Refinement for Scale Utility: The Physical Education Classroom Instrument, Short Version, PECI-S

Paul Rock Krech
and
Pamela Hodges Kulinna
Arizona State University

Donetta Cothran
Indiana University

Contact Information:
Paul Rock Krech

College of Education
Psychology in Education
Arizona State University
BOX 870611
Tempe, AZ 85287-0611

ABSTRACT

The study was designed to buttress the findings of an existing parent instrumen designed to describe the spectrum of student behaviors impacting classroom management. This study had multiple phases, including exploratory factor analyses to identify salient subscales and items, first and second-order confirmatory factor analyses, a multiple group confirmatory factor analysis, and an assessment of internal consistency of the shortened instrument. Furthermore, the relationship of original and newly identified subscales was assessed. Participants were 2275 junior high school and high school students, as well as 300 teachers in physical of education classes. The following 5 factors (i.e. subscales) emerged from the exploratory factor analysis: (I) aggressive, (II) low engagement, (III) hyperactive, (IV) harmful, and (V) defiant. Across-group subscale alpha reliability coefficients averaged around .71, while the average alpha for the three group total scale scores was .92. The rotated factor loadings for the items ranged from .54 to .83. Standardized confirmatory factor analytic loadings ranged from .16 to .96. The shortening process produced a 20-item version of the Physical Education Classroom Instrument, PECI. We maintain that this is an instrument appropriate for use with students and teachers.

Key words: Short form, multiple group confirmatory factor analysis, second order factor analysis

Refinement for Scale Utility: The Physical Education Classroom Instrument, Short Version, PECI-S

INTRODUCTION

An important component of classroom management in education is the ability of all those involved to accurately describe student misbehavior. It was posited that although different stakeholder groups are certain to have differing viewpoints, there is also a similarity in what people believe about actions that constitute disruptive behaviors. The risk of a wide range of responses leading to between-group variance or inequality of performance across groups; however, might be expected surveying different respondent groups. This phenomenon can lead to the compromise of the validity of a shortened instrument. Likewise, simplifying or shortening any scale that attempts to model a latent psychological variable may seriously impact estimates of reliability and validity of the scores produced by the instrument in various populations. In the present case, this hazard is activated by the use of three quite diverse groups of respondents: Younger students, older students, and teachers. Each group in practice appears to have similarities and differences regarding individual behaviors and interpretation of their meanings.

The initial Physical Education Classroom Instrument (PECI) examining student disruptive behaviors was developed by Kulinna, Cothran, and Regualos (2003). The earlier study was the primary effort which included instrument development, tests of content-related evidence, an exploratory factor analysis to identify factors, and internal consistency reliability analyses. The initial study produced a 59-item, six subscale instrument with reliability (Cronbach's alpha) coefficients of .91, .88, .89, .74, .72, and .53 for scales I through VI, in that order. Issues of item redundancy, scale reliability, and different validity estimates have been thoroughly addressed and rectified in the first PECI study.

Defining student disruptive behavior

Classroom management is an important factor of teacher effectiveness chiefly because student disruptive behaviors negatively impact effectiveness due to time spent correcting those behaviors. Therefore, student disruptive behavior can be defined as those behaviors that range from requiring teachers to spend instruction time managing, influencing other students to defocus from learning, to posing a threat to classroom and school safety. Teachers' reports of the most frequent and troublesome behaviors in general are related to aggressive, harmful, and defiant behavior. Teachers also spend time correcting behaviors relating to low student engagement (i.e. time not on-task), distraction of other learners, and student hyperactivity. Although most disruptive behaviors are generally minor, these behaviors still take time from teachers total instruction time.

Previous scale development

To study student behavior, previous work by Kulinna et al. (2003) focused on teachers' perceptions of disruptive actions occurring in classes. This was brought about by the lack of and available instrument to examine teachers and students' reports of the type and frequency of student behaviors that might disrupt physical education classes. It was also hypothesized that students and teachers might have different views of the importance of types of disruptive behavior. Therefore, the parent scale, The PECI was developed and validated across three groups of respondents; those being Junior high school and high school students, as well as teachers.

Purpose

The study was designed to as a formative adjunct in the development of a short form to compliment the PECI. The original PECI was designed to describe various student behaviors impacting classroom management. Usually, shortened versions of parent forms are developed

some time after the parent survey is validated, and rarely at the same time (Smith, McCarthy, & Anderson, 2000). The chief purpose of study was to formatively simplify the PECI to an optimal length for administrative utility. The goals of shortening the scale are multifaceted. The first objective was to increase the economy of instrument administration. The short-form will require approximately half the time to administer and score. The second objective was to increase the ability compare different groups of respondents (i.e. teachers, junior high school, and high school students) to determine the existence of similar views about classroom management across populations. By increasing the scale comparability of groups, we narrow subscale constructs regarding specific student in-school behaviors, while maintaining validity of the overall scale. A third objective was to produce a more concise instrument within a parsimonious model. We are aware that by narrowing the content of the subscales, we risk changing the substantive nature of the overall construct. Caution must be taken not to reduce the number of factors and items so drastically that it leads to construct under-representation that may obfuscate the intended purpose and validity of the measure (Messick, 1995).

METHODS

Participants

Of a total of 2612 respondents sampled, 34 cases (1.3%) yielded incomplete or missing data for a total of 2578 analyzable responses. Initially, all 59 scale items were analyzed by combining the three different groups of respondents. Group one (n = 1024, 39.2%) was composed of junior high school students, group two (n = 1251, 47.9%) and was composed of high school students, and group three consisted of 303 (11.8%) teachers. Students and parents provided informed consent and the disruptive behavior instrument was completed during physical education classes. There was similar representation from both sexes (Males = 1145,

50.6%; Females = 1304, 49.9%). Students had various ethnic backgrounds including European American/Caucasian (*n* = 1527, 59.3%), African American (*n* = 810, 31.5%), Hispanic (*n* = 88, 3.4%), Asian Indian (*n* = 29, 1.1%), Asian American (*n* = 16, 0.6%), and multiracial or other (*n* = 105, 4.1%).

Instrumentation

The parent PECI is a 59-item five-point scale using Likert-type indicator scale format. Responses range from 1 (*never*) to 5 (*always*) for each of the items. For instance, the item of "smart mouth toward teacher" would be rated by students and teachers on how often this happens in their classroom from the indicators 1 (*never*), 2 (*rarely*), 3 (*sometimes*), 4 (*frequently*), or 5 (*always*). Participants took approximately 20 minutes to complete the scale. Teachers completed the instrument outside of class time. The parent instrument has previously shown that it can produce reliable and valid data with students (Kulinna, Cothran, & Regualos, 2003) and teachers (Kulinna, Cothran, & Regualos, 2006).

Procedure

Each respondent completed one PECI survey. The survey data was entered into a database by a research assistant. A combination of exploratory (EFA) factor analysis, confirmatory factor analysis (CFA), reliability (coefficient alpha), and correlational analyses were utilized.

Analyses

Exploratory factor analytic procedures.

An EFA and parallel analysis (PA) were performed with the number of factors to be interpreted set to five based on previous finding with the PECI (Kulinna, et al. 2003; 2006). Both the EFA and PA suggested a five factor solution to be appropriate for analysis. The PA is

used as an adjunct with EFA for comparing eigenvalues obtained from observed data to eigenvalues obtained from random data. The PA then retains factors/components with observed data eigenvalues greater than random data eigenvalues (O'Connor, 2000). The method of factor extraction was set by specifying principal axis factoring (PAF) with promax (oblique) rotation. PAF is a factor extraction method which seeks the least number of factors which can account for the common variance (correlation) of a set of variables. PAF is preferred for purposes of latent variable analysis or structural equation modeling (SEM). Widaman (1993) notes, "principal component analysis should not be used if a researcher wishes to obtain parameters reflecting latent constructs or factors."

The EFA was performed using the aggregated three-group (junior high, high school, and teacher) response data (N = 2578). Items were retained that both held the highest factor loadings across all three groups and made substantive sense in assessing the five factors (see Table 1). These items were further tested within a CFA framework both by individual groups, and multiple group analyses. Basic statistics including mean, standard deviation, skew, and kurtosis were computed for all three samples (see Table 2).

Confirmatory factor analytic procedures. A series of CFAs were conducted to assess the fit of the proposed five factor measurement model of the PECI-S, to consider alternative models, and to assess across group scale invariance by comparing group fits. By fit, we mean the ability of the model specified (i.e. variance and covariance parameters placed on the model) to reproduce the sample variance-covariance data matrix. By "good fit" we mean a low chi-square statistic relative to degrees of freedom, a high confirmatory fit index ($CFI \geq .90$), high Tucker-Lewis Index (TLI $\geq .90$) and a low root mean squared error of approximation ($RMSEA \leq .05$) statistic (Hu and Bentler, 1995; Fan and Sivo, 2005). Due to non-multivariate normality of the

data, robust methods were utilized to conduct the CFAs. The robust methods assist in model assessment by adjusting the obtained model fit chi-square statistic based on the degree of non-normality in the sample data. The statistic that is reported by robust analysis is the Satorra-Bentler chi-square (i.e., S-B χ^2).

Baseline model. A baseline measurement model was first established so each subsequent model could be compared to it. In this case, the 20 item scale was used with five correlated factors, utilizing combined responses from all three groups (N = 2578).

Single group models. Next, separate datasets for each group, junior high students (n = 1028), high school students (n =1251), and teachers (n = 303), were developed from the aggregate data set. Each of these groups was run as a 20-item five correlated factor measurement model.

Multiple groups model. The Multiple Groups CFA (MGCFA) was used to evaluate a popular hypothesis about multiple populations. The MGCFA is different than the baseline (i.e. "merged") model in that all three groups are combined in the baseline model and treated as an aggregate dataset. The MGCFA is a multiple sample test testing CFA model fit individually and comparatively for each set of respondent data.

Tests of Invariance

Conducting tests of invariance is the next step in scale evaluation administered to multiple samples. These tests assess whether or not there is evidence of invariance among key measurement parameters across populations (Jöreskog, 1971). In this particular case, we think that of the items chosen to represent each subscale, the fit of each model for each of the three subsets of respondents should not vary significantly. This hypothesis is rather restrictive and implies not only that all covariance matrices are equal, but also that the linear structural model

that generates these matrices is identical in all respects across populations. It is unlikely to be strictly true except under rigorous circumstances. In this case, models may be essentially equivalent, but a few parameters may differ trivially across samples. A chi square difference test may be applied to assess whether the model differences are significant. If the differences are not significant, the researcher concludes that the single group models are the same as the multigroup model, leading to the conclusion that the model applies across groups.

*Model 1.*The first model allowed the factor loadings, factor correlations, and the error variances to vary or to be freely estimated across samples. This hypothesis tests that the single-factor structure or unidimensionality imposed among 20 PECI items of student disruptive behaviors across groups is equivalent. For example, the null hypothesis for each test would be "no invariance across compared groups." This tests for congeneric (basic) measurement properties across groups (junior high school students, high school students, and teachers), that is, each measure is associated with only one latent construct and that all covariation between measures is a consequence of the relationships among measures and constructs. In other words, this hypothesis tests whether the set of PECI-S items represents the same conceptual framework across respondents.

*Model 2.*The second model held the factor correlations invariant (i.e. constrained to be equal across groups), but allowed the error variances and factor loadings to be different for each sample. This hypothesis tests that the regression slopes, that is, factor loadings, for like-items relating to disruptive behaviors, the latent variables, are equal across the groups. This is essentially a test of equality of scaling units across groups, which tests that the expected change on each item given a per-unit change in the latent variable is the same across all groups of respondents.

*Model 3.*The third model (i.e. scalar invariance) allowed the factor correlations and error variances to vary but the factor loadings remained constrained to be equal across groups. This hypothesis tests that the vector of intercepts for like items is invariant across groups, which is a test of systematic response bias across groups. This is particularly important when substantive reasons indicate that one group might have a different mean score on the instrument. For example, if we believe that respondent differences in rating certain disruptive behavior items (i.e., aggression) are expected, the scalar invariance test is substantively meaningful when we detect group differences in such items.

Model 4. The final, and most restrictive model, required the factor loadings, factor correlations, and error variances for both samples to be constrained to be equal across groups. This hypothesis tests that measurement errors (i.e. reliabilities) for like-items are equivalent across groups. This model also tests for equivalent factor variances for the common factor across groups. This determines whether the groups used equivalent ranges of the construct continuum to respond to the indicators reflecting the construct of disruptive behavior.

Because each hypothesis is a more rigorous test of invariance, each less lenient model is compared with the next more highly constrained model (e.g., Model 1 is compared with Model 2, Model 2 is compared with Model 3, etc.) to determine the level at which measurement equivalence no longer exists (Jöreskog, 1971).

Internal Consistency Estimate. In an effort to determine dimensionality of the PECI-S, a second order model was fit to the data. A second order CFA model is one whereby the measurement model, that is the 20 PECI-S items, load on their respective five factors as specified. Then each of those factors loads on an overall or higher order factor. This process was conducted individually for each sample as well as for a merged sample of respondents (*N* =

2582). Both adjusted generalized least squares (AGLS) utilizing polychoric correlations for categorical (ordered or Likert: Poon & Lee, 1987) data and robust ML estimates using Pearson Product Moment (PPM) correlations were calculated to assess model fit.

Next, a reliability analysis using a comparative series of coefficient alphas (Cronbach, 1951) was conducted. Coefficient alpha can be conceptualized as the proportion of a scale's total variance attributed to a common source. Generally this would be the true score of each latent variable underlying its concomitant items. Alpha is calculated for each subscale for each the three groups of respondents. Finally, PPM correlations were calculated among PECI-S subscales as well as between PECI and PECI-S subscales to for assess factor overlap.

Results

The five-factor, 20-item solution seemed to provide the best interpretation of salient item and factor combinations for the shortened scale. This solution accounted for approximately 49% of the total variance among the original 59 items (see Table 1). Moment statistics (i.e. mean, standard deviation, skew, and kurtosis) were calculated for the 20 retained items (see Table 2). The majority of these estimates were within normal limits. Exceptions to these limits are discussed below.

Factor one is composed of items 26, 36, 37, 46, 50, and 52, with terms such as "Playing too rough and risking injury" and "Quarrelsome (Fusses with others)." This factor was named "aggressive," and identical to the original PECI factor name.

Factor two is composed of items 5, 40, and 45, contained terms such as "Gang symbols displayed" and "Drug use." Factor two was named "harmful," a shortened version of the original factor named "harmful/illegal" on the parent PECI.

Factor three is composed of items 28, 29, 32, and 39, for example “acts shy and withdrawn,” and “continually saying, "I can't do it.” We named this factor “low engagement,” again, a shorter name than the original factor name “low engagement/irresponsibility” on the parent PECI.

Factor four consists of items 6, 10, 14, and 17, with items such as “Can't sit still (hyperactive),” and “Doesn't pay attention.” We named this factor “hyperactive.” This is a new factor name for a subscale independent of the original PECI.

Finally, factor five includes items 51, 56, and 59, such as “Smart mouth toward students,” and “Moves slowly on purpose.” This factor was named “defiant.” Like factor four, this is a novel factor name for a subscale independent of the original PECI.

CFAs for groups of respondents (measurement models)

The CFA analyses were run using EQS software (Bentler & Wu, 2005). The baseline model, with 5 correlated factors and 20 items utilizing data from all three (merged) groups of respondents (junior high and high school students, and teachers) fit quite well. These analyses were conducted using identical 20-item five correlated factor models The fit of the first two groups (junior high school and high school respondents) was slightly better than the fit of the third group (teacher respondents) model and the MGCFA. Each dataset for each separate group was run as a five correlated factor model. This analysis tested the substantive meaning among subscales for each sample. The fit of these model were each appropriate (see Table 3). Factor correlations among each PECI-S subscale were then calculated (see Table 4).

Tests of Invariance

The four progressively restrictive models of factor invariance were conducted with mixed results. Fit indices for comparative groups, that is, for junior high vs. high school, junior high vs. teachers, high school vs. teachers, and MGCFA are reported. See Table 5.

*Model 1.*The between groups variant freely estimated factor loadings, error variances, and factor correlations of this model results indicate the set of PECI-S items does represent the same conceptual framework across all three groups of respondents when they are compared to each other. Chi-Square, CFI, TLI, and RMSEA estimates show evidence that a similar conceptual framework holds for junior high vs. high school, junior high vs. teachers, high school vs. teachers, and the MGCFA. See Table 5.

*Model 2.*The constrained between-groups factor correlations and the freely estimated between groups variant error variances and factor loadings results indicate the set of PECI-S items does represent equality of scaling units across all three groups. That is, the expected change on each item given a per-unit change in the latent variable is the same across all groups of respondents.

Model 3. The freely estimated factor correlations and error variances with constrained factor loadings across groups for student groups did not detect group differences in such items. However, for student groups compared to teachers, the test did signify group differences for items. This result is especially pronounced in the student vs. teacher and by the MGCFA which signifies that teachers especially have a different mean score on such items than do students.

Model 4. The across groups constrained factor loadings, factor correlations, and error variances for all three samples indicated that all three groups being non-equivalent in terms of reliability. This result is especially pronounced between each student group compared to teachers. Likewise, the poorer fit of each comparison shows that even these samples of junior

high and high school students significantly differ in instrument scores having a detrimental impact on the equivalence of scale reliabilities across samples.

Internal consistency estimates.

Internal consistency estimates (coefficient alpha) were calculated for the aggressive, harmful, low engagement, hyperactive, and defiant factors. The PECI parent scale appropriately reflects a single construct, "disruptive behaviors," with an average factor alpha = .81.
For the PECI-S, coefficient alpha estimates ranged from .57 (High school respondents on the "Low engagement" factor) to .86 (Junior high school and teacher respondents on the "Aggressive" factor) with an average alpha = .71. These estimates are available in Table 6. PECI and PECI-S subscales exhibited moderate to high PPM correlations with each other. PECI factors I, II, and IV had identical or nearly-identical factor names, and also exhibited high PPM correlations with factors I, II and III of the PECI-S. These estimates are available in Table 7.

Second-order models.

In Table 8, we see that the AGLS and robust ML estimates support a second order model for both junior high and high school students. Although the robust ML estimates support a second order factor for the teacher and merged respondent sets, the AGLS estimates do not. Despite model fit however, all four samples exhibited alphas of approximately .92 for the second order model. This suggests unidimensionality of the PECI-S when examined as a higher order model, as well as high inter-item correlations for each sample.

Discussion

Upon inspection of the basic item statistics in Table 2, means and standard deviations between student respondent groups appear to be similar. The teacher group overall showed a smaller standard deviation statistic for each item. This may be due to smaller sample size, more

homogenous agreement in disruptive behavior salience, or both. Interestingly the increased positive skew and kurtosis on junior high school respondents' endorsements of items 05, "Smoking" and 45, "Drug use" shows evidence of skewed and kurtotic responses. This might indicate that students in this group do not endorse these two behaviors as salient to their experience in the classroom.

The measurement model CFAs indicate appropriate, although not excellent fit for each group. An examination of the fit statistics for the model in which factor loadings, correlations, and error variances were fixed indicated adequate fit on comparative student-to-student invariance indices. However, the model fit for invariance is supported less for student-to-teacher comparisons even though the chi-square difference tests were significant. Thus, evidence of measurement invariance across student-teacher samples was not found, indicating the structure of the five-factor model mainly differed between both student groups' and teachers' responses on the PECI-S.

Upon examination of the alphas of the five individual PECI-S factors, they appear to be lower in comparison to the larger parent scale. This may be due to the shrinkage of intercorrelations of items within factors upon deletion of other items from the parent scale in constructing the PECI-S. This may also be due to a reduction of unidimensionality (i.e. the scale reflecting a single construct) from the PECI to the PECI-S. Although alpha estimates were lower for the PECI-S, they are still large enough to be promising in future research.

It appears both conceptually and analytically that we were successful in developing a shortened version of the PECI. From the EFA we were able to make appropriate decisions regarding factor structure and item retention. We have also produced a parsimoniously fitting model as exhibited by the measurement model results. The short scale appears to measure

classroom misbehaviors among three groups of respondents, not only overall like the 59-item scale, the substantive validity among the factors has also been preserved. Based on the above results we believe we have developed an optimal length scale producing reliable and valid scores measuring viewpoints on disruptive student behaviors. Although moderately different, this instrument can measure both students and teachers' viewpoints on issues of disruptive behaviors impacting classroom management. Although strict invariance of measures across student groups and teacher was not met, it was not necessarily expected to be a one-to-one match between either group of students and teachers. From the results of the internal consistency estimates we can see that our subscales and total scale scores show moderate to good reliability. The 59-item shortened to 20-item scale seems to be appropriate in representative disruptive behaviors and is useful for administration to more than a more homogenous group of respondents.

Recommendations

The next methodological task then will be to further validate the 20-item measure by collecting a new sample of responses, and then correlating the short form factors with the factors of the parent form. This effort will provide further information about factor overlap required to yield evidence of both construct and content validity.

References

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40*, 37 – 47.

Bentler, P. M. & Wu, E. J. C. (2005). *EQS Version 6.1 for Windows*. Encino, CA: Multivariate Software, Inc.

Bentler, P. M. (2002). *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc.

Bollen, K. A (1989). *Structural Equations with Latent Variables.* New York: Wiley.

Byrne, B. N. (2001). *Structural Equation Modeling with AMOS.* Rahwah, NJ: Lawrence Erlbaum Associates.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245 – 276.

Cronbach, L. J. (1951). Coefficient alpha and the and the internal structure of tests. *Psychometrika, 16,* 297 – 334.

Fan, X. & Sivo, S.A. (2005) Sensitivity of fit indices to mis-specified structural or measurement model components: rationale of two-index strategy revisited. *Structural Equation Modeling, 12*, 343 – 367.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179 – 185.

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

Hu, L. & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural Equation Modeling. Concepts, Issues, and Applications* (pp. 76-99). London: Sage.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 57*, 409 – 449.

Kulinna, P. H., Cothran, D., & Regualos, R. (2003). Refinement of an instrument to measure student disruptive behavior. *Measurement in Physical Education And Exercise Science, 7*, 25 – 41.

Messick, S. (1995). Validity of psychological measurement: Validation of inferences from a person's responses and performance in scientific inquiry into score meaning. *American Psychologist, 50*, 741 – 749.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32*, 396 – 402.

Poon, W.Y., & Lee, S. Y. (1987). Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika, 52*, 409 – 430.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

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

Steiger, J. H., Shapiro, A, & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential Chi-square statistics. *Psychometrika, 50*, 253 – 263

Widaman, K. F. (1993). Common factor analysis versus principal components analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research, 28*, 263 – 311.

Table 1

List of 20 Items in a Five Factor EFA Framework

Item	Factor/Item	Loading
	Aggressive (I)	
item26	Makes fun of other students	.70
item36	Cheating	.65
item37	Playing too rough and risking injury	.73
item46	Quarrelsome (Fusses with others)	.70
item50	Threatens others	.81
item52	Pushing or punching others	.81
	Harmful (II)	
item05	Smoking	.78
item40	Gang symbols displayed	.71
item45	Drug use	.82
	Low engagement (III)	
item28	Forms cliques (always hangs out with friends)	.72
item29	Acts shy and withdrawn	.82
item32	Gets upset when loses/fails	.54
item39	Continually saying, "I can't do it"	.74
	Hyperactive (IV)	
item06	Can't sit still (hyperactive)	.83
item10	Talking	.80
item14	Interrupts	.71
item17	Doesn't pay attention	.63
	Defiant (V)	
item51	"Smart mouth" toward students	.80
item56	Moves slowly on purpose	.61
item59	"Smart mouth" toward teacher	.71

Note. Extraction method: Principal axis factoring

Table 2

Basic Item Statistics for all Three Groups

		Junior High School (*N* = 1024)				High school (*N* = 1251)				Teachers (*N* = 303)			
Item	Factor and item nomenclature	Mean	*SD*	Skew	Kurtosis	Mean	*SD*	Skew	Kurtosis	Mean	*SD*	Skew	Kurtosis
item26	1 Makes fun of other students	3.07	1.35	0.02	-1.14	3.32	1.34	-0.23	-1.07	3.04	0.83	0.27	-0.03
item36	1 Cheating	2.15	1.23	0.90	-0.15	2.57	1.38	0.42	-1.04	2.59	0.86	0.27	0.12
item37	1 Playing too rough	2.56	1.29	0.49	-0.77	2.72	1.31	0.34	-0.92	2.77	0.83	0.61	0.65
item46	1 Quarrelsome (Fusses with others)	2.42	1.26	0.65	-0.50	2.53	1.20	0.48	-0.56	3.08	0.82	0.23	0.20
item50	1 Threatens others	2.13	1.27	0.97	-0.11	2.28	1.28	0.78	-0.43	2.33	0.84	0.68	0.90
item52	1 Pushing or punching others	2.49	1.35	0.58	-0.81	2.43	1.32	0.59	-0.78	2.67	0.90	0.46	0.03
item05	2 Smoking	1.18	0.67	4.41	19.81	1.41	0.90	2.58	6.42	1.21	0.49	2.49	6.51
item40	2 Gang symbols displayed	1.60	1.08	1.91	2.78	1.62	1.07	1.91	2.86	1.40	0.66	1.71	2.62
item45	2 Drug use	1.26	0.78	3.40	11.35	1.60	1.10	1.93	2.72	1.34	0.58	1.54	1.40
item28	3 Forms cliques	3.62	1.42	-0.55	-1.07	3.71	1.35	-0.68	-0.80	3.15	0.90	-0.18	0.09
item29	3 Acts shy and withdrawn	2.11	1.10	0.96	0.40	2.27	1.12	0.74	-0.03	2.46	0.63	0.27	0.35
item32	3 Gets upset when loses/fails	2.70	1.28	0.38	-0.81	2.79	1.28	0.21	-0.95	2.99	0.81	0.22	0.19
item39	3 Continually saying, "I can't do it"	2.51	1.28	0.49	-0.76	2.54	1.32	0.45	-0.93	2.67	0.79	0.50	0.34
item06	4 Can't sit still (hyperactive)	2.92	1.34	0.14	-1.13	3.08	1.36	0.00	-1.19	3.50	0.89	-0.21	-0.35
item10	4 Talking	4.35	0.98	-1.40	0.98	4.59	0.82	-2.17	4.39	3.94	0.78	-0.32	-0.38
item14	4 Interrupts	2.72	1.22	0.27	-0.78	2.72	1.23	0.30	-0.79	3.24	0.80	0.12	0.04
item17	4 Doesn't pay attention	2.95	1.20	0.09	-0.82	3.03	1.22	0.04	-0.84	3.33	0.77	0.07	0.07
item51	5 "Smart mouth" toward students	3.00	1.35	0.06	-1.14	3.19	1.34	-0.13	-1.12	2.98	0.82	0.19	0.17
item55	5 Moves slowly on purpose	2.74	1.32	0.32	-0.97	2.91	1.38	0.11	-1.18	2.72	0.91	0.45	-0.08
item59	5 "Smart mouth" toward teacher	2.60	1.37	0.43	-0.99	2.70	1.43	0.31	-1.20	2.58	0.91	0.43	-0.10

Table 3

CFA Fit Indices for Group Measurement Models

CFA Analysis	*df*	*AIC*	*SB-χ^2*	*TLI*	*CFI*	*RMSEA*	90% *CI*
Baseline (Combined)	160	752.44	1072.44	0.93	0.94	0.05	(.044 to .050)
Junior High School	160	180.76	500.76	0.92	0.92	0.05	(.049 to .057)
High School	160	333.17	653.17	0.90	0.92	0.05	(.049 to .057)
Teachers	160	-31.90	288.10	0.92	0.93	0.05	(.044 to .050)
MGCFA	480	748.06	1708.06	0.89	0.90	0.03	(.030 to .036)

Note. Satorra-Bentler (*SB-χ^2*) robust chi-square statistics reported.

Table 4

PPM Correlations Among PECI-S Subscales

	Sample/Correlation			
Subscale	I	II	III	IV
	Junior High			
II	.61	---		
III	.83	.46	---	
IV	.78	.44	.82	---
V	.94	.52	.85	.79
	High School			
II	.62	---		
III	.86	.56	---	
IV	.86	.58	.84	----
V	.96	.53	.88	.93
	Teachers			
II	.60	---		
III	.79	.42	---	
IV	.76	.49	.85	---
V	.94	.52	.86	.81
	MGCFA (*df* = 480)			
II	.39	---		
III	.80	.51	---	
IV	.76	.16	.68	---
V	.85	.48	.87	.68

Note. PECI-S Subscales: I = Aggressive, II = Harmful, III = Low engagement, IV =Hyperactive, V = Defiant

Table 5

Tests of Invariance Between Groups of Respondents

	Jr high and high school	*df*	χ^2	$\Delta\chi^2$	Δdf	*p*	CFI	TLI	RMSEA	90% CI
Model 1	Baseline model	320	1154.10	n/a	n/a	n/a	0.923	0.908	0.037	(.034 to .039)
Model 2	Factor correlations invariant	330	1190.46	36.36	10	> .001	0.920	0.908	0.037	(.034 to .039)
Model 3	Factor loadings invariant	333	1215.61	25.16	3	> .01	0.918	0.907	0.037	(.035 to .039)
Model 4	*Full restriction	363	1687.20	471.59	28	> .001	0.883	0.871	0.043	(.041 to .045)
	Jr high and teachers									
Model 1	Baseline model	320	796.724	n/a	n/a	n/a	0.918	0.903	0.036	(.033 to .040)
Model 2	Factor correlations invariant	330	859.797	63.07	10	> .001	0.909	0.896	0.038	(.035 to .041)
Model 3	Factor loadings invariant	333	835.068	24.73	3	> .01	0.914	0.902	0.037	(.033 to .040)
Model 4	Full restriction	363	2501.21	1666.15	28	> .001	0.763	0.739	0.072	(.070 to .075)
	High school and teachers									
Model 1	Baseline model	320	955.209	n/a	n/a	n/a	0.917	0.901	0.038	(.035 to .041)
Model 2	Factor correlations invariant	330	1062.11	106.90	10	> .001	0.904	0.890	0.040	(.038 to .044)
Model 3	Factor loadings invariant	333	1147.94	85.84	3	> .001	0.895	0.884	0.041	(.039 to .044)
Model 4	Full restriction	363	3025.91	1877.97	28	> .001	0.794	0.774	0.073	(.071 to .076)
	MGCFA									
Model 1	Baseline model	480	1456.52	1456.52	n/a	n/a	0.919	0.904	0.030	(.029 to .032)
Model 2	Factor correlations invariant	500	1585.42	128.90	20	> .001	0.910	0.897	0.031	(.030 to .033)
Model 3	Factor loadings invariant	530	2030.75	445.33	30	> .001	0.875	0.866	0.036	(.034 to .037)
Model 4	Full restriction	570	4332.69	2301.94	40	> .001	0.781	0.765	0.055	(.053to .056)

Note. *Full restriction = factor correlations, factor loadings, error variances invariant

Table 6

PECI-S Internal Consistency Estimates (coefficient alphas)

	Alphas				
Group/Subscale	I	II	III	IV	V
Junior High School	.86	.63	.62	.74	.78
High School	.84	.65	.57	.65	.76
Teachers	.86	.76	.61	.82	.75

Note. PECI-S Subscales: I = Aggressive, II = Harmful, III = Low engagement, IV =Hyperactive, V = Defiant

Table 7.

PPM Correlations of PECI and PECI-S Subscales

Correlations

Scale/Subscale	PECI scale (59 item)						PECI-S scale (20 item)				
PECI	I	II	III	IV	V	VI	I	II	III	IV	V
I	----										
II	0.79	----									
III	0.75	0.75	---								
IV	0.64	0.59	0.57	---							
V	0.62	0.63	0.59	0.48	---						
VI	0.65	0.64	0.60	0.54	0.42	---					
PECI-S											
I	0.95	0.75	0.71	0.59	0.56	0.62	---				
II	0.50	0.42	0.41	0.85	0.34	0.40	0.46	---			
III	0.64	0.78	0.61	0.42	0.59	0.67	0.60	0.30	---		
IV	0.67	0.66	0.80	0.44	0.76	0.48	0.61	0.31	0.54	---	
V	0.84	0.76	0.65	0.51	0.57	0.52	0.75	0.38	0.59	0.59	---

Note. Combined Samples Utilized

PECI Subscales: I = Aggressive, II = Low Engagement or Irresponsibility, III = Fails to Follow Directions,

IV = Illegal or Harmful, V = Distracts or Disturbs Others, VI = Poor Self-Management.

PECI-S Subscales: I = Aggressive, II = Harmful, III = Low engagement, IV =Hyperactive, V = Defiant

Table 8.

Second Order Model Fit Estimates

Sample	*N*	SB-χ^2	*df*	AFI	AAFI	TLI	CFI	RMSEA	90% CI	alpha
Jr High	1028	497.82	164	0.97	0.97	0.92	0.93	0.05	(.044 to .054)	0.92
HS	1251	584.91	164	0.95	0.94	0.92	0.92	0.05	(.048 to .057)	0.92
Teachers	303	3126.89	164	0.84	0.79	0.90	0.91	0.06	(.046 to .066)	0.91
Merged	2582	778.07	164	0.76	n/a	0.91	0.92	0.05	(.044 to .054)	0.92

Note. AFI = AGLS fit index, AAFI = Adjusted AGLS fit index, TLI = Tucker-Lewis Index