" may have been less clear to the teachers than the content of the others items.

The final factors, component items, factor loadings, and the correlations between factors are presented in Table 2. Seven items loaded saliently on the first factor. These items rated skills such as following rules and appropriately interacting with teachers and other students, which correspond to the research literature on social aspects of behavioral engagement. As a result, this factor was named *Problems in Social Engagement* (example item, "Works and plays cooperatively with others"). The remaining seven items loading saliently on the second factor reflected behaviors that directly influence the learning process, such as persistence and attentiveness to completing academic tasks. Thus, consistent with the research literature on the academic aspects of behavioral engagement, this factor was labeled *Problems in Academic Engagement* (e.g., "Strives for quality work"). The two factors were correlated .68.

Confirmatory Factor Analysis

A confirmatory factor analysis of the 14 remaining items in the scale supported the two-factor model. The fit statistics (Table 3) suggested acceptable model-data fit for each of the three confirmatory subsamples (i.e., first, second, and third grades). The bifactor model positing one general factor (Problems in Behavioral Engagement) measured by all 14 items and two group factors derived from EFA (Problems in Social and Academic Engagement) also fit the data well for each of the three confirmatory samples (Table 3). For all grades, the AIC and BIC (both unadjusted and sample-size adjusted) indicated that the bifactor model fit the data better than the two-factor model. Furthermore, in all grades these indices indicated that the bifactor model fit the data better than the unidimensional model. Across grades, all items loaded saliently on the general factor (average loading = .80), while the item loadings on the group factors indicated a

moderate to high degree of residual association (average loading of .41) consistent with the finding that the bifactor model is an improvement over the unidimensional model. In addition, the general Problems in Behavioral Engagement factor explained 76.5% of the variance on average across grades, while the Problems in Social and Academic Engagement factors collectively explained 23.5% of the common variance on average. Although the general factor explained more of the common variance, the group factors were able to explain nearly a quarter of the common variance above and beyond the general domain.

Reliability. Across all grades, the scale demonstrated high internal consistency for both the general factor (Problems in Behavioral Engagement) and the two group factors (Problems in Social and Academic Engagement). In the spring of third grade, Cronbach's Alpha for Problems in Behavioral Engagement was .92, Problems in Social Engagement was .90, and Problems in Academic Engagement was .88 (with similar results for the spring of first and second grade). Using the third grade cohort data, the PCES demonstrated good temporal stability over three months (from the winter to the spring), with a test-retest correlation of .83 for the general Problems in Behavioral Engagement factor, .80 for Problems in Social Engagement, and .81 for Problems in Academic Engagement. Similar results were found for the same three-month period in the second grade (.83 for Problems in Behavioral Engagement, .80 for Problems in Social Engagement, and .81 for Problems in Academic Engagement) and first grade cohorts (.83 for Problems in Behavioral Engagement, .79 for Problems in Social Engagement, and .82 for Problems in Academic Engagement). Temporal stability from first to third grades was also supported by the model fit indices from each of the three CFAs. Disparities between CFIs $\leq .01$ and RMSEAs $\leq .015$ indicate that the two models do not differ practically (Chen, 2007). The differences in the CFIs between grades ranged from 0 to .001 and disparities in RMSEAs ranged

from 0 to .002, supporting the stability of the bifactor structure across first, second, and third grades.

Relations to External Variables

Table 4 provides the correlations between the PCES scores and TerraNova standardized scores in reading and mathematics, PSSA scaled scores in reading and mathematics, suspension, and truancy, all measured in the spring of third grade. All correlations were significant ($p < .001$) and in the expected direction.[5] Applying Cohen, Cohen, West, and Aiken's (2003) cutoffs for small, moderate, and large correlations, Problems in Academic Engagement scores had small to moderate correlations with TerraNova and PSSA reading and mathematics scores (ranging from -.23 to -.26), a small correlation with truancy ($r = .15$), and a near-zero correlation with suspension ($r = .03$). In contrast, scores on Problems in Social Engagement had a small to moderate correlation with suspension ($r = .22$) and near-zero correlations with TerraNova and PSSA reading and mathematics scores (ranging from -.03 to -.06) and truancy ($r = .01$). As expected, scores on the general factor (Problems in Behavioral Engagement) had small to moderate correlations with all outcomes. Specifically, Problems in Behavioral Engagement scores had moderate correlations with TerraNova and PSSA scores (ranging from -.26 to -.33), a moderate correlation with suspension ($r = .38$), and a small correlation with truancy ($r = .15$).

Discussion

The research literature has established clear links between student engagement and educational well-being (Fredricks, Blumenfeld, & Paris, 2004). The developmental ecological model (Bronfenbrenner & Morris, 1998) identifies early primary school as a critical developmental period to foster student engagement (Mahatmya, Lohman, Matjasko, & Farb,

[5] Given that tests of statistical significance are influenced by large sample size, effect size interpretations are reported.

2012). These connections signal the need for scientifically based assessments of early primary school students' classroom engagement that are practical for routine large-scale use. This need is particularly salient for students in large, diverse urban school districts who are at higher than average risk for academic failure (U.S. Department of Education, 2009). The purpose of the present study was to investigate the scientific integrity of a short, teacher-report measure to meet this need—the PCES. The PCES was designed and tested specifically for district-wide use as a component of the student report card system for early primary grades in one of the largest urban public school districts in the country.

Factor analytic examination of the PCES revealed a robust bifactor model reflecting one general factor (Problems in Behavioral Engagement) measured by all 14 items and two group factors (Problems in Social Engagement and Problems in Academic Engagement). These findings are consistent with engagement theory, which operationalizes behavioral engagement to include academic and social engagement (Finn & Zimmer, 2012). Also in accordance with this literature, the Problems in Social Engagement dimension focuses on a student's ability to work and play well with others, respect the rights of others, follow teacher's rules and directions, and manage conflict appropriately. In contrast, the Problems in Academic Engagement dimension addresses a student's persistence and attention to school work, ability to complete academic tasks independently, and a student's display of motivation to learn and to complete quality work. These dimensions were found to be internally and temporally consistent across first, second, and third grades, providing support for the use of the PCES for guiding classroom instruction and for progress monitoring within and across grades.

An examination of the relations among PCES scores and external variables in third grade revealed that the two specific behavioral engagement factors had different patterns of

correlations with academic and behavioral outcomes, while the general factor was linked to all of the outcomes. The Problems in Academic Engagement factor was most strongly correlated with academic achievement (TerraNova and PSSA scores) and truancy and exhibited a near-zero correlation with suspension. Truancy is more likely to affect academic engagement in young children, as frequent absences would limit a child's exposure to academic material but would not necessarily diminish his/her ability to socialize with peers and adults. On the other hand, Problems in Social Engagement was most strongly correlated with suspensions, a behavioral outcome most likely linked to children's ability to manage social interactions effectively in school. This factor exhibited near-zero correlations with the achievement outcome measures and truancy, which are more closely related to academic engagement.

The components of behavioral engagement captured by the PCES correspond to national school readiness dimensions and state standards for early childhood education. The U.S. Department of Education's National Education Goals Panel (1995) identified five major dimensions of school readiness which included Social and Emotional Development and Approaches to Learning. Social and Emotional Development represents a child's ability to relate to others and adjust to the social expectations of school. Approaches to Learning denotes students' competence motivation, attention control, and persistence on tasks. These two dimensions are also prominent in the national Head Start Performance Standards and state-level standards for early childhood education (Head Start Bureau, 2004; Scott-Little, Kagan, & Frelow, 2005). The dimensions of the PCES correspond to these essential elements of school readiness and to national studies on situational-based dimensions of problems in classroom engagement (Fantuzzo et al., 2005).

Although the present study provided validity evidence for the PCES capturing theoretically and practically relevant dimensions of classroom engagement, it has some limitations that provide opportunities for future research. First, this research tested the PCES in one large urban school district. While the findings are encouraging, the generalizability of the PCES across other large urban school districts should be assessed. Second, the testing of this measure was limited to first, second, and third grades. Future research should examine the full capacity of the PCES by extending it both downward to kindergarten and prekindergarten as well as upward to fourth and fifth grades. If these dimensions are valid from prekindergarten through fifth grade, it would significantly increase the potential value of this tool for longitudinal assessment to inform intervention. Even if the PCES required some adaptation at the lower or upper levels to be more developmentally appropriate, future research could use IRT techniques of vertical equating (du Toit, 2003) to create a system for continuous growth assessment. Third, assessment of the relations between PCES scores and external variables was limited to those in school district administrative records. A comparison of the PCES with new measures designed to assess early primary school children's classroom engagement in large urban school districts would provide information on how the PCES comports with other classroom engagement tools. Additional benefit would be gained if these measures used independent observers thereby providing a perspective beyond the teacher. Finally, this study employed traditional, single-level factor analytic methods to analyze nested data. This approach has been recommended as good practice prior to conducting multilevel analyses, especially in the case of categorical data (Grilli & Rampichini, 2007). However, future research should employ multilevel categorical factor analytic approaches to further assess the robustness of the current model.

The PCES is focused on behavioral engagement in the classroom and is validated for use with first- through third-grade students, district-wide, three times per academic year. As such, it has the potential to be developed and tested to serve as part of a comprehensive early identification and intervention system that integrates measures like the PCES to foster dialogue about student needs and bring extra support and attention to their classroom adaptation. One such system endorsed by many national organizations is Response to Intervention (RTI; National Center on Response to Intervention, 2010). Central to RTI are scientifically sound, system-wide screening instruments of academics and student behavior that initiate a dynamic, multi-tiered system of assessment and intervention for children who demonstrate a need for more intensive interventions. As a practical short measure that can be easily used by teachers, the PCES may be suited for incorporation into the first tier of this process. Of course, this would require more research and testing to determine its ability to promote beneficial early classroom interventions to support student needs and teacher instruction.

Another important potential benefit of the PCES is how it could be used to foster family involvement. The NCLB Act has substantial requirements for schools, particularly those receiving Title I funds, to engage parents (NCLB Action Briefs, 2004). As stated in the NCLB Parental Involvement Action Brief, "every school district and every school receiving Title I dollars must have a written parent involvement policy, as well as a built school capacity to effectively implement the parent policy provisions" (NCLB Action Briefs, 2004, p. 1). The built capacity of the PCES can be used by school districts to provide scientifically valid data to parents in parent-teacher conferences. Teachers could use the PCES portion of the report card to alert parents if their child is having difficulties engaging socially and academically in the classroom and share what they are doing in the classroom to foster engagement. Moreover, as

part of the parent involvement plan, school administrators could generate professional development opportunities to help teachers engage family members in teaching these critical social and academic engagement skills. This professional development could also expand teachers' repertoire of classroom intervention strategies to support engagement.

In sum, the PCES provides a scientifically based means of fostering district-wide focus on students' early behavioral, social, and academic engagement. This information has the potential to be used by responsible educators as part of their effort to promote child, family, and school district engagement in the early primary grades. The PCES could provide a small but important contribution to a larger data-based dialogue among teachers, professional support staff, and parents to enhance educational trajectories for students at higher than average risk for educational disengagement and future academic failure.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317-322.

Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage Publication.

American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME]. (2014). *Standards for educational and psychological testing*. Washington, DC: Author.

Bentler, P. M. (2001). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.

Birch, S., & Ladd, G. (1997). The teacher–child relationship and children's early school adjustment. *Journal of School Psychology, 35,* 61–79.

Boyle, G. (1985). A reanalysis of the higher-order factor structure of the motivation analysis test and the eight state questionnaire. *Personality and Individual Differences, 6*(3), 367-374.

Bridgeland, J.M., DiIulio, J.J., & Morison, K.B. (2006). *The silent epidemic: perspectives of high school dropouts*. Civic Enterprises, LLC.

Bronfenbrenner, U., & Morris , P. A.(1998). The ecology of developmental processes. In W. Damon (Series Ed.) & R. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theory* (5th ed., pp. 993-1029=8). New York: John Wiley.

Byrne, B.M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaun & Associates.

Cattell, R.B. (1966). The Scree Test for the Number of Factors. *Multivariate Behavioral Research, 1*(2), 245-276.

Chang, H, N. & Romero, M. (2008*). Present, engaged, and accounted for: The critical importance of addressing chronic absence in the early grades*. Retrieved from http://www.ncpie.org/pubs/PresentEngagedAccountedFor_Sep2008.pdf.

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

Christenson, S., Reschly A., & Wylie, C., (2012). *Handbook of research on student engagement*, (Ed). NY, NY.: Springer Publishing Co.

Christenson, S. L., & Thurlow, M. L. (2004). School dropouts: Prevention considerations, interventions, and challenges. *Current Directions in Psychological Science, 13,* 36 –39. doi:10.1111/j.0963-7214.2004.01301010.x.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (3rd ed.). Mahwah, NJ: Erlbaum.

CTB/McGraw-Hill. (1997). *TerraNova, the Second Edition.* [Assessment]. Monterey, CA: Author.

CTB/McGraw-Hill. (2003). *TerraNova, the Second Edition, CAT Technical Report.* Monterey, CA: Author.

CTB/McGraw-Hill. (2006). Pennsylvania Grade 3 Assessment Mathematics and Reading. *Technical Report, Spring 2006 Operational Test*.

Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom

interactions. *Early Childhood Research Quarterly, 25*, 1-16. doi:10.1016/j.ecresq.2009.08.004

du Toit, M. (Ed.). (2003). *IRT from SSI, BILOG-MG, MULTILOG, PARSCALE, TESTFACT.* Lincolnwood, IL: Scientific Software International.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272-299.

Fantuzzo, J.W., Rouse, H.L., McDermott, P.A., Sekino, Y., Childs, S. & Weiss, A. (2005). Early childhood experiences and kindergarten success: A population-based study of a large urban setting. *School Psychology Review, 34,* 571–588.

Fincham, E D., Hokoda, A., & Sanders, R., Jr. (1989). Learned helplessness, test anxiety, and academic achievement: A longitudinal analysis. *Child Development, 60*, 138-145.

Finn, J. D., Folger, J., & Cox, D. (1991). Measuring participation among elementary grade students. *Educational and Psychological Measurement, 51,* 393-402.

Finn, J. D., & Zimmer, K. S. (2012). Student engagement: What is it? Why does it matter? In S. L. Christenson, A. L. Reschly & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 97-132). USA: Springer

Fredricks, J., & McColskey, W. (2012). The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments. In S. L. Christenson, A. L. Reschly & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 763-782). USA: Springer

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

Gibbons, R. D., Bock, R., Hedeker, D., Weiss, D., Segawa, E., Bhaumik, D., Kupfer, D., Frank, E., Grochocinski, V., & Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement, 31*, 4-19.

Gorsuch, R. L. (2003). Factor analysis. In J. A Schinka & W. F. Velicer (Eds.), *Handbook of psychology*: *Research methods in psychology* (Vol. 2, pp. 143–164). NY: Wiley & Sons.

Grilli, L., & Rampichini, C. (2007) Multilevel Factor Models for Ordinal Variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 1-25. doi:10.1080/10705510709336734

Head Start Bureau. (2004). *Program performance standards for the operation of Head Start Programs by grantees and delegate agencies*. Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services. Retrieved from http://www.acf.hhs.gov/programs/hsb/performance/1304A.html.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Articles*, 2.

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

Hughes, J. N., Luo, W., Kwok, O.-M., & Loyd, L. K. (2008). Teacher–Student Support, Effortful Engagement, and Achievement: A 3-Year Longitudinal Study. *Journal of Educational Psychology, 100*(1), 1–14. doi:10.1037/0022-0663.100.1.1

LeBoeuf, W. A., Fantuzzo, J. W., & Lopez, M. L. (2010). Measurement and Population Miss-Fits: A Case Study on the Importance of Using Appropriate Measures to Evaluate Early

Childhood Interventions. *Applied Developmental Science, 14*(1), 45-53. DOI: 10.1080/10888690903510349

Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., Kastberg, D., & Jocelyn, L. (2004). *International Outcomes of Learning in Mathematics Literacy and Problem Solving: PISA 2003 Results From the U.S. Perspective* (NCES 2005-003). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Li-Grining, C. P., Votruba-Drzal, E., Maldonado-Carrefio, C., & Haas, K. (2010). Children's early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology, 46*, 1062-1077.

Mahatmya, M., Lohman, B. J., Matjasko, J. L., & Farb, A. F. (2012). Engagement across developmental periods. In S. L.Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 45-63). New York: Springer.

Matthews, J. S., & Kizzie, K. T. (2010). African Americans and boys: Understanding the literacy gap, tracing academic trajectories, and evaluating the role of learning-related skills. *Journal of Educational Psychology*, *102*, 757-771.

McClelland, M. M., Acock A. C., Piccinin A., Rhea S. A., & Stallings M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early Childhood Research Quarterly*, *28*(2), 314-324.

McDermott, P. A. (1993). National standardization of uniform multisituational measures of child and adolescent behavior pathology. *Psychological Assessment, 5*(4), 413-424. DOI: 10.1037/1040-3590.5.4.413

Muthen, L. K., & Muthen, B. O. (2012). *Mplus Version 7: Statistical analysis with latent variables* [Computer program]. Los Angeles: Muthen & Muthen.

National Center on Response to Intervention. (2010). Essential components of RTI—A closer look at response to intervention. Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Center on Response to Intervention. Retrieved from http://www.rti4success.org/pdf/rtiessentialcomponents_042710.pdf.

National Education Goals Panel. (1995). *The National Education Goals Report*. Washington, DC: Author.

No Child Left Behind Action Brief, (2004). Parental involvement Title 1. Retrieved from http:/www.ncpie.org/nclbaction/parent_involvement.html.

Pianta, R. C., Cox, M. J., & Snow, K. L. (Eds.). (2007). *School readiness and the transition to kindergarten in the era of accountability*. Baltimore, MD: Paul H. Brookes Publishing Co.

Reise, S. P. (2012). Invited Paper: The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research, 47*(5), 667–696. doi:10.1080/00273171.2012.715555

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd ed.). Bloomington, MN: Pearson Assessments.

Sattler, J. & Hoge, R., (2006). *Assessment of children: Behavioral, social, and clinical foundations.* La Mesa, CA: Jerome M. Sattler, Inc.

Scott-Little, C., Kagan, S., & Frelow, V. (2005). *Inside the content: The breadth and depth of early learning standards*. Greensboro, NC: Regional Educational Laboratory Southeast.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment, 29*, 304-321.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464.

Smith, G.T., McCarthy, D.M., & Anderson, K.G. (2000). On the sins of short-form development. *Psychological Assessment, 12,* 102-111.

Stott, D. H., McDermott, P. A., Green, L. F., & Francis, J. M. (1988). *Learning Behaviors Scale and Study of Children's Learning Behaviors Research Edition manual*. San Antonio, TX: The Psychological Corporation.

Thacker, A. A., Dickenson, E.R., & Koger, M. E. (2004). *Relationships among the Pennsylvania System of School Assessment (PSSA) and other commonly administered assessments* (HumRRO Report No. FR-04-33). Alexandria, VA: Human Resources Research Organization.

U.S. Congress. (2001). Public Law 107-110: No Child Left Behind Act of 2001. Retrieved from http://www.ed.gov/policy/elsec/leg/esea02/index.html.

U.S. Department of Education (2005). *Common Core of Data*. Washington, DC: National Center for Education Statistics.

U.S. Department of Education. (2009) *2007-2008 Schools and Staffing Survey (SASS*). Washington, DC: National Center for Education Statistics.

U. S. Department of Education. (2009). *Achievement gaps: How Black and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress: Statistical Analysis Report.* NCES 2009-445: Washington, DC: National Center for Education Statistics.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. Psychometrika, 41, 3, 321-32

Vigdor, J., & Ludwig, J. (2007). Segregation and the Black-White test score gap. *Paper 12988.* Retrieved from http://www.nber.org/papers/w12988.

Wellborn, J. G., & Connell, J. P. (1987*). Manual for the Rochester Assessment Package for Schools*. Rochester, NY: University of Rochester.

Wood, R., Wilson, D. T., Muraki, E., Schilling, S. G., Gibbons, R., & Bock, R. D. (2003). TESTFACT 4: Test scoring, item statistics, and item factor analysis. *Scientific Software International Inc., Lincolnwood, IL. Version, 4*.

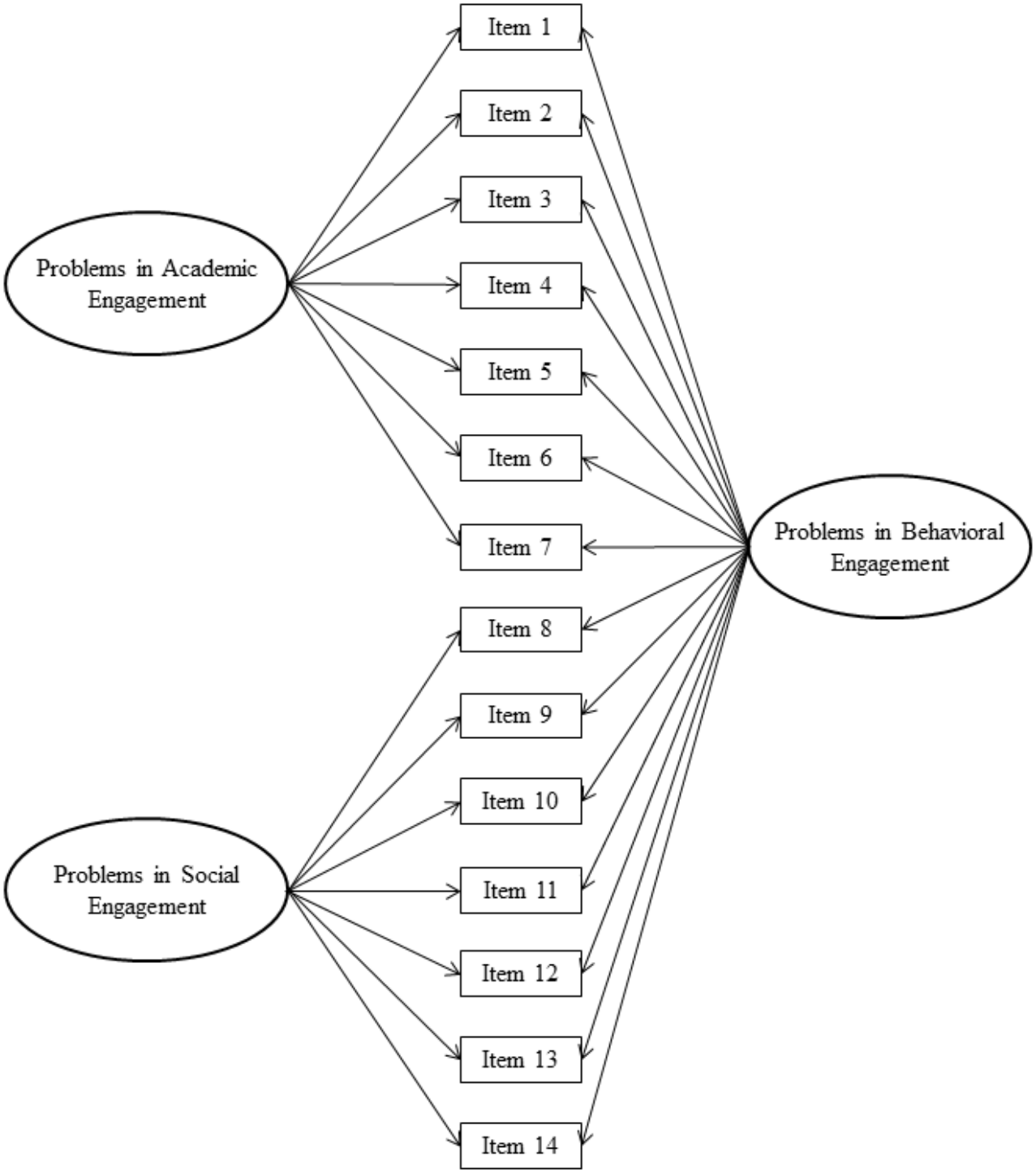

Figure 1. Confirmatory bifactor model positing one general factor of Problems in Behavioral Engagement and two group factors of Problems in Social Engagement and Problems in Academic Engagement. General and group factors are uncorrelated.

Table 1

Demographics of Enrolled Students Enrolled with Complete Data on the PCES

	Respondents with complete data		
	1st Grade (*N* = 13,923)	2nd Grade (*N* = 12,860)	3rd Grade (*N* = 12,390)
Sex (male)	51.6	51.1	50.8
African American	62.2	61.5	61.2
White	13.9	13.3	13.2
Latino	17.8	18.5	18.7
American Indian	0.1	0.2	0.1
Asian	5.1	5.5	5.6
Other	0.9	1.1	1.2

Note: Numbers represent the percentages of children within each demographic characteristic.

Table 2

Factor Loadings for Exploratory Factor Analysis with Promax Rotation of the PCES Scale, Spring of Third Grade Exploratory Subsample (n=6,245)

Scale	Social	Academic
Handles conflict appropriately	**0.99**	-0.06
Respects rights, diversity, feelings and property of others	**0.96**	-0.03
Works and plays cooperatively with others	**0.91**	0.02
Accepts responsibility for choices and actions	**0.83**	0.12
Follows school and classroom rules	**0.82**	0.15
Respects school environment and materials	**0.72**	0.25
Listens and Follows directions	**0.60**	0.39
Completes work on time	-0.11	**1.00**
Strives for quality work	0.03	**0.90**
Can work independently	-0.03	**0.90**
Demonstrates consistent effort	0.09	**0.88**
Shows positive attitudes toward learning	0.29	**0.70**
Organizes self, materials, and belongings	0.16	**0.69**
Participates in group activities	0.23	**0.63**

Note: Factor loadings > .40 are in boldface. Promax factor correlation: $r_{12} = .68$. Social = Problems in Social Engagement; Academic = Problems in Academic Engagement. Teachers were asked to identify problems in each item, so the underlying factors also represent the problems in engagements.

Table 3

Confirmatory Factor Analysis (CFA) Model Fit Statistics

Grade	Structure	*CFI*	*TLI*	*RMSEA*	90% *CI RMSEA*
1st grade	Two-factors	0.985	0.982	0.067	(.064, .069)
	Bifactor	0.995	0.993	0.041	(.039, .044)
	One-factor	0.961	0.954	0.105	(.102, .107)
2nd grade	Two-factors	0.983	0.979	0.066	(.064, .069)
	Bifactor	0.994	0.991	0.043	(.040, .045)
	One-factor	0.956	0.948	0.105	(.102, .107)
3rd grade	Two-factors	0.988	0.986	0.059	(.056, .061)
	Bifactor	0.995	0.993	0.043	(.040, .045)
	One-factor	0.966	0.960	0.099	(.096, .101)

Note: *CFI*: comparative fit index;
TLI: Tucker-Lewis index;
RMSEA: root mean-square error of approximation;
90% *CI RMSEA*: 90% confidence interval for RMSEA.

Table 4

Correlations between Engagement Scores and Academic and Behavior Outcomes in the Spring of Third Grade

Factor	TN reading	TN math	PSSA reading	PSSA math	Suspension	Truancy
General Behavioral Engagement	-0.26	-0.26	-0.33	-0.29	0.38	0.15
Social Engagement	-0.05	-0.03	-0.06	-0.05	0.22	0.01
Academic Engagement	-0.23	-0.23	-0.26	-0.26	0.03	0.15

Note. N = 11,222. TN = TerraNova. All correlations are significant at the 0.001 level.