



HHS Public Access

Author manuscript

J Clin Child Adolesc Psychol. Author manuscript; available in PMC 2022 January 01.

Published in final edited form as:

J Clin Child Adolesc Psychol. 2021 ; 50(1): 58–76. doi:10.1080/15374416.2020.1802736.

Using Evaluative Criteria to Review Youth Anxiety Measures, Part I: Self-Report

Rebecca G. Etkin, Yaara Shimshoni, Eli R. Lebowitz, Wendy K. Silverman

Child Study Center, Yale University School of Medicine, New Haven, CT

Abstract

Evidence-based assessment serves several critical functions in clinical child psychological science, including being a foundation for evidence-based treatment delivery. In this Evidence Base Update, we provide an evaluative review of the most widely used youth self-report measures assessing anxiety and its disorders. Guided by a set of evaluative criteria (De Los Reyes & Langer, 2018), we rate the measures as Excellent, Good, or Adequate across their psychometric properties (e.g., construct validity). For the eight measures evaluated, most ratings assigned were Good followed by Excellent, and the minority of ratings were Adequate. We view these results overall as positive and encouraging, as they show that these youth anxiety self-report measures can be used with relatively high confidence to accomplish key assessment functions. Recommendations and future directions for further advancements to the evidence base are discussed.

Keywords

Evidence-based; Anxiety; Assessment; Measurement; Child

Anxiety disorders are common, with lifetime prevalence estimates as high as 32% for children and adolescents (hereafter referred to as youth unless referring to a specific developmental period) (Merikangas et al., 2010). Youth anxiety and its disorders often compromise functioning across family, school, and peer contexts, and left untreated, can have detrimental effects into adulthood, including poor financial, interpersonal, and physical and mental health outcomes (Copeland, Angold, Shanahan, & Costello, 2014; Copeland, Shanahan, Costello, & Angold, 2009). Evidence-based assessments that can perform key clinical and research functions, such as identifying and quantifying symptoms, are critical in helping to redirect these likely harmful trajectories. This is because appropriate use of evidence-based treatments depends on having accurate information about the clinical problems that require targeting; such knowledge is derived through evidence-based assessments. Indeed, evidence-based assessments beget evidence-based treatments.

Despite the critical role of evidence-based assessment, historically, there has been insufficient clarity about what constitutes a strong evidence base for assessment measures. The last time a review of youth anxiety assessment measures was published in this journal,

Silverman and Ollendick (2005) noted that the field was beginning to answer questions about the state of the assessment literature such as “If I can give only one rating scale as an anxiety screen, which one should I give?...Can I use a particular instrument to help differentiate between anxiety and other disorders, such as depression? Which measure or set of measures should I include in a treatment outcome study” (p. 380). Although Silverman and Ollendick (2005) expressed some optimism for the advancements in the literature, their optimism was tempered by the absence of criteria to evaluate measures, “At this point, a set of criteria or guidelines for what is an evidence-based assessment is simply not there.” (p. 404).

Evidence-Based Assessment Criteria

We are pleased that since Silverman and Ollendick (2005), criteria for evaluating assessment research have emerged. Developed by Hunsley and Mash (2008) and expanded by Youngstrom and colleagues (2017), these criteria specify how to evaluate/rate the quality and quantity of research examining measures’ psychometric properties. As illustrated by the rubric showing these criteria (Table 1), each psychometric property may earn a rating of Adequate, Good, or Excellent, with benchmarks for a specific rating varying by the psychometric property under consideration. For example, to earn a rating of Excellent for internal consistency, most studies on a given measure must report Cronbach’s alphas above .90. In general, a rating of Adequate indicates that a measure has a minimal level of rigorous research support for a given psychometric property; a rating of Good indicates that a measure has sound empirical research support; and a rating of Excellent indicates that a measure has substantial and high-quality research support for a given psychometric property (De Los Reyes & Langer, 2018).

The development and utilization of clear and specific criteria are important to advance evidence-based assessment. Such criteria enable consistent evaluation and comparison across measures, including how well each measure can accomplish a specific purpose or function based on the quality and quantity of empirical evidence (Silverman & Kurtines, 1996; Silverman & Ollendick, 2005; Youngstrom et al., 2017). For example, if a clinical scientist seeks to answer the questions posed above about which measures are optimal to differentiate anxiety from other problems or which should be used in a treatment outcome study, they can compare the ratings for discriminative validity or treatment sensitivity, respectively, which will point to the measure(s) with the strongest empirical evidence to accomplish these goals. Additional questions can be addressed by consulting these criteria, including “How do I know if a score on a given measure will be consistent if re-administered a few days later?” (test-retest reliability) and “Which measure can I validly use to assess anxiety in my diverse sample?” (validity generalization). These criteria have further value in highlighting gaps in the research literature that might stimulate new studies. For example, if a measure earns a rating of Adequate for norms (one of the psychometric criteria) due to a lack of descriptive data within large, relevant clinical samples, this could spur research to collect additional normative data.

Current Evaluative Review

In line with the goals of the *Journal of Clinical Child & Adolescent Psychology* Evidence Base Updates Series, the current review is guided by the evaluative criteria described above and presented in Table 1 (De Los Reyes & Langer, 2018). Other recent reviews of youth anxiety measures have discussed the utility of these criteria (Spence, 2018) or assigned ratings across some of the psychometric properties (Byrne, Lebowitz, Ollendick, & Silverman, 2018; Tulbure, Szentagotai, Dobrean, & David, 2012). Our current review moves beyond these past reviews in that it is the first to rate measures along Youngstrom and colleagues' (2017) expanded set of psychometric properties and to provide justification for these ratings. Based on the information that our ratings reveal about the strengths and weaknesses of each measure we address questions regarding their use, including which measure is best suited to perform which assessment function(s), and provide suggestions to help guide future research.

Our review also differs from others in that it focuses exclusively on youth anxiety self-report measures. In a forthcoming companion article, we conduct a similar evaluative review of parent-report youth anxiety symptom measures. We begin though with self-report measures because they have long played a central role in the multi-modal, multi-informant approach considered to be the gold-standard in the clinical child and adolescent psychological assessment literature. Harkening back to the seminal work of Lang (1968) and Rachman (1978) on the assessment of anxiety within a tripartite framework, self-reports provide ecologically-valid perspectives with theoretical and clinical implications concerning idiosyncratic manifestations of anxiety (e.g., associations with behavioral assessments of anxiety; Cannon et al., 2020; Deros et al., 2018; Shimshoni, Silverman, & Lebowitz, 2017). Moreover, youth self-report measures are highly practical (e.g., quick to administer, inexpensive) and versatile in their ability to serve multiple clinical functions as referenced above, such as screening, aiding in diagnosis, and informing case conceptualization (e.g., Byrne et al., 2018; Silverman & Kurtines, 1996). This also includes their longstanding premier status as estimators of treatment efficacy in controlled clinical trials (e.g., Davis, May, & Whiting, 2011; Silverman & Ollendick, 2005; Weisz, Jensen Doss, & Hawley, 2005). Considering their versatility, it is not surprising that self-report measures are also one of the 'units of analysis' for incorporation in research, including youth anxiety research (e.g., Lebowitz, Gee, Pine, & Silverman, 2018), in the National Institute of Mental Health Research Domain Criteria.

Method

Literature Search

We conducted keyword-guided electronic database searches using PsycINFO and Google Scholar to identify articles. Several searches included names of specific measures, such as "Multidimensional Anxiety Scale for Children OR MASC." We decided which measures to search for by consulting Silverman and Ollendick (2005) and other recent reviews (e.g., Byrne et al., 2018; Spence, 2018), which identified the most frequently used and studied youth anxiety assessment methods. Of the self-report measures identified, we searched for those that specifically assess anxiety *symptoms* and not anxiety-related constructs, such as

anxiety sensitivity (Silverman, Fleisig, Rabian, & Peterson, 1991). Our other searches cast a broader net by using descriptive terms (e.g., “Child* OR Youth OR Adolescen* OR Pediatric AND Anxiety AND Inventory OR Measure OR Scale OR Questionnaire”) to ensure we retrieved articles in which the measure name was not in the title. Finally, we searched for and consulted meta-analyses and reviews of youth anxiety treatment to locate studies with which to specifically rate measures’ treatment sensitivity, a psychometric criterion (e.g., Bennett et al., 2013; Creswell & Cartwright-Hatton, 2007; Creswell et al., 2020; Manassis et al., 2014). These search procedures yielded approximately 5,000 articles.

Article Screening and Inclusion

Articles were screened by the first and last authors to ensure they that were peer-reviewed, empirical studies evaluating the psychometric properties of youth self-report anxiety symptom measures. As per our search described above, our main inclusion criterion was that the measures assess symptoms of anxiety and/or its disorders. We did not include studies of measures specifically assessing posttraumatic stress disorder and obsessive-compulsive disorder, as they are no longer classified as anxiety disorders in the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5; American Psychiatric Association, 2013). Given our focus on assessment of anxiety symptoms in childhood and adolescence, we included studies examining samples of youth as young as 6 and as old as 18. We included only studies published in English using an English-version of the measure.

Of the articles screened, 136 fit our inclusion criteria. These articles report psychometric or treatment studies of the following eight measures: State Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973), Revised Children’s Manifest Anxiety Scale (RCMAS; Reynolds & Richmond, 1978), Fear Schedule Survey for Children - Revised (FSSC-R; Ollendick, 1983), Social Anxiety Scale for Children - Revised and Social Anxiety Scale for Adolescents (SASC-R/SAS-A; La Greca & Lopez, 1998; La Greca & Stone, 1993), Social Phobia and Anxiety Inventory for Children (SPAI-C; Beidel, Turner, & Morris, 1995), Multidimensional Anxiety Scale for Children (MASC; March, Parker, Sullivan, Stallings, & Connor, 1997), Screen for Child Anxiety Related Emotional Disorders (SCARED; Birmaher et al., 1997), and Spence Children’s Anxiety Scale (SCAS; Spence, 1998). We include the FSSC-R in our review because of its extensive past use to assess phobic disorders (e.g., Ollendick et al., 2009; Silverman et al, 1999b; Weems et al., 1999) prior to the development of other measures that contain specific fear or phobia subscales (e.g., SCAS Physical Injury Fears subscale). We present further details about each of these measures in Table 2. Our review focuses on the most widely used and studied versions of these measures. Some of these measures have been updated (RCMAS-2; MASC-2; SCARED-R; March, 2013; Muris, Merckelbach, Schmidt, & Mayer, 1998; Reynolds & Richmond, 2008; see Table 2) or abbreviated (e.g., 25-item FSSC-R; 11-item SPAI-C; 8-item SCAS; Bunnell, Beidel, Liu, Joseph, & Higa-McMillan, 2015; Muris, Ollendick, Roelofs, & Austin, 2014; Reardon, Spence, Hesse, Shakir, & Creswell, 2018) but they have undergone little psychometric scrutiny or the extant studies used non-English versions of measures.

Evaluation Procedure

All studies were reviewed for information pertinent to rating measures' psychometric properties as Adequate, Good, or Excellent. We provide below a summary of the research studies that informed our ratings of each psychometric property for each measure (organized from lowest to highest quality rating and from oldest to newest measure). We prioritized citing studies that have been published since Silverman and Ollendick (2005). Our ratings, including tallies of the number of Adequate, Good, and Excellent ratings found for each measure and each psychometric property, are presented in Table 3. A few of the psychometric properties delineated in the criteria are absent from our review because they were either not relevant to youth self-report measures (interrater reliability) or, as far as our search revealed, have not been examined in any research studies (repeatability, prescriptive validity, clinical utility).

Results

Norms

As shown in Table 3, 25% ($n = 2$) of ratings are Adequate, 62.5% ($n = 5$) of ratings are Good, and 12.5% ($n = 1$) of ratings are Excellent for norms. The FSSC-R and SASC-R/SAS-A each earn a rating of Adequate. These measures were originally normed in community samples ($Ns = 217 - 590$; La Greca & Lopez, 1998; La Greca & Stone, 1993; Ollendick, 1983); the FSSC-R normative sample also included a small clinical subsample ($n = 25$). Although descriptive data for the FSSC-R and SASC-R/SAS-A are available from at least one large ($N > 100$) clinical sample (e.g., Ginsburg, La Greca, & Silverman, 1998; Weems, Silverman, Saavedra, Pina, & Lumpkin, 1999), most subsequent studies were likewise conducted with community samples, or relatively small and predominantly White clinical samples (e.g., SASC-R: $n = 57$ clinical, $n = 178$ community; Epkins, 2002).

The STAIC, SPAI-C, MASC, SCARED, and SCAS each earn a rating of Good. The STAIC, MASC, and SCAS were originally normed in community samples ($Ns = 374 - 2,052$; the SCAS normative sample also included a small clinical subsample, $n = 40$; March et al., 1997; Spence, 1998; Spielberger, 1973) and the SPAI-C and SCARED were originally normed in clinical samples ($Ns = 100 - 341$; Beidel et al., 1995; Birmaher et al., 1997). Multiple subsequent studies report descriptive data for large, diverse (> 25% ethnic minority) clinical and community samples, often broken down by factors including subscale and/or youth sex, age, ethnicity, and diagnosis. Taking the SPAI-C, for example, Higa and colleagues (2006) reported 53% biracial or multiethnic youth in a community sample ($N = 158$, 10-14 years) and Viana, Rabian, and Beidel (2008) reported 32% Black or other ethnic minority youth in a clinical sample ($N = 172$, 7-17 years).

The RCMAS earns a rating of Excellent. It was originally normed in a nationally representative sample consisting of urban and rural community subsamples obtained from every geographical region in the United States ($N = 4,972$; Reynolds & Paget, 1983), which is the largest normative sample of all the measures. Descriptive data are also available from multiple other large, diverse community and clinical samples (e.g., $N = 632$, 13-15 years, 39% ethnic minority, non-clinical youth from across the U.S.; Dierker et al., 2001; $N = 677$,

6-16 years, 59% Hispanic, youth presenting to a specialty clinic; Pina, Little, Knight, & Silverman, 2009).

Internal Consistency

For internal consistency, 75% ($n = 6$) of ratings are Good and 25% ($n = 2$) of ratings are Excellent. The STAIC, RCMAS, SASC-R/SAS-A, MASC, SCARED, and SCAS each earn a rating of Good, though we note two points. One, we rated the more widely used STAIC subscale, STAIC-Trait, as it measures a more stable aspect of anxiety and generally shows higher internal consistency than the STAIC-State (e.g., Trait: $\alpha = .82 - .89$; State: $\alpha = .71 - .76$; Papay & Spielberger, 1986). Two, we rated the RCMAS, SASC-R/SAS-A, MASC, SCARED, and SCAS based on findings that Cronbach's alphas are above .80 or .90 for total scale scores, despite variability in alpha subscale scores. For example, several clinical and community studies reported subscale alphas between .60 - .77 for the RCMAS (Varela & Biggs, 2006), .72 - .91 for the SASC-R/SAS-A (La Greca, Ingles, Lai, & Marzo, 2015; Storch, Masia-Warner, Dent, Roberti, & Fisher, 2004), .64 - .87 for the MASC (Grills-Tacquechel, Ollendick, & Fisak, 2008; Rynn et al., 2006; Wei et al., 2014), .54 - .89 for the SCARED (Boyd, Ginsburg, Lambert, Cooley, & Campbell, 2003; Gonzalez, Weersing, Warnick, Scahill, & Woolston, 2012; Rappaport, Pagliaccio, Pine, Klein, & Jarcho, 2017), and .53 - .84 for the SCAS (Spence, Barrett, & Turner, 2003; Whiteside & Brown, 2008).

The FSSC-R and SPAI-C each earn a rating of Excellent. The total scale score is most commonly used in research with these measures and Cronbach's alpha estimates are consistently above .90 in clinical and community samples (e.g., FSSC-R: $\alpha = .92 - .94$; Ollendick, 1983; Ollendick, Yule, & Ollier, 1991; SPAI-C: $\alpha = .91 - .95$; Higa et al., 2006; Storch et al., 2004).

Test-Retest Reliability and Stability

Test-retest reliability and stability both refer to the consistency of a measure's scores across time. Test-retest reliability measures consistency of scores over a few days or weeks, while stability measures consistency of scores over a few months or longer (Watson, 2004). Both metrics are assessed with intraclass correlation coefficients (ICCs) or Pearson's r . In general, higher values indicate higher test-retest reliability and stability, and it is expected that test-retest values are higher than stability values (Watson, 2004; Youngstrom, Salcedo, Frazier, & Perez Algorta, 2019). In line with past guidelines, we used the following criteria to rate measures' test-retest reliability and stability: ICCs $> .74$ and Pearson's $r > .70$ are considered Excellent; ICCs = .59 - .74 and Pearson's $r = .50 - .70$ are considered Good; and ICCs = .40 - .58 and Pearson's $r = .30 - .50$ are considered Adequate (e.g., Cohen, 2013; Landis & Koch, 1977). Our ratings reflect measures' overall performance for both test-retest reliability and stability, a point we return to in the Discussion.

Seventy-five percent ($n = 6$) of ratings are Good and 25% ($n = 2$) of ratings are Excellent for test-retest reliability and stability. The STAIC-Trait, RCMAS, FSSC-R, SASC-R/SAS-A, SCARED, and SCAS each earn a rating of Good. For the STAIC-Trait, Spielberger (1973) reported $r_s = .65 - .71$ over 6 weeks in a community sample. For the RCMAS, Wisniewski and colleagues (1987) reported $r_s = .60 - .88$ over 1-5 weeks for total and subscales scores,

and Reynolds (1981) reported $r = .68$ over 9 months, both in community samples. For the FSSC-R, Ollendick (1983) reported $r = .82$ for the total scale score over 1 week and $r = .55$ over 3 months in a community sample. For the SAS-A, La Greca and Lopez (1998) reported $r_s = .54 - .78$ over 2 months and Storch et al. (2004) reported $r_s = .55 - .62$ over 12 months, both in community samples. For the SCARED, Birmaher and colleagues (1997) reported ICCs = .70 - .90 over 5 weeks in a clinical sample (for subscale and total scale scores; 38-item version); Behrens and colleagues (2019) reported ICCs = .59 - .61 over an average of 40 days in a sample of clinical and healthy control youth. Regarding stability, Boyd and colleagues (2003) reported $r_s = .19 - .48$ over 6 months in a community sample. For the SCAS, Spence (1998) reported $r_s = .45 - .60$ over 6 months (for subscale and total scale scores) and Spence and colleagues (2003) reported $r_s = .51 - .75$ over 12 weeks, both in community samples.

The SPAI-C and MASC each earn a rating of Excellent. For the SPAI-C, Beidel and colleagues (1995) reported $r = .86$ over 2 weeks in a mixed sample of clinical and healthy control youth. Regarding stability, Beidel and colleagues (1995) reported $r_s = .63$ over 10 months in a clinical sample and Storch and colleagues (2004) reported $r = .47$ over 12 months in a community sample. For the MASC, March, Sullivan, and Parker (1999) reported ICCs = .62 - .92 over 3 weeks for the total and subscale scores in a community sample. In two studies assessing stability, March and colleagues (1997) reported ICCs = .87 - .93 over 3 months in a clinical sample and Baldwin and Dadds (2007) reported $r_s = .47 - .55$ (for subscale and total scale scores) over 12 months in a community sample.

Content Validity

For content validity, 62.5% ($n = 5$) of ratings are Good and 37.5% ($n = 3$) are Excellent. The STAIC, RCMAS, FSSC-R, SASC-R/SAS-A, and SCARED each earn a rating of Good because developers clearly defined each measure's content domains, ensured representation across items, and items were evaluated by one set of judges (i.e., experts or youth). The STAIC, RCMAS, and FSSC-R are all downward extensions of adult anxiety self-report measures. Developers relied further on the input of experts including researchers, graduate students, clinicians, and/or teachers to modify items to ensure the items were developmentally appropriate. The SASC-R developers relied similarly on experts to reduce and refine items by rating how well each item assessed social anxiety in children (La Greca & Stone, 1993). In constructing the SAS-A, the content remained the same, but SASC-R wording was revised to assure its appropriateness with adolescents (e.g., "kids" changed to "peers"; "playing" changed to "doing things"; La Greca & Lopez, 1998). Regarding the SCARED, developers administered items to a small sample of youth to check comprehension and ascertain feedback to further revise items before inclusion in the measure (Birmaher et al., 1997).

The SPAI-C, MASC, and SCAS each earn a rating of Excellent. In addition to meeting the criteria above for a rating of Good, developers refined the measures through consultation with judges and pilot tested the next iteration with youth. For example, the MASC developers used multiple procedures, including a Q-sort by experts and youth with and

without anxiety disorders, followed by pilot testing with youth to reduce and refine the initial item pool (March et al., 1997).

Construct Validity

All of the youth anxiety self-report measures earn a rating of Good (100%; $n = 8$), as the bulk of independent studies demonstrate the various facets of construct validity. In demonstrating convergent validity, for example, moderate-to-strong, positive correlations have been found between measures (e.g., $r = .73$ for the SCARED and STAIC-Trait; Monga et al., 2000; $r = .61 - .81$ for the SASC-R and SPAI-C; Epkins, 2002; $r = .61$ for the MASC and RCMAS; Rynn et al., 2006). This extends as well to specific subscales. For example, there is evidence that of the five FSSC-R subscales, the Failure and Criticism subscale correlates highest with the SPAI (Clark et al., 1994), of the four MASC subscales the Social Anxiety subscale correlates highest with the SAS-A and SPAI-C (Anderson, Jordan, Smith, & Inderbitzen-Nolan, 2009), and of the six SCAS subscales the Generalized Anxiety subscale correlates highest with the RCMAS (Spence, 1998; Spence et al., 2003).

In demonstrating divergent validity, as would be expected, the bulk of evidence reveals nonsignificant associations between the youth anxiety self-report measures and externalizing symptom measures. For example, the SCARED and MASC are not significantly correlated with measures of hyperactivity ($r = .07 - .17$, $p > .05$; Boyd et al., 2003; March et al., 1997) and the SPAI-C is not significantly correlated with Child Behavior Checklist Externalizing scale ($r = .18$, $p > .05$; Beidel et al., 1995). In contrast, most studies find the anxiety self-report measures to correlate significantly and positively with measures of depression. This is not surprising given comorbidity across these disorders, item-overlap across measures, and shared-method variance (e.g., RCMAS and Children's Depression Inventory; Carter, Silverman, Allen, & Ham, 2008). Nevertheless, several studies found that the MASC (e.g., Rynn et al., 2006), SCAS (Spence, 1998, Spence et al., 2003), and SASC-R/SAS-A (Inderbitzen & Hope, 1995; Inderbitzen-Nolan & Walters, 2000) show higher correlations with other anxiety measures than with depression measures (or nonsignificant correlations in some studies, e.g., $r = .19$, $p > .05$ for the MASC and Children's Depression Inventory; March et al., 1997), while the RCMAS and STAIC correlate as highly with depression measures (Dierker et al., 2001; Hodges, 1990; March et al., 1997).

Insufficient research attention has focused on investigating *incremental* validity, or a measure's ability to predict clinically relevant data above and beyond other self-report measures (e.g., Haynes & Lench, 2003; Johnston & Murray, 2003). One study suggests that in non-clinical adolescent samples, the SPAI-C may have incremental validity in predicting social anxiety disorder (SOC) controlling for the SAS-A and MASC Social Anxiety subscale scores (Anderson et al., 2009). Several studies have examined incremental validity of youth- versus parent-report forms. In one of these studies, youth self-reports on the SCARED Social Anxiety subscale predicted performance on a behavioral task measuring social anxiety, controlling for scores on the parent version of the SCARED Social Anxiety subscale (Bowers et al., 2020). Given the scarcity of research on incremental validity overall, no measure is rated Excellent.

Discriminative Validity

For discriminative validity, 50% ($n = 4$) of ratings are Adequate, 25% ($n = 2$) are Good, and 25% ($n = 2$) are Excellent. The STAIC, RCMAS, FSSC-R, and SASC-R/SAS-A each earn a rating of Adequate. Most studies of the STAIC, RCMAS, FSSC-R find that clinically anxious youth have significantly higher mean scores compared with healthy control youth, but not compared with youth with depressive or externalizing disorders (Hodges, 1990; Last, Francis, & Strauss, 1989; Mattison, Bagnato, & Brubaker, 1988; Perrin & Last, 1992; Strauss, Last, Hersen, & Kazdin, 1988). For the RCMAS and STAIC-Trait, Hodges (1990) found the total scale scores yielded sensitivity rates of 34% and 42% respectively, in a psychiatric inpatient sample. For the FSSC-R, Weems and colleagues (1999) found that only 62% of youth with specific phobias or SOC could be accurately classified by discriminant function analysis using the subscale scores, although specific items were better able to discriminate specific phobias relative to the subscale scores (also found by Last et al., 1989). For the SASC-R/SAS-A, mean scores are significantly higher for youth with SOC than with other anxiety disorders (Beidel, Turner, Hamlin, & Morris, 2000; Epkins, 2002; Ginsburg et al., 1998), but there is also evidence of low sensitivity (i.e., 44% in a community sample; Inderbitzen-Nolan, Davies, & McKeon, 2004).

Several studies used receiver operating characteristics (ROC) analyses, which provide information about measures' diagnostic accuracy with the area under the curve (AUC) metric (values closer to 1.00 indicate better discrimination). ROC analyses can also identify clinical cutoff scores that maximize measures' sensitivity and specificity. For the STAIC-Trait, Monga and colleagues (2000) found poor discrimination between youth with anxiety and other psychiatric disorders with the parent version ($AUC = .61$), noting that similar results were found with the youth version. For the RCMAS total scale score, Dierker and colleagues (2001) found low accuracy for identifying youth diagnosed with SOC, generalized anxiety disorder (GAD), and separation anxiety disorder (SAD) ($AUCs = .51 - .67$) in a community sample. One ROC analysis of the SASC-R/SAS-A found that it could not identify SOC in a primary care sample (Bailey, Chavira, Stein, & Stein, 2006). We did not find any ROC studies of the FSSC-R.

The SCARED and SCAS each earn a rating of Good. For the SCARED, studies find that mean scores are significantly higher for youth with anxiety disorders than those with depression, externalizing disorders, and healthy controls (e.g., Monga et al, 2000; Rappaport et al., 2017). Birmaher and colleagues (1999) also found good sensitivity (71%) and specificity (61% - 71%) of the total scale score for discriminating anxiety, depressive, and externalizing disorders in a clinical sample. For the SCAS, Brown-Jacobsen, Wallace, and Whiteside (2011) found the total scale score to be reasonably sensitive (.64) and specific (.62) in identifying the presence of any anxiety disorder, although the subscale scores, especially those for Generalized Anxiety, were weaker in identifying corresponding disorders (sensitivity: .33 - .78, specificity: .42 - .84).

Several studies found evidence of discrimination using ROC analyses. As would be expected, the SCARED total and subscale scores have low accuracy for discriminating anxiety and depression ($AUC = .60$; Birmaher et al., 1997), but moderate accuracy for discriminating anxiety and externalizing disorders ($AUC = .68 - .78$; Birmaher et al., 1997)

and other psychiatric disorders (AUCs = .66 - .86; Birmaher et al., 1997; Monga et al., 2000). As noted above, Monga and colleagues report findings with the parent version but state that findings with the youth version were similar. Rappaport and colleagues (2017) also found that the SCARED total scale score successfully identified the presence of any anxiety disorder, and the Generalized Anxiety and Social Anxiety subscale scores successfully identified corresponding disorders in treatment-seeking (AUCs = .83 - .90) and non-treatment-seeking youth (AUCs = .67 - .76). For the SCAS, Reardon and colleagues (2019) found that the Separation Anxiety and Social Phobia subscales (for girls only) were most successful in identifying youth with the corresponding disorder, while the Generalized Anxiety and Physical Injury subscales were least successful.

The SPAI-C and MASC each earn a rating of Excellent. Mean MASC scores are significantly higher for anxious than depressed youth (e.g., Rynn et al., 2006) and mean SPAI-C scores are significantly higher for youth with SOC than those with other anxiety disorders, externalizing disorders, or no diagnoses (e.g., Beidel, 1996; Beidel et al., 2000; Epkins, 2002). Several studies using ROC analyses found evidence of significant discrimination, not only with regard to discriminating between clinical and healthy control youth, for the MASC (e.g., AUC = .91 in discriminating anxiety and externalizing disorders in a psychiatric inpatient sample; Osman et al., 2009) and the SPAI-C (e.g., AUC = .65 for discriminating youth with SOC and those with both SOC and GAD in a highly comorbid outpatient sample; Viana et al., 2008).

Although earlier research found that the MASC accurately identified GAD only in non-clinical girls (AUC = .82; Dierker et al., 2001), later research found that the MASC performed poorly in identifying GAD in clinical samples with either total or subscale scores (e.g., Rynn et al., 2006; Wood, Piacentini, Bergman, McCracken, & Barrios, 2002). However, there is evidence that the MASC does perform well in identifying SAD and SOC with corresponding subscales (e.g., AUC = .80, sensitivity = .63, specificity = .64 - .82; Anderson et al., 2009; AUCs = .69 - .86; Wei et al., 2014; sensitivity = .63 - .89, specificity = .64 - .68; Wood et al., 2002). Although most research on the SPAI-C focuses on SOC specifically, one study reported AUCs of .89 for discriminating any anxiety disorder and SOC specifically from healthy control youth in a diverse sample ($N = 501$, 8-16 years, 24% Black; Pina, Little, Wynne, & Beidel, 2014). Pina and colleagues (2014) also reported good sensitivity (.70 - .72) and specificity (.90) for discriminating youth with any anxiety disorder or SOC from healthy control youth using the recommended cutoff score of 18, but found that sensitivity significantly improved at cutoff scores of 13 - 14 (.81 - .87).

Validity Generalization

All of the self-report measures earn a rating of Excellent (100%; $n = 8$) for validity generalization, as the bulk of evidence supports their use with more than one specific youth demographic group (e.g., sex, race/ethnicity, socioeconomic status), and/or in multiple settings (i.e., schools, specialty clinics, general outpatient clinics, inpatient units, residential treatment, primary care/medical units). Each of these measures has also been used cross-culturally and translated into multiple languages (see Table 2).

Several studies provide evidence of validity generalization by testing measurement invariance across demographic groups. Studies' findings provide varying levels of confidence about whether measures are equivalent between groups. This is based on whether evidence of configural (i.e., factor structure is equivalent between groups), metric (i.e., factor loadings are equivalent between groups) or scalar (i.e., scores are equivalent between groups) invariance is found. There is evidence of configural invariance and at least partial metric invariance (i.e., most items have equivalent factor loadings) in different ethnic groups for most measures, indicating that these measures can be validly used to assess anxiety and its associations with related constructs in these groups, though there are exceptions (e.g., Boyd et al., 2003; Kingery, Ginsburg, & Burstein, 2009; Wren et al., 2007).

Taking For the SCARED, for example, Gonzalez and colleagues (2012) found evidence of configural and partial metric invariance in a clinical sample of Hispanic, Black, and White youth ($N = 408$, 5-18 years), and Skriner and Chu (2014) found configural and full metric invariance in a community sample of Hispanic, Black, White, and Asian youth ($N = 881$, 11-14 years). Other studies have similarly found evidence for configural and/or metric invariance of the RCMAS (Pina et al., 2009; Varela & Biggs, 2006), FSSC-R (Varela, Sanchez-Sosa, Biggs, & Luis, 2008), SPAI-C (Pina et al., 2014), SASC-R/SAS-A (La Greca et al., 2015; Storch, Eisenberg, Roberti, & Barlas, 2003), MASC (Brown et al., 2013) and SCAS (Holly, Little, Pina, & Caterino, 2015) for White and ethnic minority youth. We found one study (Pina et al., 2009) that established scalar invariance, which provides the highest confidence that scores are comparable across groups and that score differences are true rather than an artifact of the measure (e.g., items are understood differently by different groups).

Treatment Sensitivity

Twenty-five percent ($n = 2$) of ratings are Good and 75% ($n = 6$) are Excellent for treatment sensitivity. The STAIC and SASC-R/SAS-A each earn a rating of Good because several studies found evidence that scores are sensitive to change over the course of treatment (e.g., STAIC: individual and family cognitive behavioral therapy [CBT]; Bodden et al., 2008; SASC-R: group CBT for social phobia; Gallagher, Rabian, & McCloskey, 2004). The RCMAS, FSSC-R, SPAI-C, MASC, SCARED, and SCAS are the most widely used measures in treatment outcome research and each earn a rating of Excellent. Multiple studies have found that all these measures are sensitive to change across different modalities of CBT, including individual (e.g., FSSC-R, Kearney & Silverman, 1999; SPAI-C, Herbert et al., 2009; SCARED, Lebowitz, Marin, Martino, Shimshoni, & Silverman, 2020), group (e.g., RCMAS, Silverman et al., 2019; FSSC-R, Silverman et al., 1999a; SPAI-C and SCARED, Ingul et al., 2014), family or parent-involved (e.g., RCMAS, Silverman et al., 2019; FSSC-R, Silverman et al., 1999b; MASC, Wood, Piacentini, Southam-Gerow, Chu, & Sigman, 2006; SCAS, Rapee et al., 2013), and internet-delivered (e.g., SPAI-C and SCAS, Spence, Donovan, March, Kenardy, & Hearn, 2017). These measures are also sensitive to other approaches such as clinic and school-based social skills training (e.g., SPAI-C, Beidel et al., 2007; Masia Warner, Fisher, Shrout, Rathor, & Klein, 2007) and pharmacological treatment (e.g., SPAI-C, Beidel et al., 2007; SCARED and MASC, Walkup et al., 2001).

Several recent studies have examined measures' ability to identify recovery following treatment using ROC analyses. Evans and colleagues (2017), for example, found the Separation Anxiety subscale of the SCAS to accurately identify recovery from SAD (AUC = .80) in a sample of 7-12-year-old youth participating in clinician-guided parent-delivered CBT. Caporino and colleagues (2017) examined clinical cutoffs for treatment response and remission among youth with anxiety disorders participating in the Child/Adolescent Anxiety Multimodal treatment study (CBT, medication, and combination) and found that reductions in SCARED total scores of 50% predicted response and 60% predicted remission.

Discussion

Our evaluative review in this article moves beyond past reviews of youth anxiety self-report measures (Silverman & Ollendick, 2005; Spence, 2018) by our application of criteria that allowed for rating the quality and quantity of research examining measures' psychometric properties (Hunsley & Mash, 2008; Youngstrom et al., 2017). These criteria moreover permitted us to carry out consistent evaluations and comparisons across measures, including their strengths and limitations in performing different assessment functions.

Our findings yield an overall positive picture of the psychometric properties of these widely used measures, leaving us feeling optimistic about the state of the evidence base. Of 64 ratings assigned across the eight psychometric properties and eight measures evaluated, close to 38% ($n = 24$) are ratings of Excellent, 53% ($n = 34$) are Good, and a minority 9% ($n = 6$) are Adequate. It is encouraging to see a majority of Good and Excellent ratings, as this indicates on a general level that most of the youth self-report measures can be used for most assessment functions with confidence. The SPAI-C and MASC have the most Excellent ratings of all the measures; 75% ($n = 6$) for the SPAI-C and 62.5% ($n = 5$) for the MASC. These measures, along with the SCARED and SCAS also have no Adequate ratings. The STAI-C-Trait and SASC-R/SAS-A have the fewest Excellent ratings (12.5% or $n = 1$ for each), and the RCMAS and FSSC-R ratings are most variable.

The results are likewise encouraging when examining the ratings by criteria. The few ratings of Adequate were confined to only two psychometric properties (norms and discriminative validity) with the rest being Good and Excellent. These findings are testament to the burgeoned empirical scrutiny that these measures have undergone for over more than a decade (Silverman & Ollendick, 2005). Because of the importance of tying a measure's utility to a specific assessment function, our remaining discussion focuses on what the review's findings reveal about the measures' utility vis-à-vis the psychometric criteria. We proceed from those criteria with the most to least empirical support, based on their distribution of ratings.

Criteria with the most Excellent ratings.

The one criterion to receive all ratings of Excellent was validity generalization. All Excellent is indeed excellent! This finding implies that in responding to the important question about which measures can be used with diverse samples or in diverse settings, the answer is that any of these youth anxiety self-report measures can accomplish this goal. This latitude and flexibility will be especially helpful when assessment decisions are limited by other criteria

(e.g., if selecting a measure for its discriminative abilities, one can be assured that the measure has been used in diverse samples). However, we acknowledge that the criteria for validity generalization perhaps has more latitude and flexibility than might be optimal. For example, there would be fewer ratings of Excellent if evidence of measurement invariance was required to make this rating. Although all the self-report measures covered in this review, except for the STAIC, had the support of at least one study establishing configural or metric measurement invariance across two or more ethnic or racial groups, more research is needed. Criteria that promotes comparison of measures' degree of invariance (configural, metric, and scalar) across different demographic factors (e.g., age, gender) and settings (e.g., outpatient clinic, primary care) would provide further evidence for these measures' validity generalization and help identify gaps in the literature.

Treatment sensitivity is the next criterion to receive the most number of Excellent ratings. This finding also therefore implies some degree of latitude and flexibility about which measure to use to assess change in anxiety in response to treatment. As with validity generalization, our review reveals that it could be worthwhile to fine-tune the criteria to allow for more nuanced distinctions among measures. The few studies showing self-report measures' ability to assess treatment response or remission using ROC analyses highlight a useful approach in further specifying the criteria for treatment sensitivity (Caporino et al., 2017; Evans et al. 2017). Research is further needed to expand the evidence base to permit guidance about using youth anxiety self-report measures to monitor treatment progress. Such research would address questions including: What frequency of measure-administration captures change? And, what magnitude of change is 'meaningful' at different time intervals? The latter question relates to earlier calls by Kazdin (1977) for social validation of measures used in treatment studies (e.g., anxiety reduction relating to social interactions is associated with increased social interactions) and by Blanton and Jaccard (2006) for establishing empirical linkages between quantitative changes in measures' treatment scores and meaningful life outcomes.

These gaps in knowledge notwithstanding, there is evidence that all measures are sensitive to different types and modalities of treatment, although to a lesser degree for the STAIC and SASC-R/SAS-A, which are the two measures rated Good and not Excellent. To the extent that a treatment targets a specific disorder, the SPAI-C, MASC, SCARED, and SCAS could be particularly useful because these measures were all developed based on DSM criteria and have disorder-specific subscales. Of note though is that most treatments target a broad spectrum of anxiety symptoms or disorders, and therefore treatment studies typically rely on measures' total scale rather than disorder-specific subscale scores. As such, we find it interesting that the STAIC, RCMAS, and FSSC-R, none of which are tied to specific diagnoses, have been used far less in recent years. The RCMAS stands out to us particularly because it was the most widely used measure for assessing treatment sensitivity in the past (Silverman & Ollendick, 2005). Whether it is the RCMAS or another measure, we retain the past recommendation of Silverman and Ollendick (2005) and recently of Creswell and colleagues (2020), that a common measure or set of measures be adopted in treatment studies to allow for comparisons across studies.

Criteria with mostly Good ratings.

Internal consistency, test-retest reliability and stability, content validity, and construct validity each received mostly Good ratings. The FSSC-R and SPAI-C are the only measures rated Excellent for internal consistency, the SPAI-C and MASC for test-retest reliability and stability, and the SPAI-C, MASC, and SCAS for content validity. Relating to internal consistency, most measures were Good due to the range of alphas for subscales (most total scale scores were $>.90$, which is criteria for Excellent). While we were unable to provide an evaluative summary of specific subscales, some show consistently low internal consistency (e.g., School Avoidance subscale of the SCARED; Physical Injury Fears subscale of the SCAS; Gonzalez et al., 2012; Whiteside & Brown, 2008); we therefore suggest interpretation with caution. Given limitations of Cronbach's alpha (e.g., its magnitude is tied to the number of scale items; Youngstrom et al., 2019), we expect to see growing awareness in psychological research for the utility of reporting not only Cronbach's alpha as an estimate of internal consistency but also McDonald's total omega (ω_t ; McDonald, 1999), which is considered a more appropriate estimate of total scale reliability (Revelle & Condon, 2019).

Compared with internal consistency, our review revealed that far less research has examined measures' test-retest reliability and stability. We also found providing these ratings particularly challenging. This is because, according to the criteria, a measure is rated Excellent if retest correlations over one year are $\geq .70$, but Adequate if retest correlations over 2 weeks are $\geq .70$. This raised a concern on our part about conflating test-retest reliability and stability. To address this concern, we chose to rate measures as Adequate, Good, and Excellent using widely used benchmarks for the magnitude of Pearson's r or ICCs (e.g., Cohen, 2013; Landis & Koch, 1977). Not only would one anticipate these coefficients to be lower over longer periods of time, but one would hope that measures' scores are less consistent over longer periods of time due to changes that would be expected with development or intervention. We therefore recommend revisions in the test-retest reliability and stability criteria reflective of these issues. Based on the current review, however, the SPAI-C and MASC have the most empirical support for their ability to consistently assess anxiety symptoms over short and long intervals of time.

With respect to construct validity, ratings are Good for all measures. This finding indicates that all the measures are associated with theoretically related variables in expected ways (convergence) and with unrelated variables in expected ways (divergence). Convergent and divergent validity are most widely represented in the evidence base and so they were the focus of our review. Nevertheless, all measures show Good evidence for other aspects of construct validity, including concurrent and predictive validity.

The scant research conducted on measures' incremental validity precluded us from rating any measure as Excellent for construct validity. This gap is notable given earlier recommendations for such studies (e.g., Haynes & Lench, 2003; Johnston & Murray, 2003). For example, Muris and colleagues (2002) suggested examining the incremental validity of the newer measures (MASC, SCARED, SCAS) over the older ones (STAIC, RCMAS, FSSC-R). As researchers seek to address this gap, studies of anxiety sensitivity could provide helpful examples; many studies have found that the Child Anxiety Sensitivity Index

(Silverman et al., 1991) has incremental validity in predicting youths' fears and panic symptoms above and beyond anxiety measures (e.g., McLaughlin, Stewart & Taylor, 2007; Weems, Hammond-Laurence, Silverman, & Ginsburg, 1998). It is also relevant that incremental validity is a multidimensional construct. As such, measures may show varying degrees of incremental validity when it comes to content, sensitivity to change, predictive ability, and other psychometric properties (Haynes & Lench, 2003). This raises the possibility that incremental validity may warrant its own criteria, separate from construct validity.

Criteria with Adequate ratings.

Norms and discriminative validity were the only criteria to receive ratings of Adequate and not only Good and Excellent. The FSSC-R and SASC-R/SAS-A were the two measures rated Adequate for norms due to the limited descriptive data from large (i.e., $N > 100$) and diverse clinical samples relative to those available for other measures. Given what we noted above about the declining use of the RCMAS, it is interesting that it is the only measure rated Excellent for norms and even to date, to have the largest normative sample ($N = 4,972$) relative to the other measures. We add though that it is unclear whether norms from 1978 cohorts of youth are representative of 2020 cohorts; questions can further be raised about the representativeness of a sample of about 5,000 in a country that currently contains about 70 million youth. Another question relating to the RCMAS's normative sample is its low racial/ethnic diversity (12%), which is far less representative of the current proportion of minority youth in the U.S. It is perhaps a lofty goal to use methods specified in the criteria such as census-matching to establish norms (which is most closely approximated by the nationally representative sampling conducted by RCMAS developers), but such methods would help to address some of the questions raised here.

Discriminative validity is an especially salient criterion for most clinical child scientists, as it addresses the important question of which measure to use for distinguishing anxiety from other disorders. The bulk of studies that address this question examined mean differences across different comparator groups. We are encouraged to see more studies addressing this question through more sophisticated methods, namely ROC analyses. Of the 13 ROC studies we identified, only one was noted in Silverman and Ollendick (2005). We are also encouraged to see more studies using comparator groups other than healthy controls, which establishes more clinically realistic conditions for gauging measures' discriminative ability (Hunsley & Mash, 2008; Youngstrom et al., 2017).

Despite this progress, discriminative validity still received more Adequate ratings than any other criterion. This is because the four measures rated as Adequate, the STAIC, RCMAS, FSSC-R, and SASC-R/SAS-A, could not discriminate anxiety and other psychiatric disorders including externalizing disorders, and in some cases, healthy controls. As mentioned earlier, this could partly be because these measures assess anxiety symptoms generally rather than along DSM criteria. These measures also have only either very few or no studies that used ROC analyses, further limiting our confidence in their discriminative validity. In contrast, the measures rated Good and Excellent have evidence of distinguishing anxiety from other disorders, with multiple studies supported by ROC analyses. The MASC

and SPAI-C were rated Excellent, and support is the strongest for these measures to identify and discriminate SOC and SAD (MASC only). We also recommend the SCARED and SCAS, which earned ratings of Good, especially if it is necessary to utilize a measure that is available at no cost. The SCARED also has somewhat more support for identifying GAD.

To continue advancing research on discriminative validity, we offer some additional suggestions. One is that the criteria could be expanded to include benchmarks for evaluating aspects of discriminative validity beyond AUCs, including mean differences, positive and negative predictive value, and specificity and sensitivity (e.g., Pina, Silverman, Alfano, & Saavedra, 2002). Aspects of discriminative validity should also be considered within the context of the intended assessment function; for instance, high specificity at the cost of lower sensitivity may be acceptable if the assessment goal is to rule out the presence of an anxiety disorder. We also suggest research examining measures' ability to discriminate anxiety disorders beyond SOC, SAD, and GAD, utilizing samples that are representative of the diversity and psychiatric comorbidity common to different settings (e.g., Viana et al., 2008). It is expected that discrimination will be lower between disorders that tend to be comorbid or have diagnostic or conceptual overlap (e.g., high negative affect) and the criteria for discriminative validity should reflect this. Other reviews have further suggested that broad-band measures assessing both anxiety and depressive symptoms might be better suited for such discrimination (Spence, 2018).

Additional Recommendations

Multimethod assessment. This review focuses on self-report measures, which have inherent limitations (e.g., social demand; Pina, Silverman, Saavedra, & Weems, 2001). We therefore recommend, as did Silverman and Ollendick (2005), that self-report measures be used in conjunction with other assessment methods. For example, the Yale Interactive Kinect Environment Software platform (Lebowitz & François, 2018) is a novel assessment that uses motion-tracking technology to measure behavioral avoidance, a key clinical feature of anxiety and its disorders, and has preliminary evidence of validity (e.g., significant associations with anxiety rating scales), test-retest reliability, and treatment sensitivity (e.g., reductions in avoidance for clinically anxious youth following CBT). The Unfamiliar Peer Paradigm (Cannon et al., 2020) is another new multimodal assessment consisting of standardized interaction tasks with a peer confederate that provides psychometrically sound and clinically useful data on several constructs relevant to adolescent social anxiety (e.g., observer-rated social skills and self-reported arousal were both moderately-to-highly correlated with the SPAI-C). Such methods can complement self-report measures by contributing valuable information about the functional impact of symptoms, and can be used to measure treatment mechanisms or test theoretical linkages between youth self-reports of anxiety, avoidance, and related constructs (e.g., Cannon et al., 2020; Lebowitz, Shic, Campbell, Basile, & Silverman, 2015).

Multi-informant assessment. Another recommendation is that researchers further investigate the utility of multi-informant assessment, as studies suggest that clinicians, mothers, fathers, teachers, and peers may all contribute critical information about anxiety in youth (e.g., Moreno, Silverman, Saavedra, & Phares, 2008; Ford-Paz et al., 2019; Reardon et

al., 2019; Wei et al., 2014). For example, Wei and colleagues (2014) found that using the parent-report MASC in conjunction with the youth self-report version improved the accuracy of predicting SAD, SOC, and GAD diagnoses; Reardon and colleagues (2019) found that using data from two or more informants improved the sensitivity of the SCAS Separation Anxiety and Social Phobia subscales for predicting corresponding diagnoses; and Ford-Paz and colleagues (2019) found that certain cut-scores on parent- and youth-report measures (including the SCARED) considered together accurately predicted the likelihood of a diagnosis and eliminated the need for clinical interview administration. Utilizing both youth- and parent-reports and evaluating agreement versus discrepancy may also confer clinical benefits, including predicting treatment prognosis (e.g., Becker-Haimes, Jensen-Doss, Birmaher, Kendall, & Ginsburg, 2018; De Los Reyes et al., 2015). We will cover these and related issues in our forthcoming article that reviews parent-report measures of youth anxiety.

Other anxiety and related youth self-report measures.: Several additional and noteworthy anxiety self-report measures have been published since 2005 but did not meet inclusion criteria for this review. The Youth Anxiety Measure for DSM-5 (YAM-5) stands out because it uniquely assesses symptoms of all DSM-5 anxiety disorders, including major anxiety disorders and phobias. Studies to date, utilizing Dutch or Spanish versions, provide evidence of the YAM-5's internal consistency, test-retest reliability, convergent/divergent validity, and discriminative validity (Muris, Mannens, Peters, & Meesters, 2017; Muris, Simon, et al., 2017; Simon, Verboon, Smeekens, & Muris, 2017). Since selective mutism was first classified as an anxiety disorder only in DSM-5 (American Psychiatric Association, 2013), the inability to assess this disorder is a limitation of existing self-report measures that is addressed by the YAM-5. Another disorder-specific symptom measure, the Children's Separation Anxiety Scale (Méndez, Espada, Orgilés, Hidalgo, & Garcia-Fernández, 2008), also has evidence of test-retest reliability, discriminative validity, and treatment sensitivity, but to date only in studies of non-clinical Spanish youth. Future research on disorder-specific measures could examine if they contribute unique clinical benefit above and beyond existing multidimensional anxiety measures.

Other new measures assess anxiety-related impairment and interference. This is important because quantifying symptoms may not fully capture the degree of impairment experienced, and many youths suffer impairment related to their anxiety symptoms despite not meeting criteria for a disorder (Angold, Costello, Farmer, Burns, & Erkanli, 1999). The Child Anxiety Impact Scale (Langley, Bergman, McCracken, & Piacentini, 2004) has become widely used for assessing the impact of youth anxiety across various domains of functioning, with growing evidence for its psychometric properties (e.g., $\alpha_s = .70 - .90$; convergent validity with the SCARED and MASC; AUC = .81 for identifying diagnostic recovery; Evans et al., 2017; Langley et al., 2014). The Child Anxiety Life Interference Scale (Lyneham et al., 2013) is another measure that assesses anxiety-related impairment using a brief format (9-items) and has initial evidence of retest reliability, construct validity, and treatment sensitivity. The Family Accommodation Scale for Anxiety (Lebowitz et al., 2013) measures the degree of *accommodation* associated with youth anxiety symptoms, a construct which has garnered increasing empirical attention due to its associations with clinically

meaningful outcomes such as symptom severity. Two studies demonstrate the youth self-report measure's test-retest reliability and construct validity (Lebowitz, Marin, & Silverman, 2019; Lebowitz, Scharfstein, & Jones, 2015). Using such measures in conjunction with symptom self-report measures could help paint a more complete clinical picture of youth anxiety and add value by highlighting potential treatment targets (i.e., specific impairment areas or degree of accommodation).

Limitations—By using evaluative criteria to review youth anxiety self-report measures, our article makes an important and unique contribution to the current youth anxiety assessment literature. Focusing on self-report measures was a reasonable starting point (i.e., Part I) because they are widely used, versatile, and efficient. As noted, a companion piece covering parent versions of these measures (i.e., Part II) is forthcoming. Evaluating other commonly used assessment methods, such as diagnostic interviews (e.g., Anxiety Disorders Interview Schedule for Children - IV; Silverman & Albano, 1996), was beyond the scope of our review. We also did not review other versions of our focal measures (e.g., MASC-2) because psychometrics studies were limited in number or otherwise exclusionary. Importantly, the findings we report in this article appear to hold for these updated measures; we look forward to more research on these versions, especially as they address some of the points raised above (e.g., increased diversity in the RCMAS-2 normative sample; Reynolds & Richmond, 2008). We could not evaluate short-forms of measures for these same reasons; here as well we welcome more research, including tests of their incremental validity. Finally, we did not include psychometric evidence from studies using non-English versions of the measures. Reviews and meta-analyses of measures' cross-cultural properties can be consulted for this information (e.g., Scaini, Battaglia, Beidel, & Ogliari, 2012). Measure-specific meta-analyses may also be useful to consult for additional information about topics pertaining to these measures such as gender- and development-related differences (e.g., Runyon, Chesnut, & Burley, 2018).

Like any review, our article was bounded by the scope of the extant literature. Because no studies, as far as we know, have evaluated the repeatability, prescriptive validity, and clinical utility of these measures, these psychometric properties could not be covered. These properties would thus benefit from greater research attention; indeed, not only in the youth anxiety assessment area but in the youth assessment area more generally. Regarding clinical utility, we do note that practical considerations (i.e., length, cost) differ across the measures we evaluated and could impact measure selection. Practical considerations notwithstanding, the measures' psychometric properties reviewed in this article illuminate the ways that they can be most clinically useful.

Conclusion—Utilizing evaluative criteria, we reviewed the state of youth anxiety self-report assessment evidence base more rigorously and with greater confidence than was possible 15 years ago, when we last published such a review. The results are encouraging, as the majority of measures earned ratings of Good or Excellent with regard to most psychometric properties. We also identified new questions, many of which apply to the assessment field more broadly and are not unique to youth anxiety assessment. Consistent with the aims of the Evidence Base Updates Series (Southam-Gerow & Prinstein, 2014), we

hope another update comes in 4 to 5 years as opposed to 15 years, and that it will show significant progress toward addressing the gaps and limitations we highlighted in this review. At present, we are happy to conclude that most measures rest on sound psychometric data and can be used to fulfill several key assessment functions.

Acknowledgments

This study was supported by National Institute of Mental Health grants R01MH119299 and R61MH115113. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health.

References

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. Washington, DC, US: American Psychiatric Pub.
- Anderson ER, Jordan JA, Smith AJ, & Inderbitzen-Nolan HM (2009). An examination of the MASC social anxiety scale in a non-referred sample of adolescents. *Journal of Anxiety Disorders*, 23(8), 1098–1105. [PubMed: 19643571]
- Angold A, Costello EJ, Farmer EM, Burns BJ, & Erkanli A (1999). Impaired but undiagnosed. *Journal of the American Academy of Child & Adolescent Psychiatry*, 38(2), 129–137. [PubMed: 9951211]
- Bailey KA, Chavira DA, Stein MT, & Stein MB (2006). Brief measures to screen for social phobia in primary care pediatrics. *Journal of pediatric psychology*, 31(5), 512–521. [PubMed: 16034004]
- Baldwin JS, & Dadds MR (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(2), 252–260. [PubMed: 17242629]
- Becker-Haimes EM, Jensen-Doss A, Birmaher B, Kendall PC, & Ginsburg GS (2018). Parent–youth informant disagreement: Implications for youth anxiety treatment. *Clinical Child Psychology and Psychiatry*, 23(1), 42–56. [PubMed: 28191794]
- Behrens B, Swettlitz C, Pine DS, & Pagliaccio D (2019). The Screen for Child Anxiety Related Emotional Disorders (SCARED): Informant discrepancy, measurement invariance, and test–retest reliability. *Child Psychiatry & Human Development*, 50(3), 473–482. [PubMed: 30460424]
- Beidel DC (1996). Assessment of childhood social phobia: Construct, convergent, and discriminative validity of the Social Phobia and Anxiety inventory for Children (SPA-C). *Psychological Assessment*, 8(3), 235–240.
- Beidel DC, Turner SM, Hamlin K, & Morris TL (2000). The Social Phobia and Anxiety Inventory for Children (SPAI-C): External and discriminative validity. *Behavior Therapy*, 31(1), 75–87.
- Beidel DC, Turner SM, & Morris TL (1995). A new inventory to assess childhood social anxiety and phobia: The Social Phobia and Anxiety Inventory for Children. *Psychological Assessment*, 7(1), 73–79
- Beidel DC, Turner SM, Sallee FR, Ammerman RT, Crosby LA, & Pathak S (2007). SET-C versus fluoxetine in the treatment of childhood social phobia. *Journal of the American Academy of Child & Adolescent Psychiatry*, 46(12), 1622–1632. [PubMed: 18030084]
- Bennett K, Manassis K, Walter S, Cheung A, Wilansky-Traynor P, Diaz-Granados N, ... & Bodden D (2013). Cognitive behavioral therapy age effects in child and adolescent anxiety: An individual patient data meta-analysis. *Depression and Anxiety*, 30, 829–841. [PubMed: 23658135]
- Birmaher B, Brent DA, 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 & Adolescent Psychiatry*, 38(10), 1230–1236. [PubMed: 10517055]
- Birmaher B, Khetarpal S, Brent DA, Cully M, Balach L, Kaufman J, & Neer SM (1997). The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale construction and psychometric characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36(4), 545–553. [PubMed: 9100430]

- Blanton H, & Jaccard J (2006). Arbitrary metrics in psychology. *American Psychologist*, 61(1), 27–41.
- Bodden DH, Bögels SM, Nauta MH, De Haan E, Ringrose J, Appelboom C, ... & Appelboom-Geerts KC (2008). Child versus family cognitive-behavioral therapy in clinically anxious youth: An efficacy and partial effectiveness study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(12), 1384–1394. [PubMed: 18981932]
- Boyd RC, Ginsburg GS, Lambert SF, Cooley MR, & Campbell KD (2003). Screen for Child Anxiety Related Emotional Disorders (SCARED): Psychometric properties in an African-American parochial high school sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, 42(10), 1188–1196. [PubMed: 14560168]
- Bowers ME, Reider LB, Morales S, Buzzell GA, Miller N, Troller-Renfree SV, ... & Fox NA (2020). Differences in Parent and Child Report on the Screen for Child Anxiety-Related Emotional Disorders (SCARED): Implications for Investigations of Social Anxiety in Adolescents. *Journal of Abnormal Child Psychology*, 48(4), 561–571. [PubMed: 31853719]
- Brown RC, Yaroslavsky I, Quinoy AM, Friedman AD, Brookman RR, & Southam-Gerow MA (2013). Factor structure of measures of anxiety and depression symptoms in African American youth. *Child Psychiatry & Human Development*, 44(4), 525–536. [PubMed: 23224837]
- Brown-Jacobsen AM, Wallace DP, & Whiteside SP (2011). Multimethod, multi-informant agreement, and positive predictive value in the identification of child anxiety disorders using the SCAS and ADIS-C. *Assessment*, 18(3), 382–392. [PubMed: 20644080]
- Bunnell BE, Beidel DC, Liu L, Joseph DL, & Higa-McMillan C (2015). The SPAIC-11 and SPAICP-11: Two brief child-and parent-rated measures of social anxiety. *Journal of Anxiety Disorders*, 36, 103–109. [PubMed: 26500188]
- Byrne SP, Lebowitz E, Ollendick TH, & Silverman WK (2018). Child and adolescent anxiety disorders. In Hunsley J & Mash E (Eds.). *A guide to assessments that work* (2nd Edition.). New York: Oxford University Press (pp. 217–241).
- Cannon CJ, Makol BA, Keeley LM, Qasmieh N, Okuno H, Racz SJ, & De Los Reyes A (2020). A paradigm for understanding adolescent social anxiety with unfamiliar peers: Conceptual foundations and directions for future research. *Clinical Child and Family Psychology Review*. doi: 10.1007/s10567-020-00314-4
- Caporino NE, Sakolsky D, Brodman DM, McGuire JF, Piacentini J, Peris TS, ... & Birmaher B (2017). Establishing clinical cutoffs for response and remission on the Screen for Child Anxiety Related Emotional Disorders (SCARED). *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(8), 696–702. [PubMed: 28735699]
- Carter R, Silverman WK, Allen A, & Ham L (2008). Measures matter: The relative contribution of anxiety and depression to suicidal ideation in clinically referred anxious youth using brief versus full length questionnaires. *Depression and Anxiety*, 25(8), E27–E35.
- Clark DB, Turner SM, Beidel DC, Donovan JE, Kirisci L, & Jacob RG (1994). Reliability and validity of the Social Phobia and Anxiety Inventory for adolescents. *Psychological Assessment*, 6(2), 135–140.
- Cohen J (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
- Copeland WE, Angold A, Shanahan L, & Costello EJ (2014). Longitudinal patterns of anxiety from childhood to adulthood: The Great Smoky Mountains Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(1), 21–33. [PubMed: 24342383]
- Copeland WE, Shanahan L, Costello EJ, & Angold A (2009). Childhood and adolescent psychiatric disorders as predictors of young adult disorders. *Archives of General Psychiatry*, 66(7), 764–772. [PubMed: 19581568]
- Creswell C, & Cartwright-Hatton S (2007). Family treatment of child anxiety: Outcomes, limitations and future directions. *Clinical Child and Family Psychology Review*, 10(3), 232–252. [PubMed: 17476594]
- Creswell C, Nauta M Hudson J, March S Reardon T, Arendt K, ... & Kendall PC (2020). Recommendations for reporting on treatment trials for child and adolescent anxiety disorders: an international consensus statement. *Journal of Child Psychology and Psychiatry*.

- Davis TE, May A, & Whiting SE (2011). Evidence-based treatment of anxiety and phobia in children and adolescents: Current status and effects on the emotional response. *Clinical Psychology Review*, 31, 592–602. [PubMed: 21482318]
- 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 & Adolescent Psychology*, 40(1), 1–9. [PubMed: 21229439]
- De Los Reyes A, Augenstein TM, Wang M, Thomas SA, Drabick DA, Burgers DE, & Rabinowitz J (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141(4), 858. [PubMed: 25915035]
- De Los Reyes A, & Langer DA (2018). Assessment and the journal of clinical child and adolescent psychology's evidence base updates series: Evaluating the tools for gathering evidence. *Journal of Clinical Child & Adolescent Psychology*, 47(3), 357–365. [PubMed: 29768054]
- Deros DE, Racz SJ, Lipton MF, Augenstein TM, Karp JN, Keeley LM, ... & De Los Reyes A (2018). Multi-informant assessments of adolescent social anxiety: Adding clarity by leveraging reports from unfamiliar peer confederates. *Behavior Therapy*, 49(1), 84–98. [PubMed: 29405924]
- Dierker LC, Albano AM, Clarke GN, Heimberg RG, Kendall PC, Merikangas KR, ... & Kupfer DJ (2001). Screening for anxiety and depression in early adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(8), 929–936 [PubMed: 11501693]
- Epkins CC (2002). A comparison of two self-report measures of children's social anxiety in clinical and community samples. *Journal of Clinical Child & Adolescent Psychology*, 31(1), 69–79. [PubMed: 11845653]
- Evans R, Thirlwall K, Cooper P, & Creswell C (2017). Using symptom and interference questionnaires to identify recovery among children with anxiety disorders. *Psychological Assessment*, 29(7), 835. [PubMed: 27845527]
- Ford-Paz RE, Gouze KR, Kerns CE, Ballard R, Parkhurst JT, Jha P, & Lavigne J (2019). Evidence-based assessment in clinical settings: Reducing assessment burden for a structured measure of child and adolescent anxiety. *Psychological Services*. doi: 10.1037/ser0000367
- Gallagher HM, Rabian BA, & McCloskey MS (2004). A brief group cognitive-behavioral intervention for social phobia in childhood. *Journal of Anxiety Disorders*, 18(4), 459–479. [PubMed: 15149708]
- Ginsburg GS, La Greca AM, & Silverman WK (1998). Social anxiety in children with anxiety disorders: Relation with social and emotional functioning. *Journal of Abnormal Child Psychology*, 26(3), 175–185. [PubMed: 9650624]
- Gonzalez A, Weersing VR, 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 & Adolescent Psychology*, 41(3), 361–369. [PubMed: 22397682]
- Grills-Taquechel AE, Ollendick TH, & Fisak B (2008). Reexamination of the MASC factor structure and discriminant ability in a mixed clinical outpatient sample. *Depression and Anxiety*, 25(11), 942–950. [PubMed: 18008335]
- Haynes SN, & Lench HC (2003). Incremental validity of new clinical assessment measures. *Psychological assessment*, 15(4), 456. [PubMed: 14692842]
- Higa CK, Fernandez SN, Nakamura BJ, Chorpita BF, & Daleiden EL (2006). Parental assessment of childhood social phobia: Psychometric properties of the Social Phobia and Anxiety Inventory for Children–Parent Report. *Journal of Clinical Child & Adolescent Psychology*, 35(4), 590–597. [PubMed: 17007605]
- Herbert JD, Gaudiano BA, Rheingold AA, Moitra E, Myers VH, Dalrymple KL, & Brandsma LL (2009). Cognitive behavior therapy for generalized social anxiety disorder in adolescents: A randomized controlled trial. *Journal of Anxiety Disorders*, 23(2), 167–177. [PubMed: 18653310]
- Hodges K (1990). Depression and anxiety in children: A comparison of self-report questionnaires to clinical interview. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2(4), 376–381.

- Holly LE, Little M, Pina AA, & Caterino LC (2015). Assessment of anxiety symptoms in school children: A cross-sex and ethnic examination. *Journal of Abnormal Child Psychology*, 43(2), 297–309. [PubMed: 2499313]
- Hunsley J, & Mash EJ (Eds.). (2008). *A guide to assessments that work*. New York, NY: Oxford University Press.
- Inderbitzen HM, & Hope DA (1995). Relationship among adolescent reports of social anxiety, anxiety, and depressive symptoms. *Journal of Anxiety Disorders*, 9(5), 385–396.
- Inderbitzen-Nolan H, Davies CA, & McKeon ND (2004). Investigating the construct validity of the SPAI-C: Comparing the sensitivity and specificity of the SPAI-C and the SAS-A. *Journal of Anxiety Disorders*, 18(4), 547–560. [PubMed: 15149713]
- Inderbitzen-Nolan HM, & Walters KS (2000). Social Anxiety Scale for Adolescents: Normative data and further evidence of construct validity. *Journal of Clinical Child Psychology*, 29(3), 360–371. [PubMed: 10969420]
- Ingul JM, Aune T, & Nordahl HM (2014). A randomized controlled trial of individual cognitive therapy, group cognitive behaviour therapy and attentional placebo for adolescent social phobia. *Psychotherapy and Psychosomatics*, 83(1), 54–61. [PubMed: 24281563]
- Johnston C, & Murray C (2003). Incremental validity in the psychological assessment of children and adolescents. *Psychological Assessment*, 15(4), 496. [PubMed: 14692845]
- Kazdin AE (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, 1(4), 427–452.
- Kearney CA, & Silverman WK (1999). Functionally based prescriptive and nonprescriptive treatment for children and adolescents with school refusal behavior. *Behavior Therapy*, 30(4), 673–695.
- Kingery JN, Ginsburg GS, & Burstein M (2009). Factor structure and psychometric properties of the Multidimensional Anxiety Scale for Children in an African American adolescent sample. *Child Psychiatry & Human Development*, 40(2), 287–300. [PubMed: 19165593]
- La Greca AM, Ingles CJ, Lai BS, & Marzo JC (2015). Social anxiety scale for adolescents: Factorial invariance across gender and age in Hispanic American adolescents. *Assessment*, 22(2), 224–232. [PubMed: 25059683]
- La Greca AM, & Lopez N (1998). Social anxiety among adolescents: Linkages with peer relations and friendships. *Journal of Abnormal Child Psychology*, 26(2), 83–94. [PubMed: 9634131]
- La Greca AM, & Stone WL (1993). Social Anxiety Scale for Children—Revised: Factor structure and concurrent validity. *Journal of Clinical Child Psychology*, 22(1), 17–27.
- Landis JR, & Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. [PubMed: 843571]
- Lang PJ (1968). Fear reduction and fear behavior: Problems in treating a construct. In Shlien JM (Ed.), *Research in psychotherapy* (Vol. III, pp. 90–103). Washington, DC: American Psychological Association.
- Langley AK, Bergman RL, McCracken J, & Piacentini JC (2004). Impairment in childhood anxiety disorders: Preliminary examination of the child anxiety impact scale—parent version. *Journal of Child and Adolescent Psychopharmacology*, 14(1), 105–114. [PubMed: 15142397]
- Langley AK, Falk A, Peris T, Wiley JF, Kendall PC, Ginsburg G, ... & Piacentini J (2014). The child anxiety impact scale: examining parent-and child-reported impairment in child anxiety disorders. *Journal of Clinical Child & Adolescent Psychology*, 43(4), 579–591. [PubMed: 23915200]
- Last CG, Francis G, & Strauss CC (1989). Assessing fears in anxiety-disordered children with the Revised Fear Survey Schedule for Children (FSSC-R). *Journal of Clinical Child Psychology*, 18(2), 137–141.
- Lebowitz ER, & François B (2018). Using motion tracking to measure avoidance in children and adults: Psychometric properties, associations with clinical characteristics, and treatment-related change. *Behavior Therapy*, 49(6), 853–865. [PubMed: 30316485]
- Lebowitz ER, Gee DG, Pine DS, & Silverman WK (2018). Implications of the research domain criteria project for childhood anxiety and its disorders. *Clinical Psychology Review*, 64, 99–109. [PubMed: 29526524]
- Lebowitz ER, Marin C, Martino A, Shimshoni Y, & Silverman WK (2020). Parent-based treatment as efficacious as cognitive-behavioral therapy for childhood anxiety: A randomized noninferiority

- study of supportive parenting for anxious childhood emotions. *Journal of the American Academy of Child & Adolescent Psychiatry*, 59(3), 362–372. [PubMed: 30851397]
- Lebowitz ER, Marin CE, & Silverman WK (2019). Measuring family accommodation of childhood anxiety: Confirmatory factor analysis, validity, and reliability of the parent and child Family Accommodation Scale–Anxiety. *Journal of Clinical Child & Adolescent Psychology*, 1–9.
- Lebowitz ER, Scharfstein L, & Jones J (2015). Child-report of family accommodation in pediatric anxiety disorders: Comparison and integration with mother-report. *Child Psychiatry & Human Development*, 46(4), 501–511. [PubMed: 25209390]
- Lebowitz ER, Shic F, Campbell D, Basile K, & Silverman WK (2015). Anxiety sensitivity moderates behavioral avoidance in anxious youth. *Behaviour Research Therapy*, 74, 11–17. [PubMed: 26348546]
- Lebowitz ER, Woolston J, Bar-Haim Y, Calvocoressi L, Dauser C, Warnick E, ... & Vitulano LA (2013). Family accommodation in pediatric anxiety disorders. *Depression and Anxiety*, 30(1), 47–54. [PubMed: 22965863]
- Lyneham HJ, Sbrulati ES, Abbott MJ, Rapee RM, Hudson JL, Tolin DF, & Carlson SE (2013). Psychometric properties of the child anxiety life interference scale (CALIS). *Journal of Anxiety Disorders*, 27(7), 711–719. [PubMed: 24135256]
- Manassis K, Lee TC, Bennett K, Zhao XY, Mendlowitz S, Duda S, ... & Bodden D (2014). Types of parental involvement in CBT with anxious youth: a preliminary meta-analysis. *Journal of Consulting and Clinical Psychology*, 82(6), 1163–1172. [PubMed: 24841867]
- March JS (2013). *Multidimensional anxiety scale for children 2nd edition (MASC 2)*. Toronto, Canada: Multi-Health Systems.
- March JS, Parker JDA, Sullivan K, Stallings P, & Conners K (1997). The Multidimensional Anxiety Scale for Children (MASC): Factor, structure, reliability, and validity. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36(4), 554–565. [PubMed: 9100431]
- March JS, Sullivan K, & Parker (1999). Test–retest reliability of the multidimensional anxiety scale for children. *Journal of Anxiety Disorders*, 13(4), 349–358. [PubMed: 10504106]
- Masia Warner C, Fisher PH, Shrout PE, Rathor S, & Klein RG (2007). Treating adolescents with social anxiety disorder in school: An attention control trial. *Journal of Child Psychology and Psychiatry*, 48(7), 676–686. [PubMed: 17593148]
- Mattison RE, Bagnato SJ, & Brubaker BM (1988). Diagnostic utility of the Revised Children's Manifest Anxiety Scale in children with *DSM-III* anxiety disorders. *Journal of Anxiety Disorders*, 2(2), 147–155.
- McDonald RP (1999). *Test theory: A unified treatment*. Hillsdale: Erlbaum.
- McLaughlin EN, Stewart SH, & Taylor S (2007). Childhood anxiety sensitivity index factors predict unique variance in DSM-IV anxiety disorder symptoms. *Cognitive Behaviour Therapy*, 36(4), 210–219. [PubMed: 18049946]
- Méndez X, Espada JP, Orgilés M, Hidalgo MD, & García-Fernández JM (2008). Psychometric properties and diagnostic ability of the Separation Anxiety Scale for Children (SASC). *European Child & Adolescent Psychiatry*, 17(6), 365–372. [PubMed: 18427868]
- Merikangas KR, He J. p., Burstein M, Swanson SA, Avenevoli S, Cui L, . . . Swendsen J (2010). Lifetime prevalence of mental disorders in US adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(10), 980–989. [PubMed: 20855043]
- Monga S, Birmaher B, Chiappetta L, Brent D, Kaufman J, Bridge J, & Cully M (2000). Screen for child anxiety-related emotional disorders (SCARED): Convergent and divergent validity. *Depression and Anxiety*, 12(2), 85–91.
- Moreno J, Silverman WK, Saavedra LM, & Phares V (2008). Fathers' ratings in the assessment of their child's anxiety symptoms: A comparison to mothers' ratings and their associations with paternal symptomatology. *Journal of Family Psychology*, 22, 915–919. [PubMed: 19102612]
- Muris P, Mannens J, Peters L, & Meesters C (2017). The Youth Anxiety Measure for DSM-5 (YAM-5): Correlations with anxiety, fear, and depression scales in non-clinical children. *Journal of Anxiety Disorders*, 51, 72–78. [PubMed: 28668214]

- Muris P, Merckelbach H, Ollendick T, King N, & Bogie N (2002). Three traditional and three new childhood anxiety questionnaires: Their reliability and validity in a normal adolescent sample. *Behaviour Research and Therapy*, 40(7), 753–772. [PubMed: 12074371]
- Muris P, Merckelbach H, Schmidt H, & Mayer B (1998). The revised version of the Screen for Child Anxiety Related Emotional Disorders (SCARED-R): Factor structure in normal children. *Personality and Individual Differences*, 26(1), 99–112.
- Muris P, Ollendick TH, Roelofs J, & Austin K (2014). The Short Form of the Fear Survey Schedule for Children-Revised (FSSC-R-SF): An efficient, reliable, and valid scale for measuring fear in children and adolescents. *Journal of Anxiety Disorders*, 28(8), 957–965. [PubMed: 25445086]
- Muris P, Simon E, Lijphart H, Bos A, Hale W, & Schmeitz K (2017). The youth anxiety measure for DSM-5 (YAM-5): development and first psychometric evidence of a new scale for assessing anxiety disorders symptoms of children and adolescents. *Child Psychiatry & Human Development*, 48(1), 1–17. [PubMed: 27179521]
- Ollendick TH (1983). Reliability and validity of the Revised Fear Survey Schedule for Children (FSSC-R). *Behaviour Research and Therapy*, 21(6), 395–399.
- Ollendick TH, Öst LG, Reuterskiöld L, Costa N, Cederlund R, Sirbu C, ... & Jarrett MA (2009). One-session treatment of specific phobias in youth: a randomized clinical trial in the United States and Sweden. *Journal of Consulting and Clinical Psychology*, 77(3), 504. [PubMed: 19485591]
- Ollendick TH, Yule W, & Ollier K (1991). Fears in British children and their relationship to manifest anxiety and depression. *Child Psychology & Psychiatry & Allied Disciplines*, 32(2), 321–331.
- Osman A, Williams JE, Espenschade K, Gutierrez PM, Bailey JR, & Chowdhry O (2009). Further evidence of the reliability and validity of the multidimensional anxiety. *Journal of Psychopathology and Behavioral Assessment*, 31(3), 202–214.
- Papay JP, & Spielberger CD (1986). Assessment of anxiety and achievement in kindergarten and first- and second-grade children. *Journal of Abnormal Child Psychology*, 14(2), 279–286. [PubMed: 3722623]
- Perrin S, & Last CG (1992). Do childhood anxiety measures measure anxiety? *Journal of Abnormal Child Psychology*, 20(6), 567–578. [PubMed: 1487597]
- Pina AA, Little M, Knight GP, & Silverman WK (2009). Cross-ethnic measurement equivalence of the RCMAS in Latino and White youth with anxiety disorders. *Journal of Personality Assessment*, 91(1), 58–61. [PubMed: 19085284]
- Pina AA, Little M, Wynne H, & Beidel DC (2014). Assessing social anxiety in African American youth using the social phobia and anxiety inventory for children. *Journal of Abnormal Child Psychology*, 42(2), 311–320. [PubMed: 23872906]
- Pina AA, Silverman WK, Alfano CA, & Saavedra LM (2002). Diagnostic efficiency of symptoms in the diagnosis of DSM-IV: generalized anxiety disorder in youth. *Journal of Child Psychology and Psychiatry*, 43(7), 959–967. [PubMed: 12405483]
- Pina AA, Silverman WK, Saavedra LM, & Weems CF (2001). An analysis of the RCMAS Lie Scale Scores in an ethnic sample of anxious children. *Journal of Anxiety Disorders*, 15, 443–458. [PubMed: 11583076]
- Rachman S (1978). Human fears: A three systems analysis. *Cognitive Behaviour Therapy*, 7(4), 237–245.
- Rapee RM, Lyneham HJ, Hudson JL, Kangas M, Wuthrich VM, & Schniering CA (2013). Effect of comorbidity on treatment of anxious children and adolescents: Results from a large, combined sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(1), 47–56. [PubMed: 23265633]
- Rappaport BI, Pagliaccio D, Pine DS, Klein DN, & Jarcho JM (2017). Discriminant validity, diagnostic utility, and parent-child agreement on the Screen for Child Anxiety Related Emotional Disorders (SCARED) in treatment- and non-treatment-seeking youth. *Journal of Anxiety Disorders*, 51, 22–31. [PubMed: 28886420]
- Reardon T, Creswell C, Lester KJ, Arendt K, Blatter-Meunier J, Bögels SM, ... & Hogendoorn SM (2019). The utility of the SCAS-C/P to detect specific anxiety disorders among clinically anxious children. *Psychological Assessment*, 31(8), 1006–1018. [PubMed: 31070449]

- Reardon T, Spence SH, Hesse J, Shakir A, & Creswell C (2018). Identifying children with anxiety disorders using brief versions of the Spence Children's Anxiety Scale for children, parents, and teachers. *Psychological assessment*, 30(10), 1342. [PubMed: 29902050]
- Revelle W, & Condon DM (2019). Reliability from α to ω : A tutorial. *Psychological Assessment*, 31(12), 1395. [PubMed: 31380696]
- Reynolds CR (1981). Long-term stability of scores on the Revised-Children's Manifest Anxiety Scale. *Perceptual and Motor Skills*, 53(3), 702. [PubMed: 7322774]
- Reynolds CR, & Paget KD (1983). National normative and reliability data for the Revised Children's Manifest Anxiety Scale. *School Psychology Review*, 12(3), 324–336.
- Reynolds CR, & Richmond BO (1978). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology*, 6(2), 271–280. [PubMed: 670592]
- Reynolds CR, & Richmond BO (2008). *Revised Children's Manifest Anxiety Scale: (RCMAS-2)*. Torrance, CA: Western Psychological Services.
- Runyon K, Chesnut SR, & Burley H (2018). Screening for childhood anxiety: A meta-analysis of the Screen for Child Anxiety Related Emotional Disorders. *Journal of Affective Disorders*, 240, 220–229. [PubMed: 30081293]
- Rynn MA, Barber JP, Khalid-Khan S, Siqueland L, Dembiski M, McCarthy KS, & Gallop R (2006). The psychometric properties of the MASC in a pediatric psychiatric sample. *Journal of Anxiety Disorders*, 20(2), 139–157. [PubMed: 16464701]
- Scaini S, Battaglia M, Beidel DC, & Ogliari A (2012). A meta-analysis of the cross-cultural psychometric properties of the Social Phobia and Anxiety Inventory for Children (SPAI-C). *Journal of Anxiety Disorders*, 26(1), 182–188. [PubMed: 22154123]
- Shimshoni Y, Silverman WK, & Lebowitz ER (2017). Maternal acceptance moderates fear ratings and avoidance behavior in children. *Child Psychiatry & Human Development*, 49 (3), 460–467.
- Skriner LC, & Chu BC (2014). Cross-ethnic measurement invariance of the SCARED and CES-D in a youth sample. *Psychological Assessment*, 26(1), 332. [PubMed: 24245992]
- Silverman WK, & Albano AM (1996). *Anxiety Disorders Interview Schedule for Children-IV (Child and Parent Versions)*. New York, NY: Oxford University Press.
- Silverman WK, Fleisig W, Rabian B, & Peterson RA (1991). Childhood anxiety sensitivity index. *Journal of Clinical Child & Adolescent Psychology*, 20(2), 162–168.
- Silverman WK, & Kurtines WM (1996). *Anxiety and phobic disorders: A pragmatic approach*. Springer Science & Business Media.
- Silverman WK, Kurtines WM, Ginsburg GS, Weems CF, Lumpkin PW, & Carmichael DH (1999a). Treating anxiety disorders in children with group cognitive-behavioral therapy: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 67(6), 995. [PubMed: 10596522]
- Silverman WK, Kurtines WM, Ginsburg GS, Weems CF, Rabian B, & Serafini LT (1999b). Contingency management, self-control, and education support in the treatment of childhood phobic disorders: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 67(5), 675. [PubMed: 10535234]
- Silverman WK, Marin CE, Rey Y, Kurtines WM, Jaccard J, & Pettit JW (2019). Group-versus parent-involvement CBT for childhood anxiety disorders: Treatment specificity and long-term recovery mediation. *Clinical Psychological Science*, 7(4), 840–855.
- Silverman WK, & Ollendick TH (2005). Evidence-based assessment of anxiety and its disorders in children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, 34(3), 380–411. [PubMed: 16026211]
- Simon E, Bos AE, Verboon P, Smeekens S, & Muris P (2017). Psychometric properties of the Youth Anxiety Measure for DSM-5 (YAM-5) in a community sample. *Personality and Individual Differences*, 116, 258–264.
- Southam-Gerow MA, & Prinstein MJ (2014). Evidence base updates: The evolution of the evaluation of psychological treatments for children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, 43(1), 1–6. [PubMed: 24294989]
- Spence SH (1998). A measure of anxiety symptoms among children. *Behaviour Research and Therapy*, 36(5), 545–566. [PubMed: 9648330]

- Spence SH (2018). Assessing anxiety disorders in children and adolescents. *Child and Adolescent Mental Health*, 23(3), 266–282. [PubMed: 32677290]
- Spence SH, Barrett PM, & Turner CM (2003). Psychometric properties of the Spence Children's Anxiety Scale with young adolescents. *Journal of Anxiety Disorders*, 17(6), 605–625. [PubMed: 14624814]
- Spence SH, Donovan CL, March S, Kenardy JA, & Hearn CS (2017). Generic versus disorder specific cognitive behavior therapy for social anxiety disorder in youth: A randomized controlled trial using internet delivery. *Behaviour Research and Therapy*, 90, 41–57. [PubMed: 27988427]
- Spielberger CD (1973). *State-trait anxiety inventory for children*. Consulting Psychologists Press.
- Storch EA, Eisenberg PS, Roberti JW, & Barlas ME (2003). Reliability and validity of the Social Anxiety Scale for Children--Revised for Hispanic children. *Hispanic Journal of Behavioral Sciences*, 25(3), 410–422.
- Storch EA, Masia-Warner C, Dent HC, Roberti JW, & Fisher PH (2004). Psychometric evaluation of the Social Anxiety Scale for Adolescents and the Social Phobia and Anxiety Inventory for Children: Construct validity and normative data. *Journal of Anxiety Disorders*, 18(5), 665–679. [PubMed: 15275945]
- Strauss CC, Last CG, Hersen M, & Kazdin AE (1988). Association between anxiety and depression in children and adolescents with anxiety disorders. *Journal of Abnormal Child Psychology*, 16(1), 57–68. [PubMed: 3361030]
- Tulbure BT, Szentagotai A, Dobrean A, & David D (2012). Evidence based clinical assessment of child and adolescent social phobia: a critical review of rating scales. *Child Psychiatry & Human Development*, 43(5), 795–820. [PubMed: 22438106]
- Varela ER, & Biggs BK (2006). Reliability and validity of the Revised Children's Manifest Anxiety Scale (RCMAS) across samples of Mexican, Mexican American, and European American children: a preliminary investigation. *Anxiety, Stress, and Coping*, 19(1), 67–80.
- Varela RE, Sanchez-Sosa JJ, Biggs BK, & Luis TM (2008). Anxiety symptoms and fears in Hispanic and European American children: Cross-cultural measurement equivalence. *Journal of Psychopathology and Behavioral Assessment*, 30(2), 132–145.
- Viana AG, Rabian B, & Beidel DC (2008). Self-report measures in the study of comorbidity in children and adolescents with social phobia: Research and clinical utility. *Journal of Anxiety Disorders*, 22(5), 781–792. [PubMed: 17888622]
- Walkup JT, Labellarte MJ, Riddle MA, Pine DS, Greenhill L, Klein R, ... & Klee B (2001). Fluvoxamine for the treatment of anxiety disorders in children and adolescents. *New England Journal of Medicine*, 344(17), 1279–1285.
- Watson D (2004). Stability versus change, dependability versus error: Issues in the assessment of personality over time. *Journal of Research in Personality*, 38(4), 319–350.
- Weems CF, Hammond-Laurence K, Silverman WK, & Ginsburg GS (1998). Testing the utility of the anxiety sensitivity construct in children and adolescents referred for anxiety disorders. *Journal of Clinical Child Psychology*, 27(1), 69–77. [PubMed: 9561939]
- Weems CF, Silverman WK, Saavedra LM, Pina AA, & Lumpkin PW (1999). The discrimination of children's phobias using the Revised Fear Survey Schedule for Children. *Journal of Child Psychology and Psychiatry*, 40(6), 941–952. [PubMed: 10509888]
- Wei C, Hoff A, Villabø MA, Peterman J, Kendall PC, Piacentini J, ... March J (2014). Assessing anxiety in youth with the Multidimensional Anxiety Scale for Children (MASC). *Journal of Clinical Child & Adolescent Psychology*, 43(4), 566–578. [PubMed: 23845036]
- Weisz JR, Jensen Doss A, & Hawley KM (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology*, 56, 337–363.
- Whiteside SP, & Brown AM (2008). Exploring the utility of the Spence Children's Anxiety Scales parent-and child-report forms in a North American sample. *Journal of Anxiety Disorders*, 22(8), 1440–1446. [PubMed: 18395408]
- Wisniewski JJ, Genshaft JL, Mulick JA, & Coury DL (1987). Test-retest reliability of the revised children's manifest anxiety scale. *Perceptual and Motor Skills*, 65(1), 67–70. [PubMed: 3684472]
- Wood JJ, Piacentini JC, Bergman RL, McCracken J, & Barrios V (2002). Concurrent validity of the anxiety disorders section of the Anxiety Disorders Interview Schedule for DSM-IV: Child and

- Parent Versions. *Journal of Clinical Child & Adolescent Psychology*, 31(3), 335–342. [PubMed: 12149971]
- Wood JJ, Piacentini JC, Southam-Gerow M, Chu BC, & Sigman M (2006). Family cognitive behavioral therapy for child anxiety disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 45(3), 314–321. [PubMed: 16540816]
- Wren FJ, Berg EA, Heiden LA, Kinnamon CJ, Ohlson LA, Bridge JA, ... & Bernal MP (2007). Childhood anxiety in a diverse primary care population: parent-child reports, ethnicity and SCARED factor structure. *Journal of the American Academy of Child & Adolescent Psychiatry*, 46(3), 332–340. [PubMed: 17314719]
- Youngstrom EA, Van Meter A, Frazier TW, Hunsley J, Prinstein MJ, Ong ML, & Youngstrom JK (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, 24(4), 331–363.
- Youngstrom EA, Salcedo S, Frazier TW, & Perez Algorta G (2019). Is the finding too good to be true? Moving from “more is better” to thinking in terms of simple predictions and credibility. *Journal of Clinical Child & Adolescent Psychology*, 48(6), 811–824. [PubMed: 31618104]

Table 1:

Rubric for evaluating norms, validity, and utility (De Los Reyes & Langer, 2018)

Criterion	Adequate	Good	Excellent
Norms	<i>M</i> and <i>SD</i> for total score (and subscores if relevant) from a large, relevant clinical sample	<i>M</i> and <i>SD</i> for total score (and subscores if relevant) from multiple large, relevant samples, at least one clinical and one nonclinical	Same as "good," but must be from representative sample (i.e., random sampling, or matching to census data)
Internal consistency (Cronbach's alpha, split half, etc.)	Most evidence shows alpha values of 0.70–0.79	Most reported alphas 0.80–0.89	Most reported alphas 0.90
Inter-rater reliability	Most evidence shows kappas of 0.60–0.74, or ICCs of 0.70–0.79	Most reported kappas of 0.75–0.84, ICCs of 0.80–0.89	Most kappas 0.85, or ICCs 0.90
Test–retest reliability (stability)	Most evidence shows test–retest correlations 0.70 over period of several days or weeks	Most evidence shows test–retest correlations 0.70 over period of several months	Most evidence shows test–retest correlations 0.70 over 1 year or longer
Repeatability	Bland–Altman (Bland Altman, 1986) plots show small bias, and/or weak trends; coefficient of repeatability is tolerable compared to clinical benchmarks (Vaz, Falkner, Passmore, Parsons, Andreou, 2013)	Bland–Altman plots and corresponding regressions show no significant bias, and no significant trends; coefficient of repeatability is tolerable	Bland–Altman plots and corresponding regressions show no significant bias, and no significant trends; established for multiple studies; coefficient of repeatability is small enough that it is not clinically concerning
Content validity	Test developers clearly defined domain and ensured representation of entire set of facets	Same as "adequate," plus all elements (items, instructions) evaluated by judges (experts or pilot participants)	Same as "good," plus multiple groups of judges and quantitative ratings
Construct validity (e.g., predictive, concurrent, convergent, and discriminant validity)	Some independently replicated evidence of construct validity	Bulk of independently replicated evidence shows multiple aspects of construct validity	Same as "good," plus evidence of incremental validity with respect to other clinical data
Discriminative validity	Statistically significant discrimination in multiple samples; AUCs <0.6 under clinically realistic conditions (i.e., not comparing treatment seeking and healthy youth)	AUCs of 0.60 to <0.75 under clinically realistic conditions	AUCs of 0.75 to 0.90 under clinically realistic conditions
Prescriptive validity	Statistically significant accuracy at identifying a diagnosis with a well-specified matching intervention, or statistically significant moderator of treatment	Same as "adequate," with good kappa for diagnosis, or significant treatment moderation in more than one sample	Same as "good," with good kappa for diagnosis in more than one sample, or moderate effect size for treatment moderation
Validity generalization	Some evidence supports use with either more than one specific demographic group or in more than one setting	Bulk of evidence supports use with either more than one specific demographic group or in multiple settings	Bulk of evidence supports use with either more than one specific demographic group AND in multiple settings
Treatment sensitivity	Some evidence of sensitivity to change over course of treatment	Independent replications show evidence of sensitivity to change over course of treatment	Same as "good," plus sensitive to change across different types of treatments
Clinical utility	After practical considerations (e.g., costs, respondent burden, ease of administration and scoring, availability of relevant benchmark scores, patient acceptability), assessment data are likely to be clinically actionable	Same as "adequate," plus published evidence that using the assessment data confers clinical benefit (e.g., better outcome, lower attrition, greater satisfaction), in areas important to stakeholders	Same as "good," plus independent replication

Note: ICC = intraclass correlation coefficient; AUC = area under the curve. Table reproduced with permission.

Table 2

Summary Information for Youth Anxiety Self-Report Measures

Measure & citation	Age range (years)	Number of items & response scale	Subscales	Translations	Open access?
State Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973)	8-14	40-items Response scale: 1 (hardly/lowest degree of feelings) – 3 (often/highest degree of feelings)	Trait and State scales (20 items each)	At least 27 languages (per mindingarden.com)	No
Revised Children's Manifest Anxiety Scale (RCMAS; Reynolds & Richmond, 1978) RCMAS-2 (Reynolds & Richmond, 2008)	6-19	RCMAS: 37-items RCMAS-2: 49-items Response scale: Yes/No	RCMAS: Physiological Anxiety, Worry/Oversensitivity, Social Concerns/Concentration, Lie Scale RCMAS-2: Physiological Anxiety, Worry, Social Anxiety, Defensiveness scale	Arabic, Bulgarian Chinese, French-Canadian, German, Korean, Malay, Portuguese, Spanish, Urdu, Xhosa	No
Fear Survey Schedule for Children-Revised (FSSC-R; Ollendick, 1983)	7-18	80-items Response scale: 1 (none) – 3 (a lot)	Fear of Failure and Criticism; Fear of Danger and Death; Fear of Small Animals; Medical Fears; Fear of the Unknown	Arabic, Afrikaans/Xhosa, Chinese, Dutch, Farsi, Greek, Hebrew Italian, Spanish, Swedish, Turkish	Yes
Social Anxiety Scale for Children - Revised (SASC-R; La Greca & Stone, 1993) Social Anxiety Scale for Adolescents (SAS-A; La Greca & Lopez, 1998)	8-18	SASC-R: 26-items SAS-A: 26-items (Both versions include 4 filler items) Response scale: 1 (not at all) – 5 (all the time)	SASC-R/SAS-A: Fear of Negative Evaluation; Social Avoidance and Distress in New Situations; General Social Avoidance and Distress	Belgian, Chinese, Dutch, Finnish, French, German, Greek, Portuguese, Spanish	Yes
Social Phobia and Anxiety Inventory for Children (SPAI-C; Beidel et al., 1995)	8-17	26-items Response scale: 0 (never or hardly ever) – 2 (most of the time or always)	Assertiveness/General Conversation; Traditional Social Encounters; Public Performance	German, Finnish, Icelandic, Italian, Norwegian, Portuguese, Spanish, Swedish	No
Multidimensional Anxiety Scale for Children (MASC; March et al., 1997) MASC-2 (March, 2013)	8-19	MASC: 39-items MASC-2: 50-items Response scale: 0 (never) – 3 (often)	MASC: Social Anxiety, Separation Anxiety/Panic, Harm Avoidance, Physical Symptoms MASC-2: Social Anxiety, Separation Anxiety/Phobias, Obsessions and Compulsions, Harm Avoidance, Physical Symptoms, GAD Index	Chinese, Dutch, Icelandic, Norwegian, Persian, Spanish, Swedish, Taiwanese	No
Screen for Child Anxiety Related Disorders (SCARED; Birmaher et al., 1997, 1999) SCARED-R (Muris et al., 1998)	8-18	SCARED: 41-items SCARED-R: 66-items Response scale: 0 (not true or hardly ever true) – 2 (very true or often true)	SCARED: Social Anxiety; Separation Anxiety; Generalized Anxiety; Panic Disorder/Significant Somatic Symptoms; Significant School Avoidance SCARED-R: Panic Disorder, Separation Anxiety Disorder, Generalized Anxiety Disorder, Social Phobia, Specific Phobias (animal, situational-environmental, blood-injection-injury phobia), Obsessive-Compulsive Disorder; Traumatic Stress Disorder	Arabic, Chinese, Dutch, French, German, Greek, Italian, Korean, Malay, Persian, Portuguese, Spanish, Swedish	Yes
Spence Child Anxiety Scale (SCAS; Spence, 1998)	7-19	38-items (child-version has 7 additional filler items) Response scale: 0 (never) – 3 (always)	Separation Anxiety; Social Phobia; Generalized Anxiety; Obsessive-Compulsive Problems; Panic/Agoraphobia; Physical Injury Fears (Phobias)	At least 33 languages (per scaswebsite.com)	Yes

Table 3

Summary of Ratings for Youth Anxiety Self-Report Measures

Criterion	STAI-C (Trait)	Measure							Totals (by criterion)
		RCMAS	FSSC-R	SASC-R/ SAS-A	SPAI-C	MASC	SCARED	SCAS	
Norms	Good	Excellent	Adequate	Adequate	Good	Good	Good	Good	1 E, 5G, 2A
Internal Consistency	Good	Good	Excellent	Good	Excellent	Good	Good	Good	2E, 6G, 0A
Test-retest Reliability/ Stability	Good	Good	Good	Good	Excellent	Excellent	Good	Good	2E, 6G, 0A
Content Validity	Good	Good	Good	Good	Excellent	Excellent	Good	Excellent	3E, 5G, 0A
Construct Validity	Good	Good	Good	Good	Good	Good	Good	Good	0E, 8G, 0A
Discriminative Validity	Adequate	Adequate	Adequate	Adequate	Excellent	Excellent	Good	Good	2E, 2G, 4A
Validity Generalization	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent	8E, 0G, 0A
Treatment Sensitivity	Good	Excellent	Excellent	Good	Excellent	Excellent	Excellent	Excellent	6E, 2G, 0A
Totals (by measure)	1E, 5G, 1A	3E, 4G, 1A	3E, 3G, 2A	1E, 4G, 2A	6E, 2G, 0A	5E, 3G, 0A	2E, 6G, 0A	3E, 5G, 0A	

Note. The following criteria are omitted from the table because we were unable to assign ratings: interrater reliability, repeatability, prescriptive validity, and clinical utility. E = Excellent, G = Good, A = Adequate.