Article

School Psychology International

School Psychology International
2020, Vol. 41(3) 257–275

DOI: 10.1177/0143034320912301
journals.sagepub.com/home/spi
SAGE

Investigating school children's behavioral and emotional problems using pediatric symptoms checklist-17 in a structural equation modeling framework

Jin Liu
University of South Carolina, USA

Siying Guo
Pittsburg State University, USA

Ruiqin Gao and Christine DiStefano
University of South Carolina, USA

Abstract

The Pediatric Symptom Checklist-17 was originally used in primary care settings with parents to identify their children's behavioral and emotional problems, but there has been some research supporting use of this scale in school settings. This study examined: (a) the factor structure and measurement invariance of the teacher-rated Pediatric Symptom Checklist-17 and (b) complex relationships among demographic characteristics, behavioral and emotional problems, and learning outcomes using structural equation modeling in elementary schools. A sample of 508 children in grades one and two were rated by their teachers with the Pediatric Symptom Checklist-17. Measures of Academic Progress test was utilized to measure participants' learning outcomes in reading and math. The results confirmed a three-factor structure of the Pediatric Symptom Checklist-17 (internalizing problems, externalizing problems, and

Corresponding author:
Jin Liu, 145 Wardlaw, Columbia, SC 29208, USA.
Email: liu99@mailbox.sc.edu

attention problems) and attested the measurement invariance across different demographic groups (i.e. gender, ethnicity, and grade levels). Boys were more likely to have severe attention problems which were associated with lower learning outcomes as seen by Measures of Academic Progress reading and math scores. Attention problems mediated the relationship between gender and learning outcomes. This study has implications for the use of the Pediatric Symptom Checklist-17 in school-based settings. Additionally, it highlights the potential relationships among gender, attention problems, and learning outcomes.

Keywords
Pediatric Symptom Checklist-17, validation, demographic information, learning outcomes, structural equation modeling

Introduction

It is common knowledge that behavioral and emotional problems are negatively associated with children's academic performance levels (Barry et al., 2002; Chen et al., 1995; Flook et al., 2005; Polderman et al., 2010). Thus, measuring children's behavioral and emotional problems is of interest to parents, teachers, and school psychologists. Response to Intervention (RtI) has been adopted as a multi-tiered framework that can be used to identify and support children's academic and behavioral health needs. A universal method (i.e. screening measure) is often used to identify all children who are at risk for behavioral and emotional problems within Tier 1 of RtI (Carta & Greenwood, 2013). In Tier 2, at-risk children can receive comprehensive assessments and social–emotional interventions with progress monitoring. More intensive support/interventions may be offered for children demonstrating lack of progress in Tier 3. This systematic and evidence-based approach has potential benefit in providing key social–emotional experiences for early childhood populations who lack such experiences, which has the potential to prevent the need for special education services (Greenwood et al., 2011).

A high-quality screening measure is important at the initial tier for RtI to function appropriately. Though professionals have developed behavioral and emotional scales for young children, many of these may not be appropriate for universal screening due to their length and cost. For instance, the Achenbach System of Empirically Based Assessment (Achenbach & Rescorla, 2001), one of the most frequently used systems in the United States for child behavioral assessment, includes over 100 items which makes it difficult to use as a universal screening tool. The Behavior Assessment System for Children—Third edition Behavioral and Emotional Screening System—Teacher Rating Scale (Kamphaus & Reynolds, 2015) is short enough for universal screening (20 items), but the expense is approximately $1 per form, which may be cost prohibitive for certain schools.

Our focus is the Pediatric Symptom Checklist-17 (PSC-17), which helps identify and assess a broad range of emotional and behavioral and emotional problems in children (Pediatric Symptom Checklist, n.d.). It appears to have advantages as a universal screening tool, given: (1) the length of the form (i.e. 17 items) is efficient for universal screening and (2) the form is free and can be downloaded online.

The PSC-17 and relevant studies

Studies on the PSC-17 have been primarily conducted in the United States. Gardner and colleagues(1999) shortened the full 35-item PSC to the 17-item PSC-17, and multiple studies were conducted to validate the PSC-17 in primary care settings with parents. Previous studies investigated the factor structure of the scale (e.g. Chaffin et al., 2017; Stoppelbein et al., 2011) and identified three underlying factors: internalizing, externalizing, and attention problems. Other validity evidence should be considered in addition to factor structure. For instance, measurement invariance is a significant aspect of psychometric quality to help researchers interpret the scale properties across groups in the same way. Only one study has tested the measurement invariance of the PSC-17 based on our review. Using a sample of African American and Caucasian youth with and without a chronic illness, researchers concluded there was no strong factor invariance established (Stoppelbein et al., 2011).

To date, a few studies have been conducted to validate the PSC-17 in the preschool environment to determine if the PSC-17 may be an option for universal screening. Previous research has identified the same factors as in primary care settings: internalizing, externalizing, and attention problems (e.g. DiStefano et al., 2017). Operational definitions of the three subscales follow: externalizing problems are disruptive behaviors such as aggression and hyperactivity (e.g. takes things that do not belong to him/her); internalizing problems are feels for depression, worry, and anxiety (e.g. worries a lot); and attention problems are attention-deficit issues (e.g. has trouble concentrating).

In addition, Liu et al. (2020) investigated the measurement invariance with gender in preschool settings. While this study established strong gender invariance, the factor structure or measurement invariance of the PSC-17 for older children in other grade levels has not yet been investigated. A primary purpose of the study was to provide psychometric evidence in such settings.

Demographics and behavioral and emotional problems

After validating the PSC-17, researchers started to apply the scale in primary care settings to explore the relationships between participant demographics and the PSC-17 scores, with conflicting results. One study indicated there was no difference in parent-reported psychological issues by gender (Wren et al., 2004), whereas boys were more likely to have attention and externalizing problems than girls in another study (Borowsky et al., 2003). However, it remains unclear whether children's

demographics, such as grade level, gender, and ethnicity, affect the three factors underlying the PSC-17 simultaneously. Also, previous studies failed to establish measurement invariance before investigating such relationships. We reviewed studies that investigated relationships between demographics and behavioral and emotional problems in general due to lack of relevant research using the PSC-17.

Previous research has suggested different patterns of gender differences for different types of problems. Girls experienced more internalizing problems than boys (Albano & Krain, 2005; Fanti & Henrich, 2010). In contrast, a consistent pattern has not been observed for gender differences in externalizing problems. Some studies reported no gender differences (Keiley et al., 2000; Silver et al., 2010), while others indicated that boys exhibited more externalizing problems than girls (Foster, 2005; Moffitt et al., 2001). Furthermore, there was no gender difference in characteristics of attention deficit hyperactivity disorder (ADHD) core symptoms (Graetz et al., 2005). Rucklidge (2010) also indicated that boys and girls with ADHD were more similar than different, and there were no sex-specific ADHD profiles from a review.

Studies examining ethnicity differences in child psychopathology yielded inconsistent findings regarding differences in ratings of behavioral and emotional problems among different ethnic groups. Some reported significant ethnicity differences in prevalence of internalizing and externalizing disorders, such as depression, anxiety, and aggression (Austin & Chorpita, 2004; Laird et al., 2005), while others documented no differences in symptoms across ethnic groups (e.g. Cole et al., 1998). Finally, the prevalence of ADHD symptoms varied by ethnic groups (Epstein et al., 1998; Pastor et al., 2015; Rabiner et al., 2004).

With respect to age, internalizing problems tended to increase gradually from infancy to early childhood (Gilliom & Shaw, 2004). In contrast, externalizing problems on average tended to decrease gradually over time, but their development exhibited heterogeneous trajectories. The majority of children experienced low levels of externalizing problems over time. Most of the children had moderate or high levels of externalizing problems early in their life, which then decreased throughout childhood, and only a few children maintained high levels of externalizing problems constantly (Fanti & Henrich, 2007; Shaw et al., 2003). As for the relationship between age and attention problems, there was evidence that symptoms such as inattention and especially hyperactivity-impulsivity declined with age among ADHD children (Faraone et al., 2006; Willoughby, 2003).

These studies highlighted the relationships between demographic characteristics and three types of behavioral and emotional problems. There was no consensus on such relationships in general. The current study aimed to further explore the complex relationships using the PSC-17.

Behavioral and emotional problems and learning outcomes

Interestingly, although relationships between behavioral and emotional problems and learning outcomes in general are well-established in literature, few studies have

been conducted to explore such relationships using the PSC-17. This is of interest to researchers and educators who may be interested in using the PSC-17 to predict children's learning outcomes. We reviewed studies that investigated such associations with different scales and learning outcomes measures. In addition, few studies have been conducted empirically to examine how three types of behavioral and emotional problems related to learning outcomes simultaneously. Thus, our review was separated by different types of behavioral and emotional problems.

First, children with initial externalizing problems may have more limited opportunities for classroom learning, thereby impeding academic competence (Chen et al., 1997). These children were more likely to be rejected by socially healthy peers and accepted by deviant peers, resulting in disinterest in school, which in turn undermined learning outcomes (Chen et al., 1997; Schwartz et al., 2006). In addition, Risi et al. (2003) found that aggressive behaviors increased children's likelihood of being expelled from the school. Such actions exclude them from an environment in which academic competencies and skills can be developed and improved.

Internalizing symptoms could also erode children's cognitive functioning, hamper use of appropriate learning strategies, and disturb their attentional focus and active participation during learning activities in the classroom, which might result in poor academic performance (Maughan et al., 2003; Roeser et al., 2001). However, it is worth mentioning that some researchers found no relationship between internalizing problems and poor academic performance (e.g. Reinherz et al., 1993). Previous cross-sectional studies indicated that attention problems were associated with academic underachievement (Frazier et al., 2007; Rabiner et al., 2004). Furthermore, Polderman and colleagues' (2010) systematic review indicated that early attention problems related to poorer academic performance and lower GPA. We expected to identify negative associations between the PSC-17 factor scores and learning outcomes based on the review.

Demographics and learning outcomes

Academic achievement differs in different demographic groups. As prior studies have suggested, females demonstrated larger advantages in language courses and smaller advantages in math courses compared with males (Voyer & Voyer, 2014). Age and ethnicity differences were identified in learning outcomes as well (Mead, 2006), although patterns might differ depending on the sample used and such relationships may be mediated by other variables. For instance, researchers investigated the potential impact of mediators, such as personality, confidence and mental rotation, on gender difference in academic achievement levels (Carvalho, 2016; Casey et al., 1997). Personality and other individual factors were also considered as mediators in examining ethnicity differences in academic achievement levels (Woolf et al., 2013). However, few studies focused on the potential mediating effects of behavioral and emotional problems on such relationships.

Research purposes

The scarcity of psychometric evidence and applied research using the PSC-17 in school-based settings suggests an area for future research. The purpose of the current study was to provide validity evidence of the PSC-17 rated by teachers in grades one and two as a universal screening tool and then utilize the scale to examine complex relationships among variables.

This study investigated the psychometric properties of the PSC-17 among students in grades one and two in the school environment to extend the usage of the scale to a new setting. Correct identification of a scale's structure is critical and fundamental for further validation. Measurement invariance was then tested before the investigation of how demographic characteristics were related to the PSC-17 factor scores and learning outcomes (Chen et al., 2010). Next, the relationships between the PSC-17 subscales and academic achievement levels were investigated when controlling several key demographic variables. Finally, the mediating effects of externalizing, internalizing, and attention problems were tested.

Methods

Participants

Institutional Review Board approval was obtained at the beginning of the study. Teachers from an elementary school in the U.S. provided ratings of the PSC-17 for all students in grades one and two in Fall of 2014. Teachers' participation in the project was voluntary, and they received small stipends for completing the forms. In Spring of 2015, students took the Measures of Academic Progress (MAP) test (Northwest Evaluation Association, 2011)—reading and math—at the end of the semester. A sample of 508 first grade ($n = 241$, 47.44%) and second grade ($n = 267$, 52.56%) children was obtained. Female ($n = 233$, 45.87%) and male children ($n = 275$, 54.13%) were evenly distributed. The children were predominantly White ($n = 342$, 67.32%), and the rest of them represented diverse minority groups ($n = 166$, 32.68%). The minority groups included African-American, Hispanic, Mixed, Asian, and Indian. We considered those as one group due to the minimal cases in most minority groups. All students in the sample received free or reduced priced lunch.

Measures

The PSC-17 was originally shortened from the full PSC (Gardner et al., 1999) using exploratory factor analysis. The 17 items were on a three-point ordinal scale with anchors of "Never" = 0, "Sometimes" = 1, and "Often" = 2. The three underlying factors had high internal consistencies (internalizing: .79; externalizing and attention: .83). Item scores were summed and higher scores indicated a higher level of behavioral and emotional problems in each subscale. Recently, researchers identified the same factors with high internal consistencies (internalizing problems: .82;

externalizing problems: .88; attention problems: .87) and established measurement invariance of gender in preschool (DiStefano et al., 2017, Liu et al., 2020).

The MAP is a computer adaptive test and a norm referenced measure of student growth over time (Northwest Evaluation Association, 2011). It is used primarily as a formative assessment aiming to help teachers' better target instruction based on students' test performance. The test had decent internal consistency and made accurate classification decisions. Diverse validity evidence, such as content validity and criterion-validity evidence, are provided in the test manual (Northwest Evaluation Association, 2011). Rasch Unit Scale (RIT) scores were reported by considering item difficulty to measure students' achievement. The RIT scores were transferred into norm curve equivalent (NCE) scores using national norms to allow comparison across grade levels. The NCE scores were standardized scores with a mean of 50 and standard deviation of 21.06. The MAP reading and math NCE scores were used as the learning outcomes in the current study.

Data analysis

Data cleaning was performed in SPSS 23.0 and then transformed into M-plus (Muthén & Muthén,1998-2012) for data analysis. Approximately 5% of the MAP reading and math scores were missing from the sample. The Roderick J. A. Little's chi-square statistic for testing (Estimating Statistics and Imputing Missing Values, n.d.) results indicated that these values were missing completely at random ($p > .05$). Therefore, the Expectation–Maximization method using an iterative process was used to impute the missing data. The mean NCE scores of the MAP reading and math after imputation were 38.83 (SD = 20.84) and 39.95 (SD = 20.54). There were no missing data in other variables.

In the modeling process, we considered the clustering feature of data, in which one teacher rated multiple students' behavior in the same classroom. The factor structure of the PSC-17 was examined using confirmatory factor analysis (CFA) as the factor structure has been well-established in previous settings. CFA is the foundation of the following analysis to ensure the underlying factor structure is tenable in the new setting (Kline, 2015). We then tested measurement invariance to ensure that we had the same measurement and scaling properties across groups based on the validated factor structure. The configural (same factor structure across groups), metric (same factor structure and factor loadings across groups), and scalar (same factor structure, factor loadings, and item thresholds across groups) invariances were examined across groups for gender, ethnicity, and grade level respectively.

Multiple Indicators Multiple Causes (MIMIC) model was further constructed to investigate how gender, ethnicity, and grade level associated with the PSC-17 factors (i.e. internalizing, externalizing, and attention problems). MIMIC models can be used to investigate manifest (observed) variables as predictors of latent (unobserved) variables (Kline, 2015) with the assumptions of measurement invariance across groups.

We then examined whether behavioral and emotional problems were associated with students' learning outcomes (i.e. the MAP reading and math scores) by controlling the three demographic variables to examine the predictive validity of the PSC-17. Finally, mediation analyses were conducted to identify mediating effects of three types of behavioral and emotional problems (externalizing, internalizing, and attention problems) on the relationships between demographic variables (sex, ethnicity, and grade level) and academic achievement levels (MAP reading and math scores). For instance, there was a mediating effect of attention problems on the relationship between gender and MAP reading scores. In total, 18 mediating effects ($3 \times 3 \times 2$) were estimated.

Weighted least square mean and variance adjusted was used as the estimation method for ordered ordinal data (Muthén & Muthén, 1998-2012). The following fit information was used to locate the optimal solution of the CFA and the final structural equation model. First, the ratio of chi-square value to degree of freedom can be used as characteristics of model fit. A ratio of less than 3 indicates an acceptable model fit (Schermelleh-engel et al., 2003). Then, a comparative fit index (CFI) value of .95 or higher suggests good fit (Hu & Bentler, 1999). The root mean square error of approximation (RMSEA) uses a value of .08 as its cut-off to indicate adequate model fit (MacCallum et al., 1996). The Tucker–Lewis Index (TLI) measures relative model fit and a value of .95 indicates acceptable fit. The weighted root mean square residual (WRMR) is well suited for categorical data because it incorporates the asymptotic variances into the computation (Finney & DiStefano, 2006). Values under 1 have been recommended (DiStefano et al., 2018).

For measurement invariance testing, the models (configural, metric, and scalar) were compared using the scaled chi-square difference tests between every two models to determine if invariance can be established by adding constraints across groups with the DIFFTEST option in M-plus. Although an insignificant chi-square difference test result ($p > .05$) suggests invariance, researchers have suggested it is an unrealistic criterion for building invariance with a large sample size (Chen et al., 2005). Other fit indices, RMSEA, CFI, and TLI, were examined across the configural and scalar models with the suggested cut-off values. Specifically, CFI difference less than .01 would indicate a meaningful change in testing measurement invariance (Cheung & Rensvold, 2002).

Results

Factor analysis and measurement invariance

The first set of analysis was conducted to find the optimal factor structure of the PSC-17. The three-factor CFA model fit the data well ($\chi^2_{(116)} = 276.974$, $\chi^2/df = 2.388$, RMSEA = .052, CFI = .979, TLI = .976, WRMR = 1.392). The modification indices indicated that one item (i.e. does not listen to rules) should be loaded on the factor of attention problems instead of the externalizing problems. Thus, the

Table 1. Confirmatory factor loadings of the PSC-17 ($n = 508$).

	Items	Factor loadings (final CFA)	Factor loadings (final model)
Internalizing problems	Feels sad, unhappy	.986	.986
	Feels hopeless	.934	.933
	Is down on self	.917	.917
	Seems to have less fun	.839	.839
	Worries a lot	.803	.815
Attention problems	Fidgety, unable to sit still	.849	.847
	Daydreams too much	.772	.791
	Has trouble concentrating	.970	.973
	Acts as if driven by a motor	.892	.898
	Distracted easily	.951	.947
	Does not listen to rules	.911	.907
Externalizing problem	Refuses to share	.898	.890
	Does not understand other people's feelings	.864	.877
	Fights with other children	.810	.820
	Blames others for his/her troubles	.902	.909
	Teases other	.837	.844
	Take things that do not belong to him/her	.797	.822

final CFA model was adjusted by making the minor revision. The final CFA model illustrated better fit in terms of all fit indices ($\chi^2_{(116)} = 242.175$, $\chi^2/df = 2.088$, RMSEA = .046, CFI = .984, TLI = .981, WRMR = 1.269). Attention problems and internalizing problems were positively related (.393). The externalizing problems are moderately associated with internalizing (.553) and attention problems (.786). All items were strongly loaded on the principal factor (Table 1). Cronbach's alpha among the three factors all exceeded .87, suggesting high internal consistencies.

The measurement invariance of gender, ethnicity, and grade level were estimated (Table 2). All models showed adequate model fit in terms of RMSEA, TLI, and CFI, supporting configural, metric, and scalar models across demographic groups. As the chi-square difference tests were sensitive to large sample size, the significant results (i.e. gender—scalar and metric models) were considered acceptable as CFI difference across models were smaller than .01 (Cheung & Rensvold, 2002). Thus, MIMIC models can be constructed to investigate group comparisons.

Three demographic variables and learning outcomes (i.e. MAP scores) were added to the model. The Final model showed adequate fit: $\chi^2_{(186)} = 334.937$, $\chi^2/df = 1.801$, RMSEA = .040, CFI = .979, TLI = .975, WRMR = 1.185, with almost all fit indices within the suggested cut-off bounds. In addition, the factor loadings were stronger in the final model compared with the final CFA model. The structural MIMIC model included standardized regression coefficients from each covariate

Table 2. Test of measurement invariance across gender, ethnicity, and grade level.

Models	χ^2 (df)	RMSEA	Diff χ^2 (df) *p* value	CFI	TLI
Gender					
Configural model	398.767 (232)	.053		.982	.979
Metric model	412.762 (246)	.052	13.818 (14) .463	.982	.980
Scalar model	428.303 (260)	.050	26.265 (14) .024	.982	.981
Ethnicity					
Configural model	374.826 (232)	.049	–	.987	.985
Metric model	389.003 (246)	.048	12.334 (14) .580	.987	.986
Scalar model	393.220 (260)	.046	10.920 (14) .692	.987	.987
Grade level					
Configural model	362.786 (232)	.047		.985	.982
Metric model	377.737 (246)	.046	18.066 (14) .204	.985	.983
Scalar model	385.837 (260)	.044	9.651 (14) .787	.986	.985

RMSEA: root mean square error of approximation; CFI: comparative fit index; TLI: Tucker–Lewis Index.

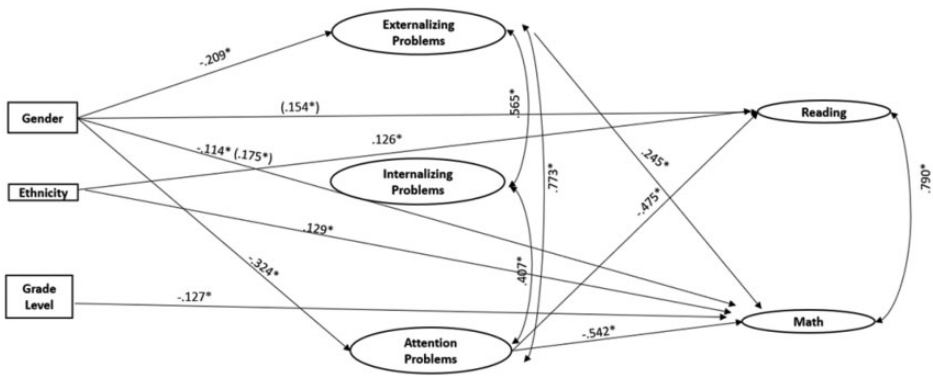

Figure 1. The PSC-17 in structural equation modeling framework.
Note: Statistically significant relationships are shown in the figure, and numbers in parentheses are mediating effects.
*$p < .05$.

being considered in the model and the identified factors (see Figure 1). Gender was directly associated with externalizing ($b = -.209$, $p < .001$) and attention problems ($b = -.324$, $p < .001$). Boys exhibited higher level of externalizing problems and attention problems than girls. In addition, boys had higher math achievement levels than

girls ($b = -.114$, $p < .001$). Ethnicity was directly associated with the MAP reading scores ($b = .126$, $p = .004$) and math scores ($b = .129$, $p = .001$). White children scored higher in reading and math than students in diverse minority groups. Older children had lower MAP math scores ($b = -.127$, $p = .009$). The severity of attention problems was directly associated with the MAP reading scores ($b = -.475$, $p < .001$) and math scores ($b = -.542$, $p < .001$), whereas internalizing problems was not significantly associated with learning outcomes. Interestingly, more externalizing problems were related with higher MAP math scores ($b = .245$, $p = .003$).

Of all tested mediation effects, only two of them were statistically significant. Specifically, gender indirectly influenced MAP reading scores ($b = .154$, $p < .001$) and math scores ($b = .175$, $p < .001$) through attention problems. In other words, girls, who tended to have lower attention problems, were more likely to have higher MAP reading and math scores.

Discussion

This study investigated the underlying factor structure and measurement invariance of gender, ethnicity, and grade level of the PSC-17 for students in grades one and two. It also investigated the complex relationships among three types of behavioral and emotional problems (internalizing problems, externalizing problems, and attention problems), demographic characteristics, and students' learning outcomes as measured by MAP reading and math scores.

Theoretical contributions

First, the three factor-structure of the PSC-17 was validated in elementary schools with a minor change. The item of "Do not listen to rules" was interpreted as a typical symptom of externalizing problems in primary care settings (Gardner et al., 1999), while it was primarily loaded on the factor of attention problems in the current study. The minor change might be due to the setting difference in the current sample (i.e. school-based settings and teacher ratings) compared with the original sample (i.e. primary care setting with parent ratings). "Listening to rules" may be more of a reflection of attention issues in classrooms.

The measurement invariance was supported by comparing model fit indices including three layers of constraints (i.e. configural, metric, and scalar) across demographic groups. The results for gender invariance were generally in concordance with previous literature on measurement invariance of the PSC-17 in the preschool setting (Liu et al., 2020).

The results of the final model indicated that certain demographic characteristics were associated with the underlying constructs measured by the PSC-17 and students' learning outcomes. Males were more likely to be reported as having attention problems and externalizing problems than females, which was consistent with previous literature (Borowsky et al., 2003; Foster, 2005; Rucklidge, 2010; Zevenbergen & Ryan, 2010).

Inclusion of gender norms when scoring protocols may provide more precise estimation of children's behavioral and emotional problems. This should be considered in the process of validating PSC-17 cut-off scores in school-based settings. Ethnicity was not significantly related to any subscales in the PSC-17. Thus, ethnicity was not a significant factor that relates to teachers' ratings of the PSC-17, which is a good feature of a screening scale. The results were generally in line with the varied results in previous literature (Cole et al., 1998; Pastor et al., 2015). In other words, such associations may differ by research settings. Finally, there was no difference in the PSC-17 subscale scores across grade one or grade two students, which suggests that the differences may not be striking in the current setting. These results should be interpreted with caution as only two grade levels (first and second) were examined. It is unclear if similar findings would emerge in samples representing a larger range of grade levels.

Next, students with more attention problems had lower achievement levels on the MAP reading and math tests. Researchers have documented that ADHD is related to poor reading and math standardized test scores (Loe & Feldman, 2007). The results of the current study were consistent with the previous literature (Frazier et al., 2007; Polderman et al., 2010; Rabiner et al., 2004; Wu et al., 2014). As attention problems impact students' reading and mathematics skills, appropriate screening can help teachers identify students at risk of attention problems so that early intervention can be provided to help students improve their attention and academic performance levels. Interestingly, more externalizing problems were related to higher math scores, which was inconsistent with previous research findings (e.g. Schwartz et al., 2006; Wu et al., 2014).

There were no significant relationships between internalizing problems and children's learning outcomes. Although the relationships are generally well documented (Chen et al., 1997), the results in the current study suggest that such relationship may not be obviously enough to be detected in the screening process in such a short period (i.e. around six months). Attention problems and externalizing problems may be more related to the short-term deficiencies in math and/or reading abilities compared with internalizing problems. In addition, some researchers did not find significant relationships between behavioral and emotional problems and academic performance either (Reinherz et al., 1993).

A significant relationship was identified between demographic variables and students' achievement scores. First, compared with boys, girls tended to have lower attention problems which in turn were related to higher MAP reading and math scores. The conclusions were consistent with findings of a recent meta-analysis (Voyer & Voyer, 2014), in which researchers concluded that females perform better in courses, such as language, mathematics, and science than males. Interestingly, boys had higher math scores than girls in the current study, and there was no significant difference between groups in reading scores if attention problems were not considered as a mediator. This result in the current study suggests the dual role of attention problems in influencing MAP scores and

mediating the relationships between gender and MAP scores. In practice, selection of mediators varies based on researchers' interests and hypotheses.

White students exhibited stronger reading and math skills than minority students in different groups. Previous studies have found such achievement gaps between White and minority students in early childhood, such as a Hispanic-White gap, Asian-White gap, and Black-White gap (Curran & Kellogg, 2016). Narrowing the racial gap in academic achievement is essential for maintaining equality in education. Finally, as students move to higher grade levels, they tended to have lower math scores. This was consistent with previous literature (Nelson et al., 2004). However, the current sample only included two grade levels, and we considered grade levels as a covariate for learning outcomes. We were thus conservative on drawing conclusions about the relationship between grade level and learning outcomes.

Practice implications

The study results could benefit school psychologists, teachers, and parents who may want to use the PSC-17 as a tool to assess children's problems. First, schools may work with teachers to incorporate this free tool into the universal screening process. School psychologists should be well-equipped with the skills to consult with other stakeholders about implementing relevant interventions for children who are at behavioral and emotional risk. As the scale is short, it may also be considered as a progress monitoring tool to investigate the effectiveness of certain interventions, although additional psychometric evidence for this purpose is needed. In addition, this tool may be especially useful to track how attentional problems at school associate with academic learning outcomes. School psychologists, teachers, and parents may work closely with those children who have severe attentional issues at schools and hopefully improve the academic achievement levels by addressing related problems.

Limitations and future research directions

The present study has several limitations. First, we only included gender, ethnicity, and grade level as factors associated with children's behavioral and emotional problems and learning outcomes. Other factors such as socio-economic status and the structure of the family should be considered to form a more comprehensive view. The study also only involved children in lower grades and future studies should be conducted in higher grade levels to figure out the behavioral and emotional patterns in different age groups as the scale was originally designed for children aged 4–16 years.

We also had a sample of children from one school, all of whom qualified for free or reduced priced lunch, which limited the generalizability of the findings to other school and socio-economic settings. Previous studies have identified the achievement gaps between White and a specific minority group. For instance, on average, White

students have higher mathematic and reading scores than Black and Hispanic students (Musu-Gillette et al., 2016; Reardon & Galindo, 2009). However, there were very few children in different minority groups, so we collapsed all minority children into one group in the current study. The achievement gaps among different ethnic groups should be examined with a qualifying sample in the future.

The current research briefly reported the mediating effects of attention problems on the relationships between gender and academic achievement levels. No detailed information was reported due to the focus of the study. Future studies can analyze the mediating role of behavioral and emotional problems on other relationships such as parental involvement and learning outcomes. Cut-off score analysis should be considered in the PSC-17 validation for it to be used as an universal screening tool to identify children's behavioral and emotional risk.

Certain relationships among variables were inconsistent with the previous literature. Researchers may conduct meta-analysis on such relationships to obtain more definitive conclusions. Finally, the PSC-17 is available in other languages, such as Spanish and Chinese. To date, no known studies have been conducted to validate the PSC-17 in the school setting outside of the United States. Researchers in other countries may consider validating the scale as a universal screening tool in their cultures. The forms can be freely accessed at this website: https://massgeneral.org/psychiatry/treatments-and-services/pediatric-symptom-checklist.

Despite the limitations, this study expanded the usage of the PSC-17 to the school environment and more diverse grade levels. Also, the study highlights the potential relationships among attention problems, gender, and learning outcomes for schools, teachers, and parents.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The current study is supported by the Institute of Education Sciences—Grant #: R324A100104.

ORCID iD

Jin Liu https://orcid.org/0000-0002-4242-6582

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. Research Center for Children, Youth, and Families.

Albano, A. M., & Krain, A. (2005). Anxiety and anxiety disorders in girls. In D. J. Bell, S. L. Foster, E. J. Mash (Eds.), *Handbook of behavioral and emotional problems in girls* (pp. 79–116). Springer.

Austin, A. A., & Chorpita, B. F. (2004). Temperament, anxiety, and depression: Comparisons across five ethnic groups of children. *Journal of Clinical Child and Adolescent Psychology*, *33*(2), 216–226. https://doi.org/10.1207/s15374424jccp3302_2

Barry, T. D., Lyman, R. D., & Klinger, L. G. (2002). Academic underachievement and attention-deficit/hyperactivity disorder: The negative impact of symptom severity on school performance. *Journal of School Psychology*, *40*(3), 259–283. https://doi.org/10.1016/s0022-4405(02)00100-0

Borowsky, I. W., Mozayeny, S., & Ireland, M. (2003). Brief psychosocial screening at health supervision and acute care visits. *Pediatrics*, *112*(1), 129–133. https://doi.org/10.1542/peds.112.1.129

Carta, J. J., & Greenwood, C. R. (2013). Future challenges and new directions in RTI. In V. Buysse & E. Peisner-Feinberg (Eds.), *Handbook of response to intervention in early childhood* (pp. 421–432). Paul H. Brookes.

Carvalho, R. G. G. (2016). Gender differences in academic achievement: The mediating role of personality. *Personality and Individual Differences*, *94*, 54–58. https://doi.org/10.1016/j.paid.2016.01.011

Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology*, *33*(4), 669. https://doi.org/10.1037//0012-1649.33.4.669

Chaffin, M., Campbell, C., Whitworth, D. N., Gillaspy, S. R., Bard, D., Bonner, B. L., & Wolraich, M. L. (2017). Accuracy of a pediatric behavioral health screener to detect untreated behavioral health problems in primary care settings. *Clinical Pediatrics*, *56*(5), 427–434. https://doi.org/10.1177/0009922816678412

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, *12*(3), 471–492. https://doi.org/10.1207/s15328007sem1203_7

Chen, H., Keith, T. Z., Weiss, L., Zhu, J., & Li, Y. Q. (2010). Testing for multigroup invariance of second-order WIS-IV structure across China, Hong Kong, Macau, and Taiwan. *Personality and Individual Differences*, *49*, 677–682. https://doi.org/10.1016/j.paid.2010.06.004

Chen, X., Rubin, K. H., & Li, D. (1997). Relation between academic achievement and social adjustment: Evidence from Chinese children. *Developmental Psychology*, *33*(3), 518. https://doi.org/10.1037/0012-1649.33.3.518

Chen, X., Rubin, K. H., & Li, Z. Y. (1995). Social functioning and adjustment in Chinese children: A longitudinal study. *Developmental Psychology*, *31*(4), 531. https://doi.org/10.1037/0012-1649.31.4.531

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/s15328007sem0902_5

Cole, D. A., Peeke, L. G., Martin, J. M., Truglio, R., & Seroczynski, A. D. (1998). A longitudinal look at the relation between depression and anxiety in children and adolescents. *Journal of Consulting and Clinical Psychology*, *66*(3), 451. https://doi.org/10.1037/0022-006x.66.3.451

Curran, F. C., & Kellogg, A. T. (2016). Understanding science achievement gaps by race/ethnicity and gender in kindergarten and first grade. *Educational Researcher*, *45*(5), 273–282. https://doi.org/10.3102/0013189x16656611

DiStefano, C., Liu, J., & Burgess, Y. (2017). Investigating the structure of the Pediatric Symptoms Checklist in the preschool setting. *Journal of Psychoeducational Assessment*, *35*(5), 494–505.

DiStefano, C., Liu, J., Jiang, N., & Shi, D. (2018). Examination of the weighted root mean square residual: Evidence for trustworthiness?. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 453–466.

Epstein, J. N., March, J. S., Conners, C. K., & Jackson, D. L. (1998). Racial differences on the Conners Teacher Rating Scale. *Journal of Abnormal Child Psychology*, *26*(2), 109–118. https://doi.org/10.1023/A:1022617821422

Estimating Statistics and Imputing Missing Values. (n.d.). https://ibm.com/support/knowledgecenter/en/SSLVMB_23.0.0/spss/mva/spssmva_estimating.html

Fanti, K. A., & Henrich, C. C. (2007). The relation of home and childcare/school environment to differential trajectories of externalizing problems. *International Journal about Parents in Education*, *1*, 117–123.

Fanti, K. A., & Henrich, C. C. (2010). Trajectories of pure and co-occurring internalizing and externalizing problems from age 2 to age 12: Findings from the National Institute of Child Health and Human Development Study of Early Child Care. *Developmental Psychology*, *46*(5), 1159. https://doi.org/10.1037/a0020659

Faraone, S. V., Biederman, J., & Mick, E. (2006). The age-dependent decline of attention deficit hyperactivity disorder: A meta-analysis of follow-up studies. *Psychological Medicine*, *36*(2), 159–165. https://doi.org/10.1017/s003329170500471x

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–314). Information Age Publishing.

Flook, L., Repetti, R. L., & Ullman, J. B. (2005). Classroom social experiences as predictors of academic performance. *Developmental Psychology*, *41*(2), 319. https://doi.org/10.1037/0012-1649.41.2.319

Foster, S. L. (2005). Aggression and antisocial behavior in girls. In D. Bell, S. Foster, & E. Mash (Eds.), *Handbook of behavioral and emotional problems in girls. Issues in clinical child psychology* (pp. 149–180). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/0-306-48674-1_5

Frazier, T. W., Youngstrom, E. A., Glutting, J. J., & Watkins, M. W. (2007). ADHD and achievement: Meta-analysis of the child, adolescent, and adult literatures and a concomitant study with college students. *Journal of Learning Disabilities*, *40*(1), 49–65. https://doi.org/10.1177/00222194070400010401

Gardner, W., Murphy, J. M., Childs, G., Kelleher, K., Pagano, M., Jellinek, M., McInerny, T. K., Wasserman R. C., Nutting, P., & Chiappetta. L. (1999). The PSC-17: A brief Pediatric Symptom Checklist with psychosocial problem subscales. A report from PROS and ASPN. *Ambulatory Child Health*, *5*, 225–236.

Gilliom, M., & Shaw, D. S. (2004). Codevelopment of externalizing and internalizing problems in early childhood. *Development and Psychopathology*, *16*(2), 313–333. https://doi.org/10.1017/s0954579404044530

Graetz, B. W., Sawyer, M. G., & Baghurst, P. (2005). Gender differences among children with DSM-IV ADHD in Australia. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(2), 159–168. https://doi.org/10.1097/00004583-200502000-00008

Greenwood, C. R., Bradfield, T., Kaminski, R., Linas, M., Carta, J. J., & Nylander, D. (2011). The response to intervention (RTI) approach in early childhood. *Focus on Exceptional Children*, *43*(9), 1–22. https://doi.org/10.17161/foec.v43i9.6912

Hu, L. T., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kamphaus, R. W., & Reynolds, C. R. (2015). *Behavior assessment system for children—Third edition (BASC-3): Behavioral and emotional screening system (BESS)*. Pearson.

Keiley, M. K., Bates, J. E., Dodge, K. A., & Pettit, G. S. (2000). A cross-domain growth analysis: Externalizing and internalizing behaviors during 8 years of childhood. *Journal of Abnormal Child Psychology*, *28*(2), 161–179. https://doi.org/10.1023/A:1005122814723

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Publications.

Laird, R. D., Pettit, G. S., Dodge, K. A., & Bates, J. E. (2005). Peer relationship antecedents of delinquent behavior in late adolescence: Is there evidence of demographic group differences in developmental processes? *Development and Psychopathology*, *17*(1), 127–144. https://doi.org/10.1017/s0954579405050078

Liu, J., DiStefano, C., Burgess, Y., & Wang, J. (2020). Pediatric Symptom Checklist-17: Testing measurement invariance of a higher-order factor model between boys and girls. *European Journal of Psychological Assessment*, *36*, 77–83.

Loe, I. M., & Feldman, H. M. (2007). Academic and educational outcomes of children with ADHD. *Journal of Pediatric Psychology*, *32*(6), 643–654. https://doi.org/10.1093/jpepsy/jsl054

MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037/1082-989x.1.2.130

Maughan, B., Rowe, R., Loeber, R., & Stouthamer-Loeber, M. (2003). Reading problems and depressed mood. *Journal of Abnormal Child Psychology*, *31*(2), 219–229. https://doi.org/10.1023/A:1022534527021

Mead, S. (2006). The evidence suggests otherwise: The truth about boys and girls. http://educationsector.org/usr_doc/ ESO_BoysAndGirls.pdf

Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2001). *Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study*. Cambridge University Press. https://doi.org/10.1017/cbo9780511490057

Musu-Gillette, L., Robinson, J., McFarland, J., KewalRamani, A., Zhang, A., & Wilkinson-Flicker, S. (2016). *Status and trends in the education of racial and ethnic groups 2016 (NCES 2016-007)*. National Center for Education Statistics.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide: Statistical analysis with latent variables* (7th ed.). Muthén & Muthén.

Nelson, J. R., Benner, G. J., Lane, K., & Smith, B. W. (2004). Academic achievement of K-12 students with emotional and behavioral disorders. *Exceptional Children*, *71*(1), 59–73. https://doi.org/10.1177/001440290407100104

Northwest Evaluation Association. (2011). *Technical manual: For Measures of Academic Progress (MAP) and Measures of Academic Progress for Primary Grades (MPG)*.

Pastor, P. N., Reuben, C. A., Duran, C. R., & Hawkins, L. D. (2015). *Association between diagnosed ADHD and selected characteristics among children aged 4–17 years: United States, 2011–2013*. (NCHS Data Brief No. 201). Centers for Disease Control and Prevention.

Pediatric Symptom Checklist. (n.d.). https://massgeneral.org/psychiatry/services/treatmentprograms.aspx?id = 2088

Polderman, T. J. C., Boomsma, D. I., Bartels, M., Verhulst, F. C., & Huizink, A. C. (2010). A systematic review of prospective studies on attention problems and academic achievement. *Acta Psychiatrica Scandinavica*, *122*(4), 271–284. https://doi.org/10.1111/j.1600-0447.2010.01568.x

Rabiner, D. L., Murray, D. W., Schmid, L., & Malone, P. S. (2004). An exploration of the relationship between ethnicity, attention problems, and academic achievement. *School Psychology Review*, *33*(4). https://doi.org/10.17105/SPR45-2.250-267

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*(3), 853–891. https://doi.org/10.3102/0002831209333184

Reinherz, H. Z., Giaconia, R. M., Pakiz, B., Silverman, A. B., Frost, A. K., & Lefkowitz, E. S. (1993). Psychosocial risks for major depression in late adolescence: A longitudinal community study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *32*(6), 1155–1163. https://doi.org/10.1097/00004583-199311000-00007

Risi, S., Gerhardstein, R., & Kistner, J. (2003). Children's classroom peer relationships and subsequent educational outcomes. *Journal of Clinical Child and Adolescent Psychology*, *32*(3), 351–361. https://doi.org/10.1207/s15374424jccp3203_04

Roeser, R. W., van der Wolf, K., & Strobel, K. R. (2001). On the relation between social–emotional and school functioning during early adolescence: Preliminary findings from Dutch and American samples. *Journal of School Psychology*, *39*(2), 111–139. https://doi.org/10.1016/s0022-4405(01)00060-7

Rucklidge, J. J. (2010). Gender differences in attention-deficit/hyperactivity disorder. *Psychiatric Clinics*, *33*(2), 357–373. https://doi.org/10.1016/j.psc.2010.01.006

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research – Online*, *8*(2), 23–74.

Schwartz, D., Gorman, A. H., Nakamoto, J., & McKay, T. (2006). Popularity, social acceptance, and aggression in adolescent peer groups: Links with academic performance and school attendance. *Developmental Psychology*, *42*(6), 1116. https://doi.org/10.1037/0012-1649.42.6.1116

Shaw, D. S., Gilliom, M., Ingoldsby, E. M., & Nagin, D. S. (2003). Trajectories leading to school-age conduct problems. *Developmental Psychology*, *39*(2), 189. https://doi.org/10.1037/0012-1649.39.2.189

Silver, R. B., Measelle, J. R., Armstrong, J. M., & Essex, M. J. (2010). The impact of parents, child care providers, teachers, and peers on early externalizing trajectories. *Journal of School Psychology*, *48*(6), 555–583. https://doi.org/10.1016/j.jsp.2010.08.003

Stoppelbein, L., Greening, L., Moll, G., Jordan, S., & Suozzi, A. (2011). Factor analyses of the Pediatric Symptom Checklist-17 with African-American and Caucasian pediatric populations. *Journal of Pediatric Psychology*, *37*(3), 348–357. https://doi.org/10.1093/jpepsy/jsr103

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*(4), 1174. https://doi.org/10.1037/a0036620

Willoughby, M. T. (2003). Developmental course of ADHD symptomatology during the transition from childhood to adolescence: A review with recommendations. *Journal of Child Psychology and Psychiatry*, *44*(1), 88–106. https://doi.org/10.1111/1469-7610.t01-1-00104

Woolf, K., McManus, I. C., Potts, H. W., & Dacre, J. (2013). The mediators of minority ethnic underperformance in final medical school examinations. *British Journal of Educational Psychology*, *83*(1), 135–159. https://doi.org/10.1111/j.2044-8279.2011.02060.x

Wren, F. J., Bridge, J. A., & Birmaher, B. (2004). Screening for childhood anxiety symptoms in primary care: Integrating child and parent reports. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(11), 1364–1371. https://doi.org/10.1097/01.chi.0000138350.60487.d3

Wu, S. S., Willcutt, E. G., Escovar, E., & Menon, V. (2014). Mathematics achievement and anxiety and their relation to internalizing and externalizing behaviors. *Journal of Learning Disabilities*, *47*(6), 503–514. https://doi.org/10.1177/0022219412473154

Zevenbergen, A. A., & Ryan, M. M. (2010). Gender differences in the relationship between attention problems and expressive language and emerging academic skills in preschool-aged children. *Early Child Development and Care*, *180*(10), 1337–1348. https://doi.org/10.1080/03004430903059292

Author biographies

Jin Liu has a PhD in Educational Research and Measurement from the University of South Carolina. Her research interests include ethics issues in assessment, peer assessment in online courses, applied research in public health, psychometric studies for scales measuring young children's behavioral and emotional problems, and applied structural equation modeling.

Siying Guo is an Assistant Professor of Criminal Justice at the Department of History, Philosophy, and Social Sciences, Pittsburg State University. Her research interests include bullying, juvenile delinquency and justice, youth development, religion and crime, life-course criminology, and quantitative methodology.

Ruiqin Gao is a PhD student at the University of South Carolina in the Educational Studies Department. Her research interests include classroom assessment and structural equation modeling.

Christine DiStefano is a Professor at the University of South Carolina where she teaches measurement theory and structural equation modeling courses. Her primary research focus is on structural equation modeling with categorical data. Her other research interests include examining behavioral and social constructs to measure behavioral risk in school settings.