

Parent and youth report of youth anxiety: evidence for measurement invariance

Melanie A. Dirks,¹ V. Robin Weersing,² Erin Warnick,³ Araceli Gonzalez,⁴ Megan Alton,¹ Christine Dauser,³ Lawrence Scahill,³ and Joseph Woolston³

¹Department of Psychology, McGill University, Montreal, QC, Canada; ²Joint Doctoral Program in Clinical Psychology, San Diego State University/University of California, San Diego, CA, USA; ³Yale Child Study Center, Yale University, New Haven, CT, USA; ⁴Department of Psychiatry, University of California, Los Angeles, CA, USA

Background: We characterized parent-youth disagreement in their report on the Screen for Child Anxiety Related Emotional Disorders (SCARED) and examined the equivalence of this measure across parent and youth report. **Methods:** A clinically referred sample of 408 parent-youth dyads (M age youth = 14.33, SD = 1.89; 53.7% male; 50.0% Non-Hispanic White (NHW), 14.0% Hispanic, 29.7% African-American) completed the SCARED. We examined (a) differences between parents and youth in the total number of symptoms reported (difference scores) and in their ratings of specific symptoms (q correlations), (b) demographic factors associated with these indices, and (c) equivalence of the pattern and magnitude of factor loadings (i.e., configural and metric invariance), as well as item thresholds and residual variances, across informants. **Results:** The mean difference score was -2.13 (SD = 14.44), with youth reporting higher levels of symptoms, and the mean q correlation was $.32$ (SD = $.24$). Difference scores were greater for African-American dyads than NHW pairs. We found complete configural, metric, and residual invariance, and partial threshold invariance. Differences in thresholds did not appear to reflect systematic differences between parent and youth report. Findings were comparable when analyses were conducted separately for NHW and ethnic minority families. **Conclusion:** Findings provide further evidence for the importance of considering youth report when evaluating anxiety in African-American families. The SCARED was invariant across informant reports, suggesting that it is appropriate to compare mean scores for these raters and that variability in parent and youth report is not attributable to their rating different constructs or using different thresholds to determine when symptoms are present. **Keywords:** Anxiety, measurement, informant disagreement.

Introduction

Best practice in the assessment of youth anxiety is to obtain reports from children and their parents (Silverman & Ollendick, 2008). Agreement between these two informants, however, is only low to moderate, a pattern that is well-documented but poorly understood (De Los Reyes, 2011). One underutilized method of examining informant disagreement is to assess the equivalence of measures across informants. Given that informants have access to different behavioral and emotional samples, as well as different frameworks and experiences that inform their interpretation of a given action or feeling (Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012), it is possible that they will use rating scales in different ways. In this study, we characterized disagreement between parent and youth (self) report on a widely used rating scale for child anxiety – the Screen for Child Anxiety Related Emotional Disorders (SCARED) – and then tested the invariance of this instrument across youth and parent report. Examining cross-informant equivalence will help us to understand why reports by different informants about the same child vary so markedly. Unpacking this disagreement will provide valuable information

for clinicians seeking to integrate reports from multiple sources (Cole, Hoffman, Tram, & Maxwell, 2000).

A common way to examine interinformant agreement is to calculate the correlations between the two groups of raters. Estimates of the correlation between parent and youth report on the SCARED range from approximately $.30$ to $.60$ (Birmaher et al., 1997, 1999; Wren, Bridge, & Birmaher, 2004; Wren et al., 2007), which is comparable to findings for other anxiety rating scales (e.g., Baldwin & Dadds, 2007; Krain & Kendall, 2000). Correlations reflect the relative ranking of two informants' ratings, but do not actually capture dyadic agreement (i.e., a parent and child might both rate the child as being highly anxious, but still show considerable disagreement in their ratings; Carlston & Ogles, 2009). Examining predictors of the disagreement within dyads is important clinically, because this information is used in diagnostic and treatment decisions. Parents and children can differ in their judgment of the overall number of problems present, as well as in their ratings of specific problems (Youngstrom, Loeber, & Stouthamer-Loeber, 2000). The former is typically operationalized using difference scores, which are calculated by subtracting the score of one informant from the other, whereas the latter can be captured using q correlations, the Pearson correlation between the items provided by two raters (Lau et al., 2004; Youngstrom et al., 2000).

Conflicts of interests statement: No conflicts declared.

Previous research has suggested that parent-youth difference scores in their report of youth's internalizing symptoms can be influenced by a number of factors including the child's gender (e.g., Berger, Jodl, Allen, McElhaney, & Kuperminc, 2005; Carlston & Ogles, 2009) and age (e.g., Berg-Nielsen, Vika, & Dahl, 2003; Stevanovic, Jancic, Topalovic, & Tadic, 2012), although not all studies find these associations (e.g., van de Looij-Jansen, Jansen, de Wilde, Donker, & Verhulst, 2011; Treutler & Epkins, 2003). A handful of studies have also examined the associations between children's ethnicity and parent-youth difference scores in their report of internalizing symptoms. Youngstrom et al. (2000) reported no difference across African-American and NHW dyads, but work conducted with a clinically referred sample found that NHW dyads had lower difference scores than African-American, Hispanic, and Asian-Pacific Islander pairings (Lau et al., 2004), indicating greater disagreement between parents and youth among minority families. These studies found that ethnicity was not associated with *q* correlations, and they did not examine whether *q* correlations varied as a function of age and gender.

In addition to characterizing parent-youth disagreement on the SCARED, we also sought to understand it by examining the invariance of this instrument across parent and youth report. Measurement invariance is the extent to which an instrument captures the same construct(s) in different groups (Dimitrov, 2010). Finding that a measure of youth psychopathology is invariant across parent and youth report means that both informants are interpreting the individual items and the underlying latent construct(s) comparably (van de Schoot, Lugtig, & Hox, 2012). Establishing the invariance of the instrument across raters is necessary for the meaningful comparison of their scores (Vandenberg & Lance, 2000). To conclude that parents and youth differ in the level of youth anxiety they report, we must have solid evidence that the measure is performing equivalently for both informants (Gomez, 2007).

Examining measurement equivalence also provides an opportunity to understand more fully how and why informants' reports differ (Cole et al., 2000; Gomez, 2007). Given mounting evidence that the disagreement between informants in their report of children's psychopathology reflects, at least partly, meaningful differences in the frameworks they use to interpret children's behavior and emotions (Dirks et al., 2012), it is possible that informants may actually be evaluating different constructs, or using different thresholds to determine when a behavior is clinically concerning. These hypotheses can be examined using formal tests for measurement invariance. Despite the psychometric importance and theoretical promise of such analyses, however, few studies have examined the cross-informant invariance of measures of youth psychopathology

(e.g., Gomez, 2007; Sanne, Torsheim, Heiervang, & Stormark, 2009; Waschbusch & Willoughby, 2008).

In summary, the goal of this article was to characterize and elucidate differences in parent and youth report of youth anxiety, building on existing work in four key ways. First, we examined agreement between parent and youth report on the SCARED. Most previous work has focused on internalizing symptoms, broadly. Understanding differences in parent and youth reports of anxiety, specifically, is important given that anxiety disorders are the most common psychiatric condition among youth, and they occur frequently in the absence of comorbid mood symptoms (Merikangas et al., 2010). Second, we examined predictors of two complementary indices of informant disagreement – difference scores and *q* correlations – concerning anxiety symptoms. Third, our ethnically diverse sample allowed us to develop the limited evidence base concerning ethnic differences in informant agreement. Fourth, we tested measurement invariance across parent and youth report. This analysis will provide additional information concerning psychometric properties of the SCARED, and advance understanding of why these two informants provide such disparate reports of youth's symptomatology.

Methods

Participants and procedures

All procedures were approved by the Research Ethics Board of the relevant institutions. Participants were drawn from families seeking services at an outpatient mental health clinic in the Northeastern United States between July 2004 and June 2009. All families complete a standardized battery of measures, including the SCARED, at intake; however, only families of children age 11 years and older could be included in our analyses because younger children do not complete self-report questionnaires. During this time frame, 849 eligible families participated in an intake. Of these, 658 (77.5%) were invited to consent to participate in research. To increase the generalizability of the sample, exclusion criteria were minimal and included the youth being in the custody of the Department of Children and Families, a clinical emergency that precluded obtaining research consent, and, prior to 2007, parents or guardians who were not English speaking. (From 2007 to 2009, Spanish consent procedures were in place provided there was a bilingual clinician present.) Of the families invited to participate, 565 (85.9%) provided written consent. There were no demographic differences between families who did and did not provide consent. The present study required the SCARED be completed by both parent and youth, data that were available for 408 families (62.0%). Missing data were typically due to the child not completing the SCARED because of intellectual

or pervasive developmental delays, or parents or children not completing the measures due to difficulties with literacy.

Demographic characteristics of our final analytic sample ($N = 408$ parent-child dyads) were as follows: *M* age youth = 14.33 years ($SD = 1.89$); 53.7% male; 50.0% Non-Hispanic White (NHW), 14.0% Hispanic, 29.7% African-American; 85.9% of families receiving Medicaid. Our analytic sample did not differ from the research-eligible sample ($N = 658$) in terms of age, gender, or type of insurance coverage at intake. The analytic sample did differ in terms of ethnicity, $\chi^2(3) = 18.72$, $p < .01$, with proportionally more NHW families in the final sample. The sample was diagnostically heterogeneous, with the most common diagnoses, based on review of medical records, including anxiety disorders (27.5%), depressive disorders (23.0%), and any disruptive behavior disorder (30.1%).

Measures

Demographic and clinical data were collected from the youth's primary caregiver (the mother in 79.7% of cases) during a standardized assessment at intake. Parents reported youth age, gender, ethnicity, and insurance coverage (Medicaid vs. other). The Screen for Child Anxiety Related Emotional Disorders, Parent and Child Report (SCARED-P/C) is a 41-item instrument designed to assess anxiety among clinically referred samples (Birmaher et al., 1999). Each item on the SCARED asks the rater to identify how often the youth experiences each symptom on a three point scale: 0 (*Not True or Hardly Ever True*), 1 (*Somewhat or Sometimes True*), or 2 (*Very True or Often True*). The items on the SCARED-P/C are identical, except for the stem 'you/your child.' Sixteen parents (4.0%; 28.1% of Hispanic parents) elected to complete the Spanish translation of the SCARED. Reliability of the SCARED-P and -C in this sample were both excellent, with $\alpha = .94$ for each.

Unlike many other rating scales of youth anxiety, which measure anxiety globally, the SCARED-P/C assesses five facets of anxiety – general anxiety, somatic/panic symptoms, separation anxiety, social phobia, and school phobia (Birmaher et al., 1997). The SCARED has strong psychometric properties, including good test-retest reliability, internal consistency, and discriminant and convergent validity (Birmaher et al., 1997, 1999). For these reasons, it is widely used by clinicians and researchers, making it particularly important to examine the equivalence of the parent and youth forms.

Data analysis

We characterized agreement between parents and their children by (a) calculating difference scores by subtracting total scores on the SCARED-C from total scores on the SCARED-P and (b) computing q

correlations by obtaining the Pearson correlation across the SCARED items for each dyad. Inspection of these variables revealed that they were distributed relatively normally. We then constructed two linear-regression models in which agreement indices served as the dependent variables and the following demographic variables were predictors: age, gender, ethnicity (dummy coded as African-American, Hispanic, and other, with NHW as the reference category), and receipt of Medicaid, which was used as a proxy for socioeconomic status (SES).

Next, we followed the steps outlined by Vandenberg and Lance (2000) to examine the equivalence of the SCARED across informants (also see van de Schoot et al., 2012). All analyses were conducted in MPlus 6.0 (Muthén & Muthén, 2010). The SCARED requires informants to respond using three discrete categories; thus, we treated items as ordered categories and used the mean- and variance-adjusted weighted least squares (WLSMV) estimator and theta parameterization (Gomez, Vance, & Gomez, 2012). To account for the nonindependence arising from the collection of data from parents and their children, we conducted the analyses within a repeated-measures framework (i.e., parent and youth report were treated as two observations on the same participant).

First, we examined whether the same number of factors and pattern of factor loadings characterized report by each informant (configural invariance), using confirmatory factor analysis (CFA) to determine whether the established five-factor model of the SCARED (somatic/panic, generalized anxiety, separation anxiety, school anxiety, and social anxiety; Birmaher et al., 1997, 1999) provided adequate fit to parent and youth data. For purposes of model identification, it was necessary to constrain factor means to 0, factor variances to 1, and residual variances to 1 for both parents and youth (Muthén & Asparouhov, 2002). Good model fit is often indicated by a nonsignificant χ^2 -test. Simulation work has indicated that when large models are fit in relatively small samples, χ^2 values may be inflated (Flora & Curran, 2004), thus we also considered the Comparative Fit Index (CFI) and the Root Mean-Square Error of Approximation (RMSEA). A CFI exceeding .95 and a RMSEA with a lower-bound confidence interval overlapping .06 indicate good fit (Hu & Bentler, 1999).

Second, we assessed metric invariance, that is, the equality of the magnitude of the factor loadings. In this model, all factor loadings were constrained to be equal across parents and youth. Thresholds were allowed to vary across groups. Factor means were still estimated at 0 and residual variances were constrained to 1. Factor variances were constrained at 1 for parent-reported items and allowed to vary for youth report. The χ^2 -difference test was used to determine whether constraining the factor loadings resulted in a significant reduction in model fit, with a nonsignificant test indicating invariance.

Next, we tested threshold invariance. Thresholds are the levels of the latent variables at which the score on the item changes (Flora & Curran, 2004). If thresholds are equivalent, the same level of the underlying variable will translate into the same score on a given item across informants. Here, all factor loadings and all thresholds (two per item, because there are three response categories) were constrained to be equal across informants. Factor means and variances were constrained to 0 and 1, respectively, for parent-reported items, but allowed to vary for youth report, and residual variances were constrained to 1 for both informants. Our final step was to test residual, or error variance, invariance, by comparing a model in which the residuals are constrained to 1 for parent report and allowed to vary for youth report to one in which all residuals are constrained to equal 1.

Results

The mean difference score was -2.13 ($SD = 14.44$; range = $-43.00, 48.00$), with youth reporting more symptoms than parents. The mean q correlation was $.32$ ($SD .24$; range = $-.31$ to 1.00). The results of the regression models examining associations between demographic predictors and indices of informant disagreement are presented in Table 1. Ethnicity predicted difference scores, with disagreement between youth and parent report being greater for African-American youth ($M = -4.30, SD = 13.30$) than for NHW participants ($M = -0.29, SD = 14.99$). Age was a predictor of q correlations, with agreement between youth and parent report increasing with age.

Having characterized the agreement of informants, we tested measurement invariance. Separate CFAs of the five-factor model of the SCARED run on parent and youth data yielded the following fit indices for

Table 1 Standardized regression coefficients linking demographic characteristics to indices of parent-youth agreement on the Screen for Anxiety Related Emotional Disorders (SCARED)

	Difference score	Q correlation
Gender ^a	-.08	.08
Age	-.09*	.15**
Ethnicity ^b		
African-American versus NHW	-.15**	-.12*
Hispanic versus NHW	-.10*	-.06
Other Ethnicity versus NHW	-.08	-.05
Medicaid status ^c	.04	.03

Difference scores were calculated such that negative scores indicated youth reported more symptoms than parents. NHW, non-Hispanic white.

^aGender dummy code with male = 0.

^bEthnicity dummy coded into three variables so that NHW was the reference category.

^cDummy coded with receipt of Medicaid = 0.

* $p < .10$; ** $p < .01$.

youth, $\chi^2(769) = 1225.67, p < .01$; RMSEA = .038 (90% CI = .034–.042), CFI = .96; and parents, $\chi^2(769) = 1410.26, p < .01$; RMSEA = .045 (90% CI = .041–.049), CFI = .95. Although the χ^2 was significant, both the RMSEA and CFI indicated good fit; thus, we accepted this model.

We then proceeded through the steps of testing measure equivalence. All fit indices and statistical tests are presented in Table 2. We found evidence for full configural and metric invariance. The threshold-invariance model provided significantly worse fit to the data than the metric-invariance model. Review of the modification indices suggested allowing 22 thresholds to vary (see Table 2). The resulting partial threshold-invariance model provided comparable fit to the metric-invariance model. We also found evidence for complete residual invariance.

As it was computationally necessary to free 22 thresholds across groups, we computed a series of χ^2 tests to examine whether the pattern of freed thresholds represented systematic differences in parent and youth report. Overall, parents and youth were equally likely to have higher thresholds, $\chi^2(1) = .72, p > .05$, and there was no association between symptom type and which informant had the higher threshold, $\chi^2(4) = 2.25, p > .05$ (For the pattern of findings, see Table S1).

Given that disagreement was greater among African-American than NHW pairs, it is possible that there may be greater invariance among this subset of the sample. The sample size was not sufficient to examine this question with just African-American families ($n = 121$). Although differences were not statistically significant, both Hispanic families ($n = 57$) and families who identified as 'other' ethnicities ($n = 25$), also demonstrated greater disagreement than NHW families. Thus, we combined these participants with the African-American dyads to form a minority-status group ($n = 203$) and we examined measurement invariance separately in this group and the NHW group ($n = 204$; note that one family did not identify their ethnicity). Results are presented in Table 2. Findings mirrored that in the full sample, with evidence for complete configural, metric, and residual invariance, and partial threshold invariance in both groups. We again used χ^2 tests to examine the pattern of freed thresholds and found that, in both groups, parents and youth were equally likely to have higher thresholds, $\chi^2(1) = 0$ (minority) and 0.29 (NHW), $ps > .05$ and there was no association between symptom type and which informant had the higher threshold, $\chi^2(4) = 4.59$ (minority) and 8.27 (NHW), $ps > .05$.

Discussion

In this study, we characterized disagreement between parent and youth report of youth anxiety symptoms on the SCARED and examined the measurement invariance of the SCARED across these

Table 2 Fit of models testing measurement invariance of the Screen for Anxiety and Related Emotional Disorders (SCARED) across parent and youth report

Models	Model Fit					Model Difference			
	χ^2	df	RMSEA (90% CI)	CFI	TLI	ΔM	Δdf	$\Delta \chi^2$	
Full sample (<i>N</i> = 408)	M1: Configural invariance	4120.47	3194	.027 (.024–.029)	.95	.95	–	–	–
	M2: Metric invariance	4091.14	3230	.026 (.023–.028)	.95	.95	M2-M1	36	49.08
	M3: Threshold invariance	4207.48	3307	.026 (.023–.028)	.95	.95	M3-M2	77	250.42*
	M3a: Partial threshold invariance ^a	4146.33	3285	.025 (.023–.028)	.95	.95	M3a-M2	55	69.06
Minority Families (<i>n</i> = 203)	M4: Residual variance invariance	4159.28	3244	.026 (.024–.029)	.95	.94	M4-M3a	41	56.12
	M1: Configural invariance	3492.01	3194	.021 (.016–.026)	.96	.96	–	–	–
	M2: Metric invariance	3516.18	3230	.021 (.015–.026)	.96	.95	M2-M1	36	45.49
	M3: Threshold invariance	3608.58	3307	.021 (.015–.026)	.96	.96	M3-M2	77	162.66*
NHW families (<i>n</i> = 204)	M3a: Partial threshold invariance ^b	3585.05	3297	.021 (.015–.026)	.97	.97	M3a-M2	67	83.28
	M4: Residual variance invariance	3547.29	3256	.021 (.015–.026)	.97	.97	M4-M3a	41	55.10
	M1: Configural invariance	3708.23	3194	.028 (.024–.032)	.94	.94	–	–	–
	M2: Metric invariance	3708.65	3230	.027 (.022–.031)	.95	.94	M2-M1	36	42.12
	M3: Threshold invariance	3802.21	3307	.027 (.023–.031)	.94	.94	M3-M2	77	171.72*
	M3a: Partial threshold invariance ^c	3772.71	3293	.027 (.022–.031)	.95	.95	M3a-M2	63	74.65
	M4: Residual variance invariance	3761.09	3252	.028 (.023–.032)	.94	.94	M4-M3a	41	48.41

RMSEA, root mean-square error of approximation; CFI, Comparative Fit Index; TLI, Tucker-Lewis Index; χ^2 = mean- and variance-adjusted weighted least squares (WLSMV) chi-square.

Minority families were those who identified as African-American, Hispanic, and 'other ethnicity.' One family did not identify their ethnicity.

$\Delta \chi^2$ was calculated using the difference-test function in MPlus 6.0, because the difference between two nested models is not distributed as chi-square when WLSMV chi-square values are used (Muthén & Muthén, 2010).

**p* < .05.

^{a-c}Model M3a was identical to model M3, except that the following thresholds (identified by modification indices) were allowed to vary across informants:

^aGeneral-anxiety factor, item 5, thresholds 1 and 2, item 14, threshold 1, item 23, threshold 1, item 28, threshold 1, item 33, thresholds 1 and 2, item 35, threshold 2; separation-anxiety factor, item 9, threshold 2, item 13, threshold 2, item 20, threshold 2; somatic/panic factor, item 18, thresholds 1 and 2, item 19, threshold 1, item 24, threshold 1, item 30, threshold 1, item 34, threshold 1; social-phobia factor, item 10, threshold 2, item 32, threshold 2, item 41, threshold 2; school-phobia factor, item 2, threshold 1, item 17, threshold 2;

^bGeneral-anxiety factor, item 5, thresholds 1 and 2, item 23, threshold 1, item 33, threshold 2, item 35, threshold 2; separation-anxiety factor, item 13, threshold 1; somatic/panic factor, item 18, threshold 2, item 19, threshold 2, item 24, threshold 1; social-phobia factor, item 32, threshold 2.

^cGeneral-anxiety factor, item 5, thresholds 1 and 2; separation-anxiety factor, item 8, thresholds 1 and 2, item 29, threshold 1; somatic/panic factor, item 9, thresholds 1 and 2, item 15, thresholds 1 and 2, item 18, threshold 1, item 19, threshold 1, item 34, threshold 1; school-phobia factor, item 2, threshold 1, item 36, threshold 2.

two informants. On average, youth reported higher levels of anxiety symptoms than parents, a pattern consistent with previous work conducted with the SCARED (Wren et al., 2004, 2007). The magnitude of the difference between the report of youth and their parents varied as a function of demographic characteristics. Unlike other studies of adolescents (e.g., Berger et al., 2005; Carlston & Ogles, 2009), we found no association between gender and either difference scores or *q* correlations, differences that may be due to sample characteristics. Berger et al. (2005) recruited a community sample. Clinically referred youths often experience higher levels of symptomatology, and their parents may be more aware of these symptoms, characteristics that may attenuate gender differences. Carlston and Ogles (2009) worked with a much larger clinical sample, which would have yielded greater power to detect variability in parent-youth disagreement across boys and girls. Similar to other studies (Berg-Nielsen et al., 2003; Stevanovic et al., 2012), we found that agreement between parents and youth increased with age, although the level of agreement remained low. The average *q* correlation among 16–18 year

olds was .37 suggesting that parents and older adolescents are still providing markedly different information.

Level of agreement between parent and youth report was also associated with ethnicity. Consistent with work by others (Lau et al., 2004; Youngstrom et al., 2000), we did not find a significant association between ethnicity and *q* correlations. When informant disagreement was indexed by difference scores, however, NHW dyads showed greater agreement than African-American pairs. There was also evidence for a comparable difference between NHW and Hispanic participants, but a relatively small number of Hispanic dyads likely limited power to detect this association. In both cases, minority youth reported higher levels of symptoms than their parents. These findings suggest the importance of considering youth report when using anxiety rating scales to inform diagnostic and treatment decisions. Our data suggest that this is particularly important when working with African-American youth. Among NHW families, parents and youth endorsed, on average, a comparable number of symptoms. If total score was used to make a diagnostic decision, for

example, by using the established clinical cutoff on the SCARED (25; Birmaher et al., 1999), 15% of NHW youth in our sample reported the presence of clinically concerning anxiety that would be “missed” if only parent report was obtained, whereas nearly a quarter (22%) of African-American youths who scored over this cutoff would remain unidentified.

Caution must be exercised when generalizing these findings. Ethnic minority youth were under-represented in our sample, relative to NHW youth. Minority families were not more likely to refuse consent to participate in research; thus, their greater exclusion from the analytic sample likely resulted, at least in part, because we required data from both parents and children. Ethnic minority families are likely to be experiencing greater economic disadvantage than NHW families (Taylor, Kochlar, Fry, Velasco, & Motel, 2011), and lower SES is associated with a number of risk factors (e.g., poorer reading ability, severity of mental health problems) that make it more likely that one or both informants did not complete the SCARED (see Kazdin, Holland, Crowley, & Breton, 1997). The issues that contributed to parents and children not completing the questionnaire would also limit their ability to provide this information solely for clinical purposes. For this reason, our sample may be representative of the families for whom it is possible to obtain both parent and youth report of youth psychopathology.

More generally, however, patterns of inter-rater agreement across ethnicity appear to be sensitive to sample differences. Although our findings were broadly consistent with work in one clinical sample (Lau et al., 2004), studies conducted with other clinical and primary-care samples yielded different patterns of parent-youth disagreement across ethnic groups (Carlston & Ogles, 2009; Wren et al., 2007). It is possible that these differences are attributable to general sample characteristics (e.g., age of participants or recruitment source). The observed variability might also be due to differences in composition of the ethnic groups being studied. Use of broad categories such as Hispanic, the convention in much psychological research, masks important differences within these groups (Safren et al., 2000). Our findings suggest that disagreement between parent and youth in their report of youth psychopathology varies as a function of race/ethnicity, and future work should unpack the specific factors contributing to this variability (Anderson & Mayes, 2010), which may include concerns about stigma and cross-cultural differences in recognition and identification of symptoms, as well as differences in facility with the language in which the assessment is being conducted (Gonzalez, Weersing, Warnick, Scahill, & Woolston, 2012). More precise elucidation of these mechanisms will help us better understand how our assessment tools perform in different settings.

Clinical use of psychopathology rating scales would also benefit from a greater understanding of

why reports from different informants vary so markedly. In particular, examining whether these instruments are invariant across informants will allow us to interpret mean-level differences in symptom reports appropriately, and could pinpoint differences in informants' reports, such as whether symptoms are making differential contributions to the underlying construct(s), that are of interest both clinically and theoretically. In our analysis of the SCARED, we found full configural and metric invariance, suggesting that parents and youth are evaluating comparable constructs: the factor structure and the magnitude of the factor loadings are equivalent across groups. This evidence for ‘weak’ invariance suggests that it is appropriate to compare associations between the latent constructs and external variables (Dimitrov, 2010). In addition, we found evidence for residual invariance, suggesting that items were measured with similar precision in each group (Dimitrov, 2010).

To compare latent-factor means across groups, it is necessary to establish threshold invariance (Dimitrov, 2010). We found evidence for partial threshold invariance: it was necessary to free 22 thresholds. Although this slightly exceeds the cutoff for ‘acceptable’ partial invariance (not more than 20% of parameters freed; Byrne, Shavelson, & Muthén, 1989), two lines of evidence suggest that the degree of variability observed across thresholds is not meaningful. First, although the χ^2 -difference test was significant when all thresholds were constrained, the change in CFI was less than .01, indicating no significant difference in fit (Vandenberg & Lance, 2000). Second, there did not appear to be systematic differences in thresholds across groups: parents and youth were equally likely to have the higher threshold, and there was no association between which informant had the higher threshold and symptom type.

Taken together, these analyses indicate that although parents and youth showed significant disagreement in their report on the SCARED, they used this instrument in similar ways. This cross-informant equivalence means that scores from parents and youth can be meaningfully compared. Completing the analyses separately for NHW and minority families yielded an identical pattern of results – full configural, metric, and residual invariance, and partial threshold invariance, with no evidence for systematic differences in thresholds across informants – validating comparison of report by parents and youth in these more precisely defined groups. The number of African-American and Hispanic families in our sample precluded testing measurement invariance separately in these groups. It will be important for future work to examine this issue, as well as examining the equivalence of measures across groups defined by constructs that may explain variability in and across ethnic groups, such as acculturation.

Beyond their psychometric importance, formal tests of measurement invariance provide a useful tool for evaluating the extent to which differences between raters are based on 'true' differences in symptomatology reported and not systematic differences in the underlying construct assessed. Although informants may be bringing unique perspectives to bear on their understanding of what constitutes clinically significant anxiety, these differences are not generally apparent in their interpretation of and responses to the SCARED. It seems likely given evidence for comparable factor loadings and thresholds across informants, that the differences in parent and youth report we observed are due, at least partly, to the symptoms to which they have access (see Gomez, 2007).

In summary, this study provides further evidence that, among clinically referred samples, disagreement between parent and youth report is greater for African-American families. Such data indicate the importance of using standardized instruments to obtain youth report concerning their anxiety, and internalizing symptoms more broadly (Lau et al., 2004), when working with these families. Although differences between youth and parent report on the SCARED are pronounced, these two informants are using this instrument in comparable ways, suggesting that they share similar conceptualizations of anxiety, and use similar criteria to determine when

symptoms are present. Formal tests of measurement invariance are a valuable strategy for elucidating why differences between raters are present, and future work should apply this technique to examine informant disagreement in ratings of other types of psychopathology.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 Thresholds freed in measurement invariance analysis

Acknowledgments

Financial support for this project was provided in part by an award to VRB, from the William T. Grant Foundation. The authors wish to thank Richard Zinbarg for his assistance with the statistical analyses. The authors have declared they have no potential or competing conflicts of interests.

Correspondence

Melanie A. Dirks, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC H3A1B1, Canada. Email: melanie.dirks@mcgill.ca

Key points

- We characterized parent-youth discrepancies in their report on the Screen for Child Anxiety Related Emotional Disorders (SCARED), and examined the equivalence of this measure across these two informants to determine whether they are conceptualizing youth anxiety comparably.
- Youth reported more symptoms than parents, and this difference was greater for African-American than for Non-Hispanic White dyads, suggesting the importance of considering youth report of anxiety when working with African-American families.
- The SCARED was invariant across youth and parent report, indicating that these informants are using this instrument comparably and differences in their report are not attributable to their rating different constructs or using different thresholds to determine when symptoms are present.

References

- Anderson, E.R., & Mayes, L.C. (2010). Race/ethnicity and internalizing disorders in youth: A review. *Clinical Psychology Review, 30*, 338–348.
- Baldwin, J.S., & Dadds, M.R. (2007). Reliability and validity of parent and child versions of the Multidimensional Anxiety Scale for Children in community samples. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*, 252–260.
- Berger, L.E., Jodl, K.M., Allen, J.P., McElhaney, K.B., & Kuperminc, G.P. (2005). When adolescents disagree with others about their symptoms: Differences in attachment organization as an explanation of discrepancies between adolescent, parent, and peer reports of behavior problems. *Development and Psychopathology, 17*, 509–528.
- Berg-Nielsen, T.S., Vika, A., & Dahl, A.A. (2003). When adolescents disagree with their mothers: CBCL-YSR discrepancies related to maternal depression and adolescent self-esteem. *Child: Care, Health and Development, 29*, 207–213.
- Birmaher, B., Brent, D.A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A replication study. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*, 1230–1236.
- Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., & McKenzie Neer, S. (1997). The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale construction and psychometric characteristics. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 545–553.
- Byrne, B.M., Shavelson, R.J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466.

- Carlston, D., & Ogles, B. (2009). Age, gender, and ethnicity effects on parent-child discrepancy using identical item measures. *Journal of Child and Family Studies, 18*, 125–135.
- Cole, D.A., Hoffman, K., Tram, J.M., & Maxwell, S.E. (2000). Structural differences in parent and child reports of children's symptoms of depression and anxiety. *Psychological Assessment, 12*, 174–185.
- De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 40*, 1–9.
- Dimitrov, D.M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*, 121–149.
- Dirks, M.A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L.S. (2012). 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, 53*, 558–574.
- Flora, D.B., & Curran, P.J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466–491.
- Gomez, R. (2007). Australian parent and teacher ratings of the DSM-IV ADHD symptoms: Differential symptom functioning and parent-teacher agreement and differences. *Journal of Attention Disorders, 11*, 17–27.
- Gomez, R., Vance, A., & Gomez, A. (2012). Children's Depression Inventory: Invariance across children and adolescents with and without depressive disorders. *Psychological Assessment, 24*, 1–10.
- Gonzalez, A., Weersing, V.R., Warnick, E., Scahill, L., & Woolston, J. (2012). Cross-ethnic measurement equivalence of the SCARED in an outpatient sample of African American and Non-Hispanic White youths and parents. *Journal of Clinical Child and Adolescent Psychology, 41*, 361–369.
- Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55.
- Kazdin, A.E., Holland, L., Crowley, M., & Breton, S. (1997). Barriers to treatment participation scale: Evaluation and validation in the context of child outpatient treatment. *Journal of Child Psychology and Psychiatry, 38*, 1051–1062.
- Krain, A.L., & Kendall, P.C. (2000). The role of parental emotional distress in parent report of child anxiety. *Journal of Clinical Child Psychology, 29*, 328–335.
- Lau, A.S., Garland, A.F., Yeh, M., McCabe, K.M., Wood, P.A., & Hough, R.L. (2004). Race/ethnicity and inter-informant agreement in assessing adolescent psychopathology. *Journal of Emotional and Behavioral Disorders, 12*, 145–156.
- van de Looij-Jansen, P.M., Jansen, W., de Wilde, E.J., Donker, M.C.H., & Verhulst, F.C. (2011). Discrepancies between parent-child reports of internalizing problems among preadolescent children: Relationships with gender, ethnic background, and future internalizing problems. *The Journal of Early Adolescence, 31*, 443–462.
- Merikangas, K.R., He, J., Burstein, M., Swanson, S.A., Avenevoli, S., Cui, L., ... & Swendsen, J. (2010). Lifetime prevalence of mental disorders in U. S. adolescents: Results from the National Comorbidity Survey Replication – Adolescent Supplement (NCS-A). *Journal of the American Academy of Child and Adolescent Psychiatry, 49*, 980–989.
- Muthén, B.O., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. Mplus Web Note No. 4. Available from: <http://www.statmodel.com/examples/webnote.shtml#web4>. [last accessed 5 February 2013].
- Muthén, L.K., & Muthén, B.O. (2010). *MPlus Users' Guide*, 6th ed. Los Angeles: Muthén and Muthén.
- Safren, S.A., Gonzalez, R.E., Horner, K.J., Leung, A.W., Heimberg, R.G., & Juster, H.R. (2000). Anxiety in ethnic minority youth: Methodological and conceptual issues and review of the literature. *Behavior Modification, 24*, 147–183.
- Sanne, B., Torsheim, T., Heiervang, E., & Stormark, K.M. (2009). The Strengths and Difficulties Questionnaire in the Bergen Child Study: A conceptually and methodically motivated structural analysis. *Psychological Assessment, 21*, 352–364.
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*, 486–492.
- Silverman, W.K., & Ollendick, T.H. (2008). Child and adolescent anxiety disorders. In J. Hunsley, & E.J. Mash (Eds.), *A guide to assessments that work* (pp. 181–206). New York: Oxford University Press.
- Stevanovic, D., Jancic, J., Topalovic, M., & Tadic, I. (2012). Agreement between children and parents when reporting anxiety and depressive symptoms in pediatric epilepsy. *Epilepsy & Behavior, 25*, 141–144.
- Taylor, P., Kochlar, R., Fry, R., Velasco, G., & Motel, S. (2011). *Wealth gaps rise to record highs between Whites, Blacks and Hispanics*. Washington, DC: Pew Research Center.
- Treutler, C., & Epkins, C. (2003). Are discrepancies among child, mother, and father reports on children's behavior related to parents' psychological symptoms and aspects of parent-child relationships? *Journal of Abnormal Child Psychology, 31*, 13–27.
- Vandenberg, R.J., & Lance, C.E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70.
- Waschbusch, D., & Willoughby, M. (2008). Parent and teacher ratings on the IOWA Conners Rating Scale. *Journal of Psychopathology and Behavioral Assessment, 30*, 180–192.
- Wren, F.J., Berg, E.A., Heiden, L.A., Kinnamon, C.J., Ohlson, L.A., Bridge, J.A., ... & Bernal, M.P. (2007). Childhood anxiety in a diverse primary care population: Parent-child reports, ethnicity and SCARED factor structure. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 332–340.
- 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 and Adolescent Psychiatry, 43*, 1364–1371.
- Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology, 68*, 1038–1050.

Accepted for publication: 9 June 2013

Published online: 25 October 2013