



HHS Public Access

Author manuscript

Psychol Assess. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

Psychol Assess. 2015 December ; 27(4): 1324–1336. doi:10.1037/pas0000101.

DISC Predictive Scales (DPS): Factor Structure and Uniform Differential Item Functioning Across Gender and Three Racial/Ethnic Groups for ADHD, Conduct Disorder, and Oppositional Defiant Disorder Symptoms

Margit Wiesner,

Department of Educational Psychology, University of Houston

David E. Kanouse,

RAND Health, RAND Corporation, Santa Monica, California

Marc N. Elliott,

RAND Health, RAND Corporation, Santa Monica, California

Michael Windle, and

Department of Behavioral Sciences and Health Education, Emory University

Mark A. Schuster

Division of General Pediatrics, Boston Children's Hospital/Harvard Medical School, Boston, Massachusetts

Abstract

The factor structure and potential uniform differential item functioning (DIF) among gender and three racial/ethnic groups of adolescents (African American, Latino, White) were evaluated for attention deficit/hyperactivity disorder (ADHD), conduct disorder (CD), and oppositional defiant disorder (ODD) symptom scores of the DISC Predictive Scales (DPS; Leung et al., 2005; Lucas et al., 2001). Primary caregivers reported on *DSM-IV* ADHD, CD, and ODD symptoms for a probability sample of 4,491 children from three geographical regions who took part in the Healthy Passages study (mean age = 12.60 years, $SD = 0.66$). Confirmatory factor analysis indicated that the expected 3-factor structure was tenable for the data. Multiple indicators multiple causes (MIMIC) modeling revealed uniform DIF for three ADHD and 9 ODD item scores, but not for any of the CD item scores. Uniform DIF was observed predominantly as a function of child race/ethnicity, but minimally as a function of child gender. On the positive side, uniform DIF had little impact on latent mean differences of ADHD, CD, and ODD symptomatology among gender and racial/ethnic groups. Implications of the findings for researchers and practitioners are discussed.

Keywords

DISC Predictive Scales; factor structure; differential item functioning; MIMIC modeling; measurement invariance

The assessment of child and adolescent attention deficit/hyperactivity disorder (ADHD), conduct disorder (CD), and oppositional defiant disorder (ODD) has drawn much attention in recent decades (Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000), motivated to a large degree by their high prevalence, substantial comorbidity, and adverse correlates and outcomes (Chen, Killea-Jones, & Vega, 2005; Dalsgaard, Mortensen, Frydenberg, & Thomsen, 2002; Fergusson, Horwood, & Ridder, 2005; Loeber, Burke, Lahey, Winters, & Zera, 2000; Maughan, Rowe, Messer, Goodman, & Meltzer, 2004; Nock, Kazdin, Hiripi, & Kessler, 2007; Waschbusch, 2002). Evaluation of these disorders with structured diagnostic interviews, such as the comprehensive and extensively tested Diagnostic Interview Schedule for Children-IV (DISC-IV; Shaffer et al., 2000), can be time-consuming. Hence, shorter inventories have been developed that can be used as efficient screening tools for Diagnostic and Statistical Manual of Mental Disorders (DSM; American Psychiatric Association, 1994) diagnoses. An example is the DISC Predictive Scales instrument (DPS; Lucas et al., 2001; Leung et al., 2005), which has parallel parent and youth versions to report on youth behaviors.

The DPS was developed with the aim of retaining only the subset of DISC items that are the most salient predictors of full-length DISC DSM diagnoses (according to forward stepwise logistic regressions) and also ensure good sensitivity and specificity (according to receiver-operating characteristic [ROC] curves). DISC items that were significant predictors ($p < .05$) of the corresponding full-length DISC DSM diagnosis in the regression models were further examined and retained if they maximized sensitivity and specificity. As reported in Leung et al. (2005) and Lucas et al. (2001), the DPS scores achieved good-to-excellent sensitivity and specificity, and accounted for large areas under the ROC curve (from 0.72 to 0.99). Lucas et al. (2001) reported on the reliability and criterion validity of the DPS scores in relation to *DSM-III-R* diagnoses; Leung et al. (2005) presented reliability and criterion validity information for an updated version of the DPS scores in relation to *DSM-IV* diagnoses. Following these two original reports, subsequent studies also examined the psychometric properties of the DPS scores (McReynolds, Wasserman, Fisher, & Lucas, 2007; Roberts, Stuart, & Lam, 2008). Overall, the literature provides much evidence supporting the reliability and criterion validity of the DPS scores. Relatively little work has focused on its factor structure and the possibility of DIF.

With regard to factor structure, Rubio-Stipec et al. (1996) conducted exploratory factor analyses with items from the DISC 2.3 (i.e., an older version of the DISC) but Lucas et al. (2001) noted that the resulting scales do not fully map onto the DPS. A confirmatory factor analysis (CFA) will thus add important information about the factor structure of the DPS. Regarding possible DIF, invariance of measurement properties across groups must be established to ensure that group comparisons with DPS scores are valid. This can be done by testing for DIF in a latent variable analysis, such as multiple indicators multiple causes (MIMIC) modeling (for an overview of alternative approaches, see Zumbo, 2007). MIMIC models can be described as CFA models with covariates (Brown, 2006); the measurement model relates the manifest variables (items) to the latent factors, and the structural model is used to estimate the direct effects of one or more covariates on item responses and latent factors. If these tests reveal that the difficulty parameter of an item (i.e., the probability of

endorsing any particular item response option) differs among individuals *after* their level on the latent factor is controlled, then the item is said to exhibit “uniform DIF” (Camilli & Shepard, 1994; Mellenbergh, 1989).

MIMIC analyses are increasingly used to screen for uniform DIF in mental health scales, including ADHD (Gomez, 2010) and depression symptom scores (Gomez, Vance, & Gomez, 2012; Grayson, Mackinnon, Jorm, Creasey, & Broe, 2000). Some studies used other statistical techniques to test for potential DIF with ADHD, ODD, or CD symptom scores (Burns, Walsh, Gomez, & Hafetz, 2006; Gelhorn et al., 2009; Gomez, 2007; Gomez, Burns, & Walsh, 2008; Hillemeier, Foster, Heinrichs, Heier, & the Conduct Problems Prevention Research Group, 2007). Results from both sets of studies revealed minimal gender DIF for parent-reported *DSM-IV* ODD (Burns et al., 2006) and ADHD symptom scores (Burns et al., 2006; Gomez, 2007). Evidence for uniform DIF among gender groups was stronger for youth-reported *DSM-IV* CD symptom scores (Gelhorn et al., 2009). Weak to moderate uniform DIF was found among four racial/ethnic groups (Australian, Malaysian, Malaysian-Chinese, Malaysian-Malay) of children for parent-reported *DSM-IV* ODD symptom scores (Gomez et al., 2008). Evidence of uniform DIF among two racial/ethnic groups in the United States (African American, White) was considerably stronger for parent-reported *DSM-III-R* ADHD symptom scores (Hillemeier et al., 2007). Overall, the literature on uniform DIF for these disorders is limited in several ways, including that Latino youth are rarely represented in this area of research. Also, prior studies have used items that are not identical with DPS items and mostly examined these issues at younger ages (often elementary school-age children). We are unaware of research on DIF across gender or racial/ethnic groups with DPS scores. This study was designed to address these issues by focusing on the ADHD, CD, and ODD items of the DPS.

Study Aims

This study had two aims: to examine the factor structure of the DPS and to conduct an exploratory analysis of whether item scores in the DPS show uniform DIF as a function of gender and race/ethnicity. First, data from a large random sample of 7th graders and their primary caregivers from three geographical areas in the U.S. were used to examine the factor structure of the DPS. It was expected that the CFA would provide support for a 3-factor solution (congruent with the ADHD, CD, and ODD subscales derived by the scale authors). Second, a MIMIC analysis was conducted to explore possible uniform DIF of these item scores as a function of gender and race/ethnicity (African American, Latino, White). Examining the validity of comparing rates of ADHD, CD, and ODD symptoms, as assessed by DPS scores, among these demographic subpopulations is critical both for theoretical and applied purposes. Frameworks such as the Attribution Bias Context Model (De Los Reyes & Kazdin, 2005) have been posited in recent years in recognition of the accumulating evidence that informant bias in children’s mental health assessment should not be considered solely a form of measurement error because informants also systematically differ in various ways (e.g., what they attribute to be the cause of the mental health symptoms; the social context in which the child is observed). Another strand of work has revealed that ethno-cultural factors influence how parents perceive and interpret their offspring’s mental health problems and that this may partially explain the well-documented racial/ethnic disparities in children’s

mental health problems and services utilization in the U.S. (Bussing, Schoenberg, & Perwien, 1998; Roberts, Alegría, Roberts, & Chen, 2005; Snowden & Yamada, 2005). In sum, there are many reasons why item scores of the DPS might exhibit uniform DIF, especially as a function of race/ethnicity. Because the literature on this issue for the DPS is scant, we did not formulate a priori hypotheses about which DPS item score(s) would show uniform DIF but rather engaged in a systematic exploratory examination.

Method

Participants

This research used data from the second wave of Healthy Passages, a study funded by the Centers for Disease Control and Prevention (CDC) that assesses adolescent health—related behaviors, outcomes, and risk and protective factors for a cohort of 5,147 fifth graders and their primary caregivers (PCGs). The students were in fifth grade at baseline and in seventh grade at the second wave of Healthy Passages. Baseline data were collected from 2004 to 2006 and data for the second wave of assessment two years later. Qualitative (i.e., focus groups, cognitive interviews) and quantitative studies were conducted during study development to evaluate the appropriateness of survey language, translation, field procedures, and language-specific study materials. Background, project history, and conceptual framework of this study are described elsewhere (Schuster et al., 2012; Windle et al., 2004).

Students were recruited from fifth-grade classrooms in public schools in each of three geographic areas: (a) 25 contiguous public school districts in Los Angeles County, CA, (b) 10 contiguous public school districts in and around Birmingham, AL, and (c) the largest public school district in Houston, TX. Eligible schools had an enrollment of at least 25 fifth graders and collectively represented over 99% of students enrolled in regular academic classrooms in the three geographic areas. A cluster probability sampling procedure was used to select schools from each site. Within the randomly sampled schools, all English- and Spanish-speaking fifth graders enrolled in regular academic classrooms were invited to participate. Design weights were constructed to reflect different school selection probabilities based on racial/ethnic composition. Nonresponse weights were created to model nonresponse as a function of school, student gender, and student race/ethnicity. These two sets of weights were combined into a final weight that represented the population of fifth graders in the public schools in each site's geographic area. Of the 11,532 fifth graders enrolled in the 118 randomly selected schools, 6,663 of their PCGs who either agreed to be contacted about the study or who were unsure were invited to participate; 5,147 (77.3%) of them completed an interview at baseline. Retention was nearly 93% at Wave 2.

After excluding 321 children who were not among the three main target racial/ethnic groups (i.e., African American, Latino, White), 2 children whose race/ethnicity was unknown, and 333 children with missing PCG data on the Wave 2 DPS instrument, a final analytic sample of 4,491 (87.3%) remained for this study. The sociodemographic characteristics of the total analytic sample and of each racial/ethnic group are presented in Table 1. Latino children were more likely to be excluded from analysis due to missing data on the Wave 2 DPS than were African American or White children (8.1%, 6.2%, and 6.1%, respectively, $p < .001$).

Procedures

All three Healthy Passages research sites used standardized data collection materials and protocols, including training manuals, field manuals, and validation procedures. Institutional review boards at each study site and the CDC approved the study. Materials about the study and the Permission to Contact Form were distributed to eligible students in their classrooms. Students were asked to take them home. A home visit was scheduled if PCGs agreed to learn more about the study. After obtaining informed PCG consent and child assent, interview teams conducted separate interviews with the child and the PCG either in their home or at another location (e.g., on-campus site). Computer-assisted personal interviews (CAPI) as well as audio-computer assisted self-interviews (A-CASI) were used in the data collection (for more details, see Windle et al., 2004). On average, it took about 1.5 hr to complete the interview at the second wave of assessment. PCGs received \$60 and children a \$30 gift card from a national chain store as an honorarium for participation in the second wave of assessment.

Measures

This study used only sociodemographic data (mostly gathered during the CAPI with the primary caregiver) and the DPS items from the second wave of Healthy Passages (i.e., when students were in seventh grade). Although the DPS items were also administered at baseline (i.e., when students were in fifth grade), endorsement rates were very low for several of these items at baseline. This was most pronounced for the items assessing CD symptoms (half of the CD items had endorsement rates below 2.5% at baseline). This greatly limited our ability to conduct latent variable analyses of their factor structure because estimation tends to be unstable as the joint occurrence of two symptoms with extremely low base rates becomes very small. For example, Muthén, Hasin, and Wisnicki (1993) noted that with a typical joint probability of 0.01, even a sample size of $N = 4,000$ is barely large enough for stable estimation with a set of binary items that have base rates ranging from 1% to 26%. Hence, only data from the second wave, in which endorsement rates for DPS items were more favorable (Table 2), were used in this study.

ADHD, CD, and ODD symptoms—The presence of ADHD, CD, and ODD symptoms in the past year was assessed by PCGs with 26 items adapted from the Diagnostic Interview Schedule for Children Predictive Scales (DPS; Leung et al., 2005; Lucas et al., 2001). The DPS is a widely used screening tool that is based on the Diagnostic Interview Schedule for Children (Chen et al., 2005; Leung et al., 2005). It has been shown to identify children who display symptoms of 11 *DSM-IV* (American Psychiatric Association, 1994) diagnoses (Leung et al., 2005). The adapted DPS (the impairment-related questions were not administered) was approved by C.P. Lucas for the Healthy Passages study. It was translated into Spanish using a standard translation procedure (i.e., translation/back-translation and committee approach) recommended for ensuring linguistic equivalence (Sireci, Yang, Harter, & Ehrlich, 2006). About 24% ($N = 1,091$) of PCGs completed the assessment in Spanish language; 76% ($N = 3,400$) used the English language version.¹ The ADHD subscale was based on 8 items, the CD subscale on eight items, and the ODD subscale on 10 items. PCGs rated the presence of each symptom on a dichotomous scale (1 = *yes*, 0 = *no*) during the A-CASI portion of the Wave 2 field interview. A bilingual voice actress was used to record

both the English and Spanish A-CASI portions of the interview. Subscale scores were calculated by summing affirmative responses across the items. The prevalence (%) of affirmative responses for each item, item-subscale correlations, and internal consistency for all subscale scores are reported in Table 2.

Statistical Analysis

Analyses were conducted using *Mplus 7.0* (Muthén & Muthén, 1998–2012) with weights (to account for differential probabilities of selection of students according to their school and differential nonresponse) and a cluster variable (to account for clustering of students within schools). Models were tested using robust weighted least squares (WLSMV) estimation, which accommodates binary data and provides robust standard errors and adjusted test statistics (Muthén, du Toit, & Spisic, 1997).

Evaluation of the models was based on multiple criteria that considered statistical, practical, and substantive fit: The comparative fit index (CFI) ranges in value from zero to one; CFI values greater than .90 and .95 typically reflect acceptable and good model fit, respectively, of a target model relative to the null model (Bentler, 1990; Hu & Bentler, 1999). The root mean squared error of approximation (RMSEA) is a measure of a model's approximate lack of fit in the population. Values less than .05 indicate good fit and values as high as .08 represent acceptable errors of approximation in the population (Browne & Cudeck, 1993; Steiger, 1990). Finally, the weighted root mean square residual (WRMR) is a more recently developed residual-based fit index proposed by Muthén and Muthén (1998–2001); values below one are indicative of good model fit according to a simulation study (Yu, 2002). The performance of the WRMR has not been extensively evaluated in Monte Carlo simulation research; therefore, less emphasis was placed on this experimental fit index. Values for the corrected chi-square statistic were reported only for comparison purposes because this statistic is not invariant to sample size and is known to be an overly sensitive index of model fit when there are large numbers of constraints, especially with large samples (Bentler, 1990; Marsh, Balla, & McDonald, 1988). Additional criteria such as local model misfit (e.g., inspection of residuals) and interpretability of parameter estimates were also used.

¹Preliminary analyses examined descriptive statistics, internal consistency, factor structure, and potential uniform DIF of DPS item scores for each language version. Internal consistencies of the three subscale scores were very similar for both language versions (ranging from .67 to .81 for the English-language and from .69 to .81 for the Spanish-language version). Next, a null model, a one-factor CFA model, and the hypothesized 3-factor CFA model (see Figure 1) were estimated for each language version. Results showed that the hypothesized 3-factor CFA model had excellent fit for the English-language ($N = 3,400$, $WLSMV\chi^2(296) = 1132.53$, $CFI = .955$, $RMSEA = .029$, $RMSEA\ 90\% \text{ CI} = .027, .031$, $WRMR = 2.240$) and the Spanish-language ($N = 1,091$, $WLSMV\chi^2(296) = 463.18$, $CFI = .964$, $RMSEA = .023$, $RMSEA\ 90\% \text{ CI} = .019, .027$, $WRMR = 1.334$) version. Correlations among the three latent factors ranged from .65 to .84 for the English-language and from .71 to .86 for the Spanish-language version, all at $p < .001$. MIMIC analyses with Oort's correction (see Results section for details) and language version as covariate revealed uniform DIF for some DPS item scores. However, some of these tests relied on extremely low base rates in the Spanish-language data. Overall, these preliminary analyses revealed a fairly high (albeit not perfect) degree of measurement equivalence. Findings from this study should be evaluated mindful of this potential measurement limitation (Knight, Roosa, & Umaña-Taylor, 2009); as a precaution, all main analyses of this study were repeated using data from only the English-language version (see Results section).

Results

CFA Models

To address the first study aim, three CFA model specifications were tested to evaluate the adequacy of the hypothesized 3-factor structure for the DPS item scores in this sample: a *null model*, a *one-factor CFA model* (all items loaded on a single latent factor), and the hypothesized *3-factor CFA model* (each item was freely estimated to load only on one factor; all latent factors were allowed to correlate with one another). Figure 1 depicts the hypothesized 3-factor CFA model for ease of comprehension. The factor loading of the first item of a latent factor was fixed to the value 1.0 for scaling purposes. Fit statistics for these CFA models are shown in the upper section of Table 3 (see M1, M2, and M3).

Based on the criteria of model fit, small residuals, and adequacy and interpretability of parameter estimates, the hypothesized 3-factor CFA model provided a good fit to the data. The parameter estimates for the 3-factor CFA model are shown in Table 4. All factor loadings were statistically significant (all $p < .001$) and ranged from .52 to .86 in the completely standardized solution. The proportion of explained variance for the latent response variables y^* ranged from .27 to .75 (it was above .50 for 18 of the 26 DPS items). Combined, these results indicate adequate convergent validity for these item scores, some of which captured relatively rare symptoms. The correlations among the latent factors from the completely standardized solution were .67 (ADHD with CD), .69 (ADHD with ODD), and .85 (CD with ODD)², all $p < .001$. These correlations indicate that discriminant validity was good for the ADHD latent factor with respect to both the CD and ODD latent factors, but marginal between the CD and ODD latent factors.³

MIMIC Models

The second study aim, screening for uniform DIF of the DPS items, was addressed via MIMIC models (Kim, Yoon, & Lee, 2012; Woods, 2009) in which covariates were added to the 3-factor CFA model. The MIMIC approach was chosen because it allows for the simultaneous inclusion of several covariates with two or more groups, in this case, gender (1 = *boy*, 0 = *girl*), African American race/ethnicity (1 = *yes*, 0 = *no*), and Latino race/ethnicity (1 = *yes*, 0 = *no*); White race/ethnicity served as reference category. Effects of these covariates were controlled for each other. Another reason for choosing the MIMIC approach was that prevalences of several CD symptoms in the overall sample were quite low, which is in keeping with the nature of these symptoms (see Table 2). Invariance testing with

²The high correlation between the ODD and CD latent factors indicated considerable redundancy. Therefore, we additionally tested a two-factor CFA model which consisted of a ADHD latent factor (all ADHD items were specified to load on the ADHD factor) and a combined ODD/CD latent factor (all ODD and CD items were specified to load on the combined ODD/CD latent factor). Both latent factors in this model were allowed to covary. This alternative two-factor CFA model also provided a good fit to the data (WLSMV $\chi^2(298) = 1408.33$, CFI = .955, RMSEA (90% CI) = .029 (.027, .030), WRMR = 2.472). We did not retain the two-factor CFA model as the final model because the proportions of explained variance in the latent response variables y^* were lower for several items compared to the 3-factor CFA model (e.g., the proportion of explained variance in the latent response variable y^* had dropped to .37 for Item 25 in the two-factor CFA model) and the latent factor structure was misaligned both with clinical diagnostic practice and the hypothesized factor structure of the DPS. We readily acknowledge that there might be situations where the more parsimonious two-factor CFA model specification is preferred.

³Although there is no firm rule for discriminant validity pertaining to CFA model tests, correlations with other latent factors $< .7$ are frequently accepted as evidence of discriminant validity; correlations $> .85$ are usually viewed as problematic because they are indicative of some redundancy of the latent factors (Brown, 2006; Kline, 2011).

multigroup CFA would have been problematic for rarely occurring symptoms (e.g., cruelty toward animals) due to empty or near-empty cells in some subgroups. MIMIC modeling inherently assumes the absence of “nonuniform DIF” (Camilli & Shepard, 1994; Mellenbergh, 1989), that is, each item discrimination parameter (aka, factor loading)⁴ is presumed to be group-invariant. Because the DPS, and its parent instrument DISC, have been refined over decades drawing on rich clinical diagnostic information and extensive evaluations of criterion validity, this assumption was considered plausible for our analysis.

Using the approach from Kim et al., (2012), a *baseline MIMIC model* in which the latent factors validated in the CFA analysis, but none of the DPS items, were simultaneously regressed on all covariates (i.e., a full-invariance MIMIC model) was estimated. For ease of comprehension, the baseline MIMIC model is shown in Figure 2. Next, the baseline MIMIC model was compared to multiple *less constrained MIMIC models* in which one direct effect of a covariate on a DPS item was added (i.e., testing uniform DIF with $df = 1$ for one item and covariate at a time). Note that all covariates still had direct effects on each latent factor in the less constrained MIMIC models. The factor variances were fixed to the value 1 for model identification so that it was possible to freely estimate the factor loadings of all items. For each model comparison, the WLSMV model chi-square difference test was conducted using the *Difftest* feature in *Mplus*; a significant difference test indicated uniform DIF for the given item. To control for Type I error inflation (i.e., false detection of uniform DIF for invariant items) during model comparisons, the Oort adjustment to the chi-square difference test was used (Oort, 1998; see Kim et al., 2012). Oort’s correction adjusts the critical chi-square value to account for potential model misspecification in the full-invariance baseline model and was found to control Type I error rates at or below the nominal level (it also compared favorably to Bonferroni corrections) and to maintain high power across different study conditions in a recent simulation study from Kim et al. (2012). An application of this adjustment procedure is provided in Ogg, McMahan, Dedrick, and Mendez (2013).

Based on an adjusted critical chi-square value of 17.988 for a nominal alpha of .05, uniform DIF was identified for several DPS item scores during this model testing process. Hence, the baseline 3-factor MIMIC model was modified and direct effects of covariates on these DPS items were freely estimated in the final 3-factor MIMIC model. Fit statistics for both MIMIC models are shown in the lower section of Table 3 (see M4, M5). Model fit of both MIMIC models was good. Although not discussed in more detail, it might be asked whether substantive findings and model fit differed as a function of PCG gender. We were unable to estimate the models for just the male PCGs because of their small group size. However, substantive findings and model fit were robust when the models were re-estimated using only data from female PCGs (see Table 3 notes). Model fit and the vast majority of substantive findings also were closely replicated when, for the reasons described in Footnote

⁴The reader is reminded that uniform DIF is to be distinguished from nonuniform DIF. Uniform DIF exists when only the item difficulty parameter differs across groups; nonuniform DIF exists when the item discrimination parameter differs across groups (i.e., group membership interacts with the latent trait factor level) (cf., Chan, 2000). A nontechnical description of how CFA with binary manifest variables is equivalent to a two-parameter normal ogive item response theory model is provided in Brown (2006). In CFA models with categorical manifest variables, the item difficulty parameters are analogous to item thresholds, and the item discrimination parameters correspond to factor loadings (cf. Muthén, Kao, & Burstein, 1991).

1, models were re-estimated using only the data gathered with the English-language version (see Table 3 notes).

Parameter estimates for the direct effects of covariates on individual DPS items that were included in the final 3-factor MIMIC model are shown in Table 5. Substantively, the results of the final 3-factor MIMIC model indicated uniform DIF for the following item scores: Male gender of the child was positively related to the presence of two ODD symptoms (Items 9 and 15; all $p < .001$) but inversely related to two other ODD symptoms (Items 10 and 11; all $p = .001$). Relative to being White, Latino race/ethnicity of the child was positively linked to the presence of one ADHD symptom (Item 6; $p < .001$) and three ODD symptoms (Items 9, 11, and 12; all $p < .01$). It also was inversely related to the presence of two ADHD symptoms (Items 3 and 5; all $p < .001$) and three other ODD symptoms (Items 16, 17, and 18; all $p < .001$). Compared to White race/ethnicity, African American race/ethnicity of the child was positively related to the presence of one ADHD symptom (Item 6; $p < .001$) and four ODD symptoms (Items 9, 12, 14, and 15; all $p < .001$). At the same time, it was inversely related to the presence of another ADHD symptom (Item 3; $p < .001$) and two other ODD symptoms (Items 16 and 18; all $p < .001$). No uniform DIF was observed for CD item scores.

Next, a sensitivity analysis of the associations of the covariates to the means of the latent factors was conducted in order to explore the extent to which group comparisons at the scale (aka, latent factor) level might be biased as a result of the identified uniform DIF of some DPS item scores. Findings from this sensitivity analysis are shown in Table 6, which contrasts the estimated direct effects of all covariates on the three latent factors from the baseline 3-factor MIMIC model (i.e., which did not include any direct effects of covariates on individual DPS items) with those from the final 3-factor MIMIC model (in which the direct effects of covariates on individual DPS items described in the preceding paragraph were included). Overall, little difference was found regardless of whether latent mean differences as a function of the three covariates were adjusted for uniform DIF in the MIMIC analysis or not. Substantively, the analyses showed that boys had significantly higher means on the ADHD, CD, and ODD latent factors compared to girls (all $p < .001$). Latino children had significantly lower means on the ODD latent factor ($p < .001$) but did not differ significantly from White children on the mean levels of the ADHD and CD factors ($p > .17$). Finally, African American children had significantly higher means on the ADHD ($p < .001$) and CD ($p < .05$) latent factors but significantly lower means on the ODD latent factor ($p = .001$) relative to White children.

Discussion

This study applied CFA and MIMIC modeling to data from a large representative community sample in the U.S. to examine uniform DIF of the ADHD, CD, and ODD item scores of the DPS as a function of gender and race/ethnicity of seventh graders. Results from CFA models were based on the parent version of the instrument and indicated that the hypothesized 3-factor structure provided a good approximation of the data. Lucas et al. (2001; see also Leung et al., 2005) had selected DISC items for inclusion in the DPS that emerged as significant predictors of DSM diagnoses in secondary data analyses. Given the

scant published evidence on the factor structure of the DPS, this result was reassuring and an important study contribution.

The MIMIC models revealed uniform DIF of various item scores, especially ODD item scores but also some ADHD item scores, whereas this was not the case for CD item scores. The last finding was inconsistent with a prior study from Gelhorn et al. (2009) with youth-reported CD symptoms. It is possible that power for detecting uniform DIF was more limited in our study due to relatively low base rates of parent-reported CD symptoms.⁵ Furthermore, uniform DIF was mostly observed for comparisons by race/ethnicity of the child but minimally for comparisons by child gender, which was largely consistent with the limited literature (Burns et al., 2006; Gomez, 2007; Hillemeier et al., 2007). In contrast to previous studies, uniform DIF effects of gender and race/ethnicity were controlled for each other. The inclusion of Latino youth also was a unique study contribution.

It is often difficult to identify the reasons why subsets of items within a scale function differently across groups (Zumbo, 2007). Misspecification of an underlying multidimensional model has been offered as one explanation (Ackerman, 1992). Secondary (in the worst case, nuisance) factors might exist that are correlated with the primary factor of interest and systematically related to the variance of items. Although we are being very speculative, we offer three interpretations that might explain why some DPS item scores showed uniform DIF. Yet, nuisance factors or measurement artifacts clearly remain a viable alternative interpretation.

First, it has been found that ethno-cultural factors can influence thresholds for the acceptability of youths' behavior (Weisz, McCarthy, Eastman, Chaiyasit, & Sunwanlert, 1997) as well as parents' interpretation of youths' mental health symptoms (Roberts et al., 2005). For example, given the importance that Latino cultures often place on values such as *respeto* and *simpatía*⁶ (Calzada, Fernandez, & Cortes, 2010; Triandis, Marín, Lisansky, & Betancourt, 1984), it is conceivable that individual ODD symptoms such as "refused to do what s/he was told to do" (Item 9) and "argued or talked back" (Item 18) might be endorsed by Latino PCGs at a different rate compared to White PCGs even when overall mean levels of ODD symptomatology on the latent factor are held constant, as found in the current study. However, the opposite patterns of uniform DIF observed across individual ODD items indicate that these effects likely resulted from more complex processes. Second, it has been documented that racial/ethnic minority children in the U.S., especially those of Latino race/ethnicity, have high rates of mental health services underutilization (Alegría, Vallas, & Pumariega, 2010; Kataoka, Zhang, & Wells, 2002; Snowden & Yamada, 2005). Some have suggested that this might be the result of racial/ethnic (aka, cultural) differences in parents' decision thresholds guiding whether treatment is warranted for specific mental health problems Alegría et al., 2004; Bussing et al., 1998; Chavez, Shrout, Alegría, Lapatin, &

⁵The two highest base rates in the Gelhorn et al. (2009) study were found for items "steal no confront" (47.2% male and 34.8% female youths endorsed this item) and "destruction of property" (21.5% male and 6.6% female youths endorsed this item). Another interesting comparison concerns the item "cruel to animals," which was endorsed by 9.6% male and 1.3% female youths in the Gelhorn et al. (2009) study.

⁶*Respeto* represents the obedience, duty, and deference of an individual's position within a hierarchical structure; *simpatía* emphasizes the importance of displaying behaviors that promote smooth and pleasant social relationships (Castillo, Perez, Castillo, & Ghosheh, 2010, pp. 164, 165).

Canino, 2010; Yeh et al., 2005; see also De Los Reyes & Kazdin, 2005). Our finding that the ADHD item “taking medication for hyperactivity” (Item 3)⁷ was less likely to be endorsed for African American and Latino children relative to White children, even when their overall mean level of ADHD symptomatology on the latent factor was held constant, fits well with other research on this issue (Eiraldi, Mazzuca, Clarke, & Power, 2006; Rowland et al., 2002). Third, race/ethnicity is often confounded with various sociodemographic factors in the United States, including family educational level and household income (see, e.g., Harrell, Langton, Berzofsky, Couzens, & Smiley-McDonald, 2014; Lahey et al., 1995). The descriptive information shown in Table 1 indicates that White race/ethnicity is strongly confounded with high socioeconomic status (SES), whereas SES is lowest among Latino children in this sample. Therefore, it is possible that the observed uniform DIF among racial/ethnic groups represents differences in parents’ SES as much as (or more than) differences in ethno-cultural factors. In conclusion, it is important to gain a better understanding of the sources of the observed uniform DIF of specific DPS item scores in future research. Ideally, this should involve use of a confirmatory approach (Zumbo, 2007), and evidence is beginning to emerge that shows how this might be done in applied research (Sandilands, Oliveri, Zumbo, & Ercikan, 2013).

Lastly, latent mean differences of the three factors as a function of gender and race/ethnicity of the child were robust regardless of whether estimates were adjusted for uniform DIF or not. This was most likely the case because uniform DIF of individual items did not consistently favor one group over others (e.g., after holding constant the mean levels of the ODD latent factor, Latino race/ethnicity of the child was positively related to three ODD items and inversely to three other ODD items compared to the reference category White race/ethnicity). In other words, item-level group differences found for specific DPS items might have balanced one another out at the total scale-level, resulting in robust patterns of latent mean differences across the two MIMIC model specifications. This tentative conclusion needs to be cross-validated in other research. Substantively, it is of interest that for some (e.g., boys had higher levels of CD symptoms than girls; African American children had higher levels of ADHD symptoms than White children), though not all, latent mean differences in the three factors among demographic subgroups were consistent with patterns of group differences obtained with the DPS for a nationally representative sample of 12–17 year-olds (Chen et al., 2005). The inconsistent group differences between both studies (no significant differences between male and female youth in levels of ODD or among African American, Latino, and White youth in levels of ODD and CD symptoms were found in the Chen et al., 2005, study) likely stem from a combination of informant (parent vs. youth DPS version) and other method effects (e.g., nonadjustment vs. adjustment for measurement error, assessment of *DSM-III-R* vs. *DSM-IV* symptoms, different age ranges of assessed youths).

⁷An anonymous reviewer noted that Item 3 (“taking medication for hyperactivity”) differs from the other items of the DPS ADHD-subscale in that it does not assess a behavioral manifestation of ADHD symptomatology. Due to the absence of independent clinical diagnoses, this study was unable to provide additional information about the sensitivity and specificity of Item 3 scores. However, the results reported in this study indicate that the base rate of Item 3 was much lower compared to those of the other DPS ADHD items and its item-subscale correlation was of a moderate effect size (see Table 2). In the 3-factor CFA model, the latent ADHD factor accounted for 43% of the variance in the latent continuous response variable for Item 3 (see Table 4). We concur with the reviewer that further investigation of the sensitivity and specificity of Item 3 would be highly informative for the field.

There are several limitations of this study. First, PCGs provided responses to the DPS items. Independent clinical diagnoses and youth-self-report DPS data for ADHD, CD, and ODD symptoms were not available for the seventh graders but would be useful to address concerns about informant bias (for a discussion, see De Los Reyes & Kazdin, 2005; Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012). Second, this study focused on ADHD, CD, and ODD symptoms. Similar research should be conducted for other DPS mental health symptoms that were not assessed in the Healthy Passages study (e.g., obsessive-compulsive disorder). Third, the study was conducted with a large representative community sample of seventh graders from three major racial/ethnic groups attending public schools in three metropolitan areas in the United States. Results may not generalize to special populations and clinical samples, other racial/ethnic groups (e.g., Asian Americans, Native Americans), younger or older youths, and different geographical regions. Fourth, there is no consensus in the expert literature on the best way of testing for uniform DIF in MIMIC modeling (Woods, 2009). Although the Oort adjustment, used in this study to control Type I error inflation, has performed well in a recent simulation study (Kim et al., 2012), further simulation work is needed to more comprehensively evaluate its performance and statistical power (e.g., with disproportionate group sizes, more noninvariant items, larger samples, or more items). Fifth, MIMIC modeling, although advantageous and justifiable in this study for the reasons described earlier, is not readily suited to detect non-uniform DIF. Some experts have begun to propose extensions through inclusion of latent moderated structures that may allow us to address this limitation in the future (Woods & Grimm, 2011). Finally, longitudinal measurement invariance testing would provide another important extension of this cross-sectional analysis.

Summarizing, findings demonstrated uniform DIF for several DPS item scores, but this had little impact on latent mean differences of the ADHD, CD, and ODD factors. These results have various implications for applied purposes. Although it is fairly common in the literature on educational achievement and aptitude measures to delete items exhibiting uniform DIF from the test, this practice might be less indicated in the assessment of mental health problems (Gitchel, Turner, & Rumrill, 2010) because deleting items with uniform DIF from mental health assessment scales could negatively affect their content validity. We concur with these authors and do not advocate the exclusion of DPS items that displayed uniform DIF, especially considering that the number of DPS items per subscale is limited. Rather, we see the main contribution of our study as a first step toward better understanding the cognitive and/or ethno-cultural processes behind PCGs' responses to items about their child's mental health symptoms and investigating whether these processes are comparable across demographic groups. In the long run, identification of uniform DIF of some DPS item scores can assist practitioners in better targeting efforts designed to improve parents' recognition of mental health symptoms among their offspring toward those symptoms that tend to be underrecognized by specific subgroups.

Pending cross-validation, our findings imply that practitioners and researchers using the parent-version of the DPS in the U.S. can compare the scale scores of ADHD, CD, and ODD symptoms across three racial/ethnic groups (African American, Latino, White children) and gender with minimal bias. However, they should be cautious with comparing these groups at the item-level, particularly for ODD symptoms but also for a few ADHD symptoms, because

some of these items vary with regard to the degree that they measure severity equally across the three racial/ethnic groups (this is much less of an issue for item-level comparisons by gender). Whether this also applies to the youth self-report version of the DPS is an open question.

Acknowledgments

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

The Healthy Passages study was supported by cooperative agreements (CCU409679, CCU609653, CCU915773, U48DP000046, U48DP000057, U48DP000056, U19DP002663, U19DP002664, and U19DP002665) with the Centers for Disease Control and Prevention.

References

- Ackerman TA. A didactic explanation of item bias, item impact, and item avlidity from a multidimensional perspective. *Journal of Educational Measurement*. 1992; 29:67–91. <http://dx.doi.org/10.1111/j.1745-3984.1992.tb00368.x>.
- Alegría M, Canino G, Lai S, Ramirez RR, Chavez L, Rusch D, Shrout PE. Understanding caregivers' help-seeking for Latino children's mental health care use. *Medical Care*. 2004; 42:447–455. [PubMed: 15083105]
- Alegría M, Vallas M, Pumariega AJ. Racial and ethnic disparities in pediatric mental health. *Child and Adolescent Psychiatric Clinics of North America*. 2010; 19:759–774. <http://dx.doi.org/10.1016/j.chc.2010.07.001>. [PubMed: 21056345]
- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th ed. Author; Washington, DC: 1994.
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107:238–246. <http://dx.doi.org/10.1037/0033-2909.107.2.238>. [PubMed: 2320703]
- Brown, TA. *Confirmatory factor analysis for applied research*. Guilford Press; New York, NY: 2006.
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing structural equation models*. Sage; Newbury Park, CA: 1993. p. 136-162.
- Burns GL, Walsh JA, Gomez R, Hafetz N. Measurement and structural invariance of parent ratings of ADHD and ODD symptoms across gender for American and Malaysian children. *Psychological Assessment*. 2006; 18:452–457. <http://dx.doi.org/10.1037/1040-3590.18.4.452>. [PubMed: 17154767]
- Bussing R, Schoenberg NE, Perwien AR. Knowledge and information about ADHD: Evidence of cultural differences among African-American and white parents. *Social Science & Medicine*. 1998; 46:919–928. [http://dx.doi.org/10.1016/S0277-9536\(97\)00219-0](http://dx.doi.org/10.1016/S0277-9536(97)00219-0). [PubMed: 9541077]
- Calzada EJ, Fernandez Y, Cortes DE. Incorporating the cultural value of respeto into a framework of Latino parenting. *Cultural Diversity and Ethnic Minority Psychology*. 2010; 16:77–86. <http://dx.doi.org/10.1037/a0016071>. [PubMed: 20099967]
- Camilli, G.; Shepard, LA. *Methods for identifying biased test items*. Sage; Thousand Oaks, CA: 1994.
- Castillo LG, Perez FV, Castillo R, Ghosheh MR. Construction and initial validation of the Marianismo Beliefs Scale. *Counselling Psychology Quarterly*. 2010; 23:163–175. <http://dx.doi.org/10.1080/09515071003776036>.
- Chan D. Detection of differential item functioning on the Kirton Adaptation-Innovation Inventory using multiple-group mean and covariance structure analysis. *Multivariate Behavioral Research*. 2000; 35:169–199. http://dx.doi.org/10.1207/S15327906MBR3502_2. [PubMed: 26754082]
- Chavez LM, Shrout PE, Alegría M, Lapatin S, Canino G. Ethnic differences in perceived impairment and need for care. *Journal of Abnormal Child Psychology*. 2010; 38:1165–1177. <http://dx.doi.org/10.1007/s10802-010-9428-8>. [PubMed: 20521095]
- Chen KW, Killeya-Jones LA, Vega WA. Prevalence and co-occurrence of psychiatric symptom clusters in the U.S. adolescent population using DISC predictive scales. *Clinical Practice and*

- Epidemiology in Mental Health. 2005; 1:22. <http://dx.doi.org/10.1186/1745-0179-1-22>. [PubMed: 16255774]
- Dalsgaard S, Mortensen PB, Frydenberg M, Thomsen PH. Conduct problems, gender and adult psychiatric outcome of children with attention-deficit hyperactivity disorder. *The British Journal of Psychiatry*. 2002; 181:416–421. <http://dx.doi.org/10.1192/bjp.181.5.416>. [PubMed: 12411268]
- De Los Reyes A, Kazdin AE. Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*. 2005; 131:483–509. <http://dx.doi.org/10.1037/0033-2909.131.4.483>. [PubMed: 16060799]
- Dirks MA, De Los Reyes A, Briggs-Gowan M, Cella D, Wakschlag LS. Annual research review: Embracing not erasing contextual variability in children's behavior—Theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry*. 2012; 53:558–574. <http://dx.doi.org/10.1111/j.1469-7610.2012.02537.x>. [PubMed: 22360546]
- Eiraldi RB, Mazzuca LB, Clarke AT, Power TJ. Service Utilization among ethnic minority children with ADHD: A model of help-seeking behavior. *Administration and Policy in Mental Health and Mental Health Services Research*. 2006; 33:607–622. <http://dx.doi.org/10.1007/s10488-006-0063-1>. [PubMed: 16862391]
- Fergusson DM, Horwood LJ, Ridder EM. Show me the child at seven: The consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology and Psychiatry*. 2005; 46:837–849. <http://dx.doi.org/10.1111/j.1469-7610.2004.00387.x>. [PubMed: 16033632]
- Gelhorn H, Hartman C, Sakai J, Mikulich-Gilbertson S, Stallings M, Young S, Crowley T. An item response theory analysis of *DSM-IV* conduct disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2009; 48:42–50. <http://dx.doi.org/10.1097/CHI.0b013e31818b1c4e>. [PubMed: 19034046]
- Gitchel D, Turner R, Rumrill P. Differential Item Functioning in rehabilitation research. *Work: Journal of Prevention, Assessment & Rehabilitation*. 2010; 36:361–369.
- Gomez R. Testing gender differential item functioning for ordinal and binary scored parent rated ADHD symptoms. *Personality and Individual Differences*. 2007; 42:733–742. <http://dx.doi.org/10.1016/j.paid.2006.08.011>.
- Gomez R. Equivalency for father and mother ratings of the ADHD symptoms. *Journal of Abnormal Child Psychology*. 2010; 38:303–314. <http://dx.doi.org/10.1007/s10802-009-9370-9>. [PubMed: 19941051]
- Gomez R, Burns GL, Walsh JA. Parent ratings of the oppositional defiant disorder symptoms: Item response theory analyses of cross-national and cross-racial invariance. *Journal of Psychopathology and Behavioral Assessment*. 2008; 30:10–19. <http://dx.doi.org/10.1007/s10862-007-9071-z>.
- Gomez R, Vance A, Gomez A. Children's Depression Inventory: Invariance across children and adolescents with and without depressive disorders. *Psychological Assessment*. 2012; 24:1–10. <http://dx.doi.org/10.1037/a0024966>. [PubMed: 21859217]
- Grayson DA, Mackinnon A, Jorm AF, Creasey H, Broe GA. Item bias in the Center for Epidemiologic Studies Depression Scale: Effects of physical disorders and disability in an elderly community sample. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2000; 55:273–282. <http://dx.doi.org/10.1093/geronb/55.5.P273>.
- Harrell, E.; Langton, L.; Berzofsky, M.; Couzens, L.; Smiley-McDonald, H. Household poverty and nonfatal violent victimization, 2008-2012. U. S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics; Washington, DC: 2014. NCJ Report 248384
- Hillemeier MM, Foster EM, Heinrichs B, Heier B, the Conduct Problems Prevention Research Group. Racial differences in parental reports of attention-deficit/hyperactivity disorder behaviors. *Journal of Developmental and Behavioral Pediatrics*. 2007; 28:353–361. <http://dx.doi.org/10.1097/DBP.0b013e31811ff8b8>. [PubMed: 18049317]
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55. <http://dx.doi.org/10.1080/10705519909540118>.

- Kataoka SH, Zhang L, Wells KB. Unmet need for mental health care among U.S. children: Variation by ethnicity and insurance status. *The American Journal of Psychiatry*. 2002; 159:1548–1555. <http://dx.doi.org/10.1176/appi.ajp.159.9.1548>. [PubMed: 12202276]
- Kim ES, Yoon M, Lee T. Testing measurement invariance using MIMIC: Likelihood ratio test with a critical value adjustment. *Educational and Psychological Measurement*. 2012; 72:469–492. <http://dx.doi.org/10.1177/0013164411427395>.
- Kline, RB. *Principles and practice of structural equation modeling*. 3rd ed.. Guilford Press; New York, NY: 2011.
- Knight, GP.; Roosa, MW.; Umaña-Taylor, AJ. Studying ethnic minority and economically disadvantaged populations: Methodological challenges and best practices. *American Psychological Association*; Washington, DC: 2009. <http://dx.doi.org/10.1037/11887-000>
- Lahey BB, Loeber R, Hart EL, Frick PJ, Applegate B, Zhang Q, Russo MF. Four-year longitudinal study of conduct disorder in boys: Patterns and predictors of persistence. *Journal of Abnormal Psychology*. 1995; 104:83–93. <http://dx.doi.org/10.1037/0021-843X.104.1.83>. [PubMed: 7897057]
- Leung PW, Lucas CP, Hung SF, Kwong SL, Tang CP, Lee CC, Shaffer D. The test-retest reliability and screening efficiency of DISC Predictive Scales-version 4.32 (DPS-4.32) with Chinese children/youths. *European Child & Adolescent Psychiatry*. 2005; 14:461–465. <http://dx.doi.org/10.1007/s00787-005-0503-6>. [PubMed: 16341503]
- Loeber R, Burke JD, Lahey BB, Winters A, Zera M. Oppositional defiant and conduct disorder: A review of the past 10 years, part I. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2000; 39:1468–1484. <http://dx.doi.org/10.1097/00004583-200012000-00007>. [PubMed: 11128323]
- Lucas CP, Zhang H, Fisher PW, Shaffer D, Regier DA, Narrow WE, Friman P. The DISC Predictive Scales (DPS): Efficiently screening for diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2001; 40:443–449. <http://dx.doi.org/10.1097/00004583-200104000-00013>. [PubMed: 11314570]
- Marsh HW, Balla JR, McDonald RP. Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*. 1988; 103:391–410. <http://dx.doi.org/10.1037/0033-2909.103.3.391>.
- Maughan B, Rowe R, Messer J, Goodman R, Meltzer H. Conduct disorder and oppositional defiant disorder in a national sample: Developmental epidemiology. *Journal of Child Psychology and Psychiatry*. 2004; 45:609–621. <http://dx.doi.org/10.1111/j.1469-7610.2004.00250.x>. [PubMed: 15055379]
- McReynolds LS, Wasserman GA, Fisher P, Lucas CP. Diagnostic screening with incarcerated youths: Comparing the DPS and voice DISC. *Criminal Justice and Behavior*. 2007; 34:830–845. <http://dx.doi.org/10.1177/0093854807299918>.
- Mellenbergh GJ. Item bias and item response theory. *International Journal of Educational Research*. 1989; 13:127–143. [http://dx.doi.org/10.1016/0883-0355\(89\)90002-5](http://dx.doi.org/10.1016/0883-0355(89)90002-5).
- Muthén B, du Toit SHC, Spisic D. Robust inference using weighted least-squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. 1997 Unpublished technical report. Available for download from <http://www.statmodel.com>.
- Muthén BO, Hasin D, Wisnicki KS. Factor analysis of ICD-10 symptom items in the 1988 National Health Interview Survey on alcohol dependence. *Addiction*. 1993; 88:1071–1077. <http://dx.doi.org/10.1111/j.1360-0443.1993.tb02126.x>. [PubMed: 8401161]
- Muthén BO, Kao C, Burstein L. Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement*. 1991; 28:1–22. <http://dx.doi.org/10.1111/j.1745-3984.1991.tb00340.x>.
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. Muthén & Muthén; Los Angeles, CA: 1998-2001.
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. Seventh Edition. Muthén & Muthén; Los Angeles, CA: 1998-2012.
- Nock MK, Kazdin AE, Hiripi E, Kessler RC. Lifetime prevalence, correlates, and persistence of oppositional defiant disorder: Results from the National Comorbidity Survey Replication. *Journal of Child Psychology and Psychiatry*. 2007; 48:703–713. <http://dx.doi.org/10.1111/j.1469-7610.2007.01733.x>. [PubMed: 17593151]

- Ogg J, McMahan MM, Dedrick RF, Mendez LR. Middle school students' willingness to engage in activities with peers with ADHD symptoms: A multiple indicators multiple causes (MIMIC) model. *Journal of School Psychology*. 2013; 51:407–420. <http://dx.doi.org/10.1016/j.jsp.2013.01.002>. [PubMed: 23816232]
- Oort FJ. Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling*. 1998; 5:107–124. <http://dx.doi.org/10.1080/10705519809540095>.
- Roberts N, Stuart H, Lam M. High school mental health survey: Assessment of a mental health screen. *The Canadian Journal of Psychiatry / La Revue canadienne de psychiatrie*. 2008; 53:314–322.
- Roberts RE, Alegría M, Roberts CR, Chen IG. Mental health problems of adolescents as reported by their caregivers. *The Journal of Behavioral Health Services & Research*. 2005; 32:1–13. <http://dx.doi.org/10.1007/BF02287324>. [PubMed: 15632794]
- Rowland AS, Umbach DM, Stallone L, Naftel AJ, Bohlig EM, Sandler DP. Prevalence of medication treatment for attention deficit-hyperactivity disorder among elementary school children in Johnston County, NC. *American Journal of Public Health*. 2002; 92:231–234. <http://dx.doi.org/10.2105/AJPH.92.2.231>. [PubMed: 11818297]
- Rubio-Stipeck M, Shrout PE, Canino G, Bird HR, Jensen P, Dulcan M, Schwab-Stone M. Empirically defined symptom scales using the DISC 2.3. *Journal of Abnormal Child Psychology*. 1996; 24:67–83. <http://dx.doi.org/10.1007/BF01448374>. [PubMed: 8833029]
- Sandilands D, Oliveri ME, Zumbo BD, Ercikan K. Investigating sources of differential item functioning in international large-scale assessments using a confirmatory approach. *International Journal of Testing*. 2013; 13:152–174. <http://dx.doi.org/10.1080/15305058.2012.690140>.
- Schuster MA, Elliott MN, Kanouse DE, Wallander JL, Tortolero SR, Ratner JA, Banspach SW. Racial and ethnic health disparities among fifth-graders in three cities. *The New England Journal of Medicine*. 2012; 367:735–745. <http://dx.doi.org/10.1056/NEJMsa1114353>. [PubMed: 22913683]
- Shaffer D, Fisher P, Lucas CP, Dulcan MK, Schwab-Stone ME. NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2000; 39:28–38. <http://dx.doi.org/10.1097/00004583-200001000-00014>. [PubMed: 10638065]
- Sireci SG, Yang Y, Harter J, Ehrlich EJ. Evaluating guidelines for test adaptations: A methodological analysis of translation quality. *Journal of Cross-Cultural Psychology*. 2006; 37:557–567. <http://dx.doi.org/10.1177/0022022106290478>.
- Snowden LR, Yamada AM. Cultural differences in access to care. *Annual Review of Clinical Psychology*. 2005; 1:143–166. <http://dx.doi.org/10.1146/annurev.clinpsy.1.102803.143846>.
- Steiger JH. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*. 1990; 25:173–180. http://dx.doi.org/10.1207/s15327906mbr2502_4. [PubMed: 26794479]
- Triandis HC, Marín G, Lisansky J, Bentancourt H. Simpatía as a cultural script of Hispanics. *Journal of Personality and Social Psychology*. 1984; 47:1363–1375. <http://dx.doi.org/10.1037/0022-3514.47.6.1363>.
- Waschbusch DA. A meta-analytic examination of comorbid hyperactive-impulsive-attention problems and conduct problems. *Psychological Bulletin*. 2002; 128:118–150. <http://dx.doi.org/10.1037/0033-2909.128.1.118>. [PubMed: 11843545]
- Weisz, JR.; McCarthy, CA.; Eastman, KL.; Chaiyasit, W.; Sunwanlert, S. Developmental psychopathology and culture: Ten lessons from Thailand. In: Luthar, S.; Burack, J.; Cicchetti, D.; Weiner, J., editors. *Perspectives on adjustment, risk, and disorder*. Cambridge University Press; Developmental psychopathology New York: 1997. p. 568-592.
- Windle M, Grunbaum JA, Elliott M, Tortolero SR, Berry S, Gilliland J, Schuster M. Healthy passages. *American Journal of Preventive Medicine*. 2004; 27:164–172. <http://dx.doi.org/10.1016/j.amepre.2004.04.007>. [PubMed: 15261905]
- Woods CM. Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research*. 2009; 44:1–27. <http://dx.doi.org/10.1080/00273170802620121>. [PubMed: 26795105]

- Woods CM, Grimm KJ. Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement*. 2011; 35:339–361. <http://dx.doi.org/10.1177/0146621611405984>.
- Yeh M, McCabe K, Hough RL, Lau A, Fakhry F, Garland A. Why bother with beliefs? Examining relationships between race/ethnicity, parental beliefs about causes of child problems, and mental health service use. *Journal of Consulting and Clinical Psychology*. 2005; 73:800–807. <http://dx.doi.org/10.1037/0022-006X.73.5.800>. [PubMed: 16287380]
- Yu, C-Y. Unpublished doctoral dissertation. University of California; Los Angeles: 2002. Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. Available for download from <http://www.statmodel.com>
- Zumbo BD. Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*. 2007; 4:223–233. <http://dx.doi.org/10.1080/15434300701375832>.

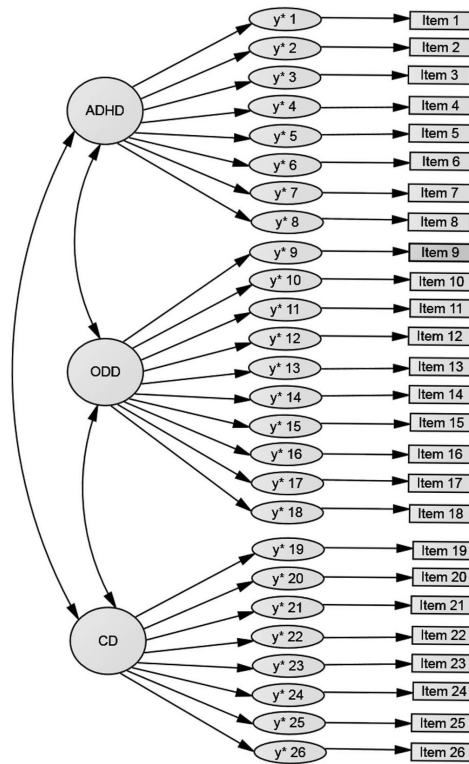


Figure 1. Hypothesized 3-factor CFA model. The y^* refer to the latent continuous response variables for the given DPS item. The abbreviated item label for each DPS item is shown in Table 2. The parameter estimates for the 3-factor CFA model are shown in Table 4.

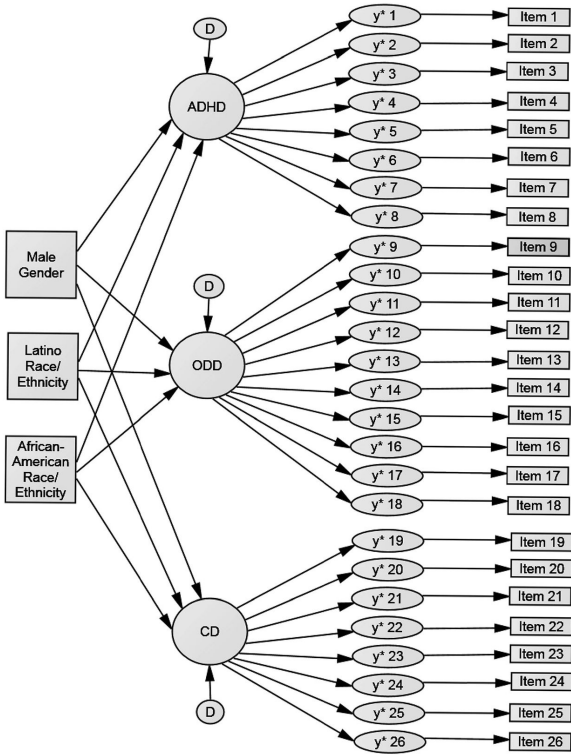


Figure 2. Baseline 3-factor MIMIC model (without DIF). The y^* refer to the latent continuous response variables for the given DPS item. The abbreviated item label for each DPS item is shown in Table 2. “D” refers to the disturbance of the latent factor.

Table 1
Sociodemographic Characteristics of the Total Analytic Sample and by Racial/Ethnic Subgroup

Variable	Total analytic sample (<i>N</i> = 4,491)	African-American (<i>N</i> = 1,646)	Latino (<i>N</i> = 1,666)	White (<i>N</i> = 1,179)
Child race/ethnicity				
African-American	1,646(30.6)			
Latino	1,666(45.9)			
White	1,179 (23.4)			
Child male gender	2,202(51.0)	779 (51.4)	823 (49.4)	600 (53.6)
Child age (years) ^a	12.60 (0.66)	12.63 (.74)	12.60 (.62)	12.58 (.62)
PCG female gender	4,153 (93.1)	1,552(94.9)	1,542(93.2)	1,059 (90.7)
PCG currently married/living with partner (yes)	2,770 (64.8)	656 (41.5)	1,182 (72.1)	932 (53.8)
PCG working part-time or full-time (yes)	3,179 (70.1)	1,220(74.9)	1,078 (64.3)	881 (75.4)
PCG highest educational attainment				
Not graduated from high school	845 (23.8)	147 (9.7)	678 (44.0)	20 (2.6)
GED/high school degree	919 (22.2)	446 (30.3)	394 (24.0)	79 (8.3)
Some years of college education	2,660 (53.9)	1,016(60.0)	577 (32.0)	1,067 (89.1)
PCG age (years) ^a	40.19 (7.30)	39.40 (8.93)	39.09 (6.07)	43.39 (6.48)

Note. Unless otherwise indicated, the total unweighted number (weighted percentage in parentheses) is shown. Percentages by racial/ethnic subgroup are weighted within-group percentages. Percentages may not add up to 100% due to rounding.

^aWeighted mean (weighted standard deviation in parentheses).

Table 2
Descriptive Information for DPS Items (N = 4,491)

Item	Prevalence (% yes)	N (yes)	I-S <i>r</i>	KR20- subscale
ADHD subscale				.76
1. Trouble finishing homework	37.8	1,697	.52	
2. Not listening to people	34.8	1,562	.50	
3. Taking medication for hyperactivity	7.5	338	.34	
4. Forgetting what s/he planned to do	27.4	1,231	.53	
5. Difficulty to keep mind on task	32.2	1,446	.57	
6. Often getting up from seat	18.2	818	.44	
7. Making a lot of easy mistakes	17.3	779	.49	
8. Talking much more than other children	30.6	1,373	.32	
ODD subscale				.81
9. Refused to do what s/he was told to do	23.0	1,034	.50	
10. Grouchy or easily annoyed	45.7	2,050	.54	
11. Mad at people/about things	36.3	1,632	.53	
12. Got even with others	10.5	473	.44	
13. Cursed/used dirty language	29.7	1,334	.46	
14. Mean on purpose	10.8	485	.46	
15. Did forbidden things on purpose	17.9	803	.53	
16. Lost temper	59.3	2,664	.46	
17. Blamed others for own mistakes	35.9	1,612	.48	
18. Argued or talked back	49.4	2,218	.53	
CD subscale				.69
19. Bullied someone	7.8	352	.39	
20. Tried/been physically cruel to someone	1.9	86	.40	
21. Lied to get something s/he wanted	9.6	433	.38	
22. Broke something on purpose	2.3	103	.38	
23. Been physically cruel to animal	0.8	37	.33	
24. Expelled from school	3.2	143	.34	
25. Been in severe physical fight	8.5	380	.36	
26. Stole from those s/he lives with	4.6	207	.42	

Note. The full wording of items can be obtained from C.P. Lucas. Summary statistics (except coefficient of reliability) adjusted for sample weights and clusters. ADHD = attention deficit/ hyperactivity disorder; CD = conduct disorder symptoms; ODD = oppositional defiant disorder symptoms; Prevalence (yes) = prevalence of PCGs (in %) indicating that the given symptom had occurred for his/her child; I-S *r* = item-subscale correlation; KR20 = Kuder-Richardson 20 coefficient of reliability.

Table 3
Fit Statistics for Estimated CFA and MIMIC Models for DPS Items (N = 4,491)

	WLSMV χ^2	df	CFI	RMSEA (90% CI)	WRMR
Confirmatory factor analysis (CFA) models					
M1: Null model	24821.85	325		.130 [.128, .131]	14.298
M2: One-factor CFA model	2304.38	299	.918	.039 [.037, .040]	3.254
M3: Three-factor CFA model ^a	1329.33	296	.958	.028 [.026, .029]	2.370
Multiple indicators multiple causes (MIMIC) models					
M4: Baseline three-factor MIMIC model (without DIF) ^b	1723.11	365	.942	.029 [.027, .030]	2.520
M5: Final three-factor MIMIC model (with DIF) ^c	1285.59	344	.960	.025 [.023, .026]	2.112

Note. All models were adjusted for sample weights and clustering. The estimated tetrachoric correlations among the 26 latent continuous response variables y^* ranged from .18 to .82 according to the three-factor CFA model. WLSMV χ^2 = robust weighted least squares chi-square test statistic; CFI = comparative fit index; RMSEA = root mean squared error of approximation; WRMR = weighted root mean square residual; DIF = differential item functioning; CI = confidence interval.

^aModel fit and substantive findings were nearly identical when the three-factor CFA model was re-estimated using data from only the female primary caregivers ($N = 4,153$, WLSMV $\chi^2(296) = 1275.24$, CFI = .955, RMSEA = .028, RMSEA 90% CI = .027, .030, WRMR = 2.313) and from only the English-language version ($N = 3,400$, WLSMV $\chi^2(296) = 1132.53$, CFI = .955, RMSEA = .029, RMSEA 90% CI = .027, .031, WRMR = 2.240), respectively.

^bModel fit and substantive findings were nearly identical when the baseline three-factor MIMIC model was re-estimated using data from only the female primary caregivers ($N = 4,153$, WLSMV $\chi^2(365) = 1659.56$, CFI = .938, RMSEA = .029, RMSEA 90% CI = .028, .031, WRMR = 2.454). Model fit and the vast majority of substantive findings also were closely replicated when this model was re-estimated using data from only the English-language version ($N = 3,400$, WLSMV $\chi^2(365) = 1418.24$, CFI = .942, RMSEA = .029, RMSEA 90% CI = .028, .031, WRMR = 2.315).

^cModel fit and substantive findings were nearly identical when the final three-factor MIMIC model was re-estimated using data from only the female primary caregivers ($N = 4,153$, WLSMV $\chi^2(344) = 1244.28$, CFI = .957, RMSEA = .025, RMSEA 90% CI = .024, .027, WRMR = 2.061). Model fit and the vast majority of substantive findings also were closely replicated when this model was re-estimated using data from only the English-language version ($N = 3,400$, WLSMV $\chi^2(344) = 1130.89$, CFI = .957, RMSEA = .026, RMSEA 90% CI = .024, .028, WRMR = 2.015).

Table 4
Parameter Estimates for Three-Factor CFA Model (N = 4,491)

Item	Unstandardized factor loading (SE)	Completely standardized factor loading	Thresh-hold	Proportion explained variance
ADHD latent factor				
1. Trouble finishing homework	1.00	.74	.31	.55
2. Not listening to people	.97 ^{***} (.03)	.72	.39	.51
3. Taking medication for hyperactivity	.88 ^{***} (.04)	.66	1.44	.43
4. Forgetting what s/he planned to do	.98 ^{***} (.03)	.73	.60	.53
5. Difficulty to keep mind on task	1.08 ^{***} (.03)	.80	.46	.65
6. Often getting up from seat	.99 ^{***} (.03)	.74	.91	.54
7. Making a lot of easy mistakes	1.04 ^{***} (.03)	.77	.94	.60
8. Talking much more than other children	.70 ^{***} (.04)	.52	.51	.27
ODD latent factor				
9. Refused to do what s/he was told to do	1.00	.80	.74	.64
10. Grouchy or easily annoyed	.95 ^{***} (.02)	.76	.11	.58
11. Mad at people/about things	.96 ^{***} (.03)	.77	.35	.59
12. Got even with others	1.01 ^{***} (.03)	.81	1.25	.65
13. Cursed/used dirty language	.81 ^{***} (.03)	.65	.53	.42
14. Mean on purpose	1.03 ^{***} (.03)	.82	1.24	.68
15. Did forbidden things on purpose	1.03 ^{***} (.03)	.83	.92	.68
16. Lost temper	.80 ^{***} (.02)	.64	-.24	.41
17. Blamed others for own mistakes	.85 ^{***} (.03)	.68	.36	.46
18. Argued or talked back	.88 ^{***} (.02)	.70	.02	.50
CD latent factor				
19. Bullied someone	1.00	.79	1.42	.62
20. Tried/been physically cruel to someone	1.10 ^{***} (.05)	.86	2.07	.75
21. Lied to get something s/he wanted	1.01 ^{***} (.04)	.79	1.30	.63
22. Broke something on purpose	1.00 ^{***} (.05)	.78	2.00	.61
23. Been physically cruel to animal	1.01 ^{***} (.07)	.79	2.40	.63
24. Expelled from school	.86 ^{***} (.04)	.68	1.86	.46
25. Been in severe physical fight	.84 ^{***} (.04)	.66	1.38	.44
26. Stole from those s/he lives with	1.00 ^{***} (.05)	.79	1.69	.62

Note. Unstandardized factor loadings for Items 1, 9, and 19 were fixed to the value “1” for identification purposes. Thresholds indicate the point on the latent response variable y^* where $y = 1$ if the threshold is exceeded (and where $y = 0$ if the threshold is not exceeded). The proportion of explained variance refers to the proportion of variance in the latent continuous response variables y^* that is explained by the latent factor. Estimated variances of the latent factors were .55 for ADHD, .64 for ODD, .62 for CD. Estimated covariances among the latent factors were .41 (ADHD with ODD), .39 (ADHD with CD), and .53 (ODD with CD).

^{***}
 $p < .001$ (two-tailed).

Table 5
Parameter Estimates for Direct Predictive Effects of Male Gender, Latino Race/Ethnicity, and African-American Race/Ethnicity on DPS Items With DIF From Final Three-Factor Model (N = 4,491)

Item	Unstandardized <i>b</i>	<i>SE</i>	Completely standardized β
Direct predictive effects of male gender on DPS item			
Item 9: Refused to do what s/he was told to do	.15***	.04	.08
Item 10: Grouchy or easily annoyed	-.13**	.04	-.06
Item 11: Mad at people/about things	-.16**	.05	-.08
Item 15: Did forbidden things on purpose	.16***	.04	.08
Direct predictive effects of Latino race/ethnicity on DPS item			
Item 3: Taking medication for hyperactivity	-.80***	.08	-.38
Item 5: Difficulty to keep mind on task	-.27***	.05	-.13
Item 6: Often getting up from seat	.34***	.07	.16
Item 9: Refused to do what s/he was told to do	.38***	.07	.19
Item 11: Mad at people/about things	.19***	.05	.10
Item 12: Got even with others	.22**	.08	.11
Item 16: Lost temper	-.40***	.05	-.20
Item 17: Blamed others for own mistakes	-.46***	.07	-.22
Item 18: Argued or talked back	-.75***	.05	-.35
Direct predictive effects of African-American race/ethnicity on DPS item			
Item 3: Taking medication for hyperactivity	-.48***	.08	-.21
Item 6: Often getting up from seat	.54***	.08	.24
Item 9: Refused to do what s/he was told to do	.51***	.06	.23
Item 12: Got even with others	.31***	.07	.14
Item 14: Mean on purpose	.32***	.06	.15
Item 15: Did forbidden things on purpose	.41***	.07	.18
Item 16: Lost temper	-.41***	.05	-.18
Item 18: Argued or talked back	-.58***	.05	-.25

Note. Reference category for race/ethnicity was White.

**
 $p < .01$ (two-tailed).

 $p < .001$.

Table 6
Sensitivity Analysis of Relations Between Covariates and DPS Latent Factors (N = 4,491)

Covariate	Baseline three-factor MIMIC model (without DIF)			Final three-factor MIMIC model (with DIF)		
	Unstandardized Coefficient (95% CI)	SE	p	Unstandardized Coefficient (95% CI)	SE	p
ADHD latent factor						
Male gender	.35 [.27, .44]	.05	.000	.35 [.27, .44]	.05	.000
Latino	-.07 [-.19, .05]	.06	.223	.05 [-.07, .18]	.06	.394
African-American	.30 [.18, .42]	.06	.000	.33 [.20, .45]	.07	.000
ODD latent factor						
Male gender	.15 [.08, .22]	.04	.000	.15 [.07, .23]	.04	.000
Latino	-.39 [-.51, -.28]	.06	.000	-.25 [-.38, -.12]	.07	.000
African-American	-.21 [-.31, -.10]	.05	.000	-.20 [-.32, -.09]	.06	.001
CD latent factor						
Male gender	.30 [.21, .40]	.05	.000	.30 [.21, .40]	.05	.000
Latino	-.15 [-.36, .06]	.11	.172	-.15 [-.35, .06]	.11	.172
African-American	.23 [.04, .42]	.10	.020	.23 [.04, .42]	.10	.019

*Note.*All estimates were adjusted for sample weights and clustering. Reference category for race/ethnicity was White. ADHD = attention deficit/hyperactivity disorder; CD = conduct disorder symptoms; ODD = oppositional defiant disorder symptoms; DIF = differential item functioning; CI = confidence interval.