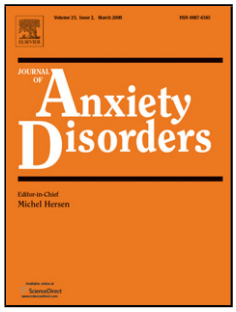


Title: Discriminant validity, diagnostic utility, and parent-child agreement on the Screen for Child Anxiety Related Emotional Disorders (SCARED) in treatment- and non-treatment-seeking youth

Authors: B.I. Rappaport, D. Pagliaccio, D.S. Pine, D.N. Klein, J.M. Jarcho

PII: S0887-6185(16)30393-0
DOI: http://dx.doi.org/10.1016/j.janxdis.2017.08.006
Reference: ANXDIS 1966

To appear in: *Journal of Anxiety Disorders*

Received date: 2-12-2016
Revised date: 13-8-2017
Accepted date: 27-8-2017

Please cite this article as: Rappaport, BI., Pagliaccio, D., Pine, DS., Klein, DN., & Jarcho, J.M., 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* http://dx.doi.org/10.1016/j.janxdis.2017.08.006

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

Rappaport, B. I.[1]
Pagliaccio, D.[1]
Pine, D. S.[1]
Klein, D. N.[2]
Jarcho, J. M.[2,*]

[1] Emotion and Development Branch, The National Institute of Mental Health, USA

[2]Department of Psychology, Stony Brook University, Stony Brook, USA

*Corresponding author

Present Address:
Stony Brook University
College of Arts and Sciences
Department of Psychology
Psychology B Building
Stony Brook, NY 11794-2500

TEL: 631.632.7800
FAX: 631.632.7876

johanna.jarcho@stonybrook.edu

HIGHLIGHTS

- Properties of the SCARED are examined in treatment and non-treatment-seeking youth.
- We confirm discriminant validity and diagnostic utility in treatment-seeking youth.
- Informant agreement was modest across both samples.
- Healthy children report higher scores than their parents.
- Anxious youth report equal or lower scores than their parents.

ABSTRACT (152 words)

The Screen for Child Anxiety and Related Emotional Disorder (SCARED) may be differentially sensitive to detecting specific or comorbid anxiety diagnoses in treatment-seeking and non-treatment-seeking youth. We assessed the SCARED's discriminant validity, diagnostic

utility, and informant agreement using parent- and self-report from healthy and treatment-seeking anxious youth (Study 1, *N*=585) or from non-treatment-seeking anxious youth (Study 2, *N*=331) diagnosed with generalized anxiety disorder (GAD), social anxiety disorder (SAD), or comorbid GAD+SAD. Among treatment-seeking youth, the SCARED showed good diagnostic utility and specificity, differentiating healthy, comorbid, and non-comorbid anxious youth. Child-parent agreement was modest: healthy child self-reports were higher than parent-reports whereas anxious child self-reports were similar or lower than parent-reports. Less consistent results emerged for diagnostic utility, specificity, and informant agreement among non-treatment-seeking youth. Given the number of non-treatment seeking anxious youth (*N*=33), generalizability of these findings may be limited. Together, results suggest informants may provide distinct information about children's anxiety symptoms.

Keywords: Anxiety; Children; Adolescents; Psychometrics; Screen for Child Anxiety Related Emotional Disorders (SCARED)

1. Introduction

The Screen for Child Anxiety and Related Emotional Disorders (SCARED) is a multi-informant questionnaire originally created as a screening tool for pediatric anxiety disorders (Birmaher et al., 1999; Birmaher et al., 1997). While studies have largely confirmed the SCARED's psychometric properties in clinically referred samples (Bodden, Bogels, & Muris, 2009; Dirks et al., 2014; Hariz et al., 2013; Monga et al., 2000; Muris, Dreessen, Bogels, Weckx, & van Melick, 2004; Muris & Steerneman, 2001; Stevanovic, 2012; Weitkamp, Romer, Rosenthal, Wiegand-Grefe, & Daniels, 2010), fewer have compared accuracy of diagnostic cutoffs obtained from the SCARED with diagnoses from standardized clinical interviews

(Bodden et al., 2009; Gardner, Lucas, Kolko, & Campo, 2007; Hariz et al., 2013; Monga et al., 2000; Muris et al., 2004; Muris & Steerneman, 2001; Weitkamp et al., 2010). Among such studies, modest sample size has largely prohibited tests of discriminant validity between specific subtypes of anxiety disorder diagnoses. Additionally, while comorbidity of multiple anxiety disorders is common (Kroenke, Spitzer, Williams, Monahan, & Lowe, 2007), there have been no tests of the SCARED's validity in patients with specific comorbid anxiety disorders. The present study addresses these limitations by testing the discriminant validity (i.e. how well the SCARED discriminates between non-anxious and anxious youth with specific and comorbid diagnoses), diagnostic utility (i.e. how well it correctly identifies the presence or absence of an anxiety disorder), and informant agreement (i.e. similarity in child- and parent-reports) of the SCARED in large samples of treatment- and non-treatment-seeking youth who are well-characterized via diagnostic interviews as having generalized anxiety, social anxiety, comorbid generalized and social anxiety, or as being free of psychopathology.

Although initially designed and implemented as a screening tool for use in clinical settings, the SCARED is often used in community settings. However, widely accepted clinical cutoffs were established with a small set of treatment-seeking patients (Birmaher et al., 1999), which may lead to underestimating the presence of diagnoses in community samples. This is largely due to treatment-seeking individuals having more severe symptoms than the broader population of people who meet criteria for a disorder (Stein & Kean, 2000). In fact, only four studies have tested the validity of clinical cutoffs for the SCARED in community samples that have undergone diagnostic interviews (Canals, Hernandez-Martinez, Cosi, & Domenech, 2012; DeSousa, Salum, Isolan, & Manfro, 2013; Muris et al., 2001; Su, Wang, Fan, Su, & Gao, 2008).

Thus, the first goal of this study is to assess the SCARED's validity in both treatment- and non-treatment-seeking populations.

The SCARED also utilizes a multi-informant approach aimed at establishing convergent validity for the child's anxiety symptoms. Yet, correlational analyses typically yield only modest levels of informant agreement (Birmaher et al., 1999; Birmaher et al., 1997; Cosi et al., 2010; Muris, Merckelbach, van Brakel, & Mayer, 1999; Su et al., 2008; Weitkamp et al., 2010). From a clinical perspective, it is critical to determine if informants provide complementary information that improves accuracy of clinical cutoffs, or if informant discrepancies lead to diagnostic discrepancies that vary based on the informant (Klein, 1991). Thus, a second goal of this study is to assess discriminant validity and diagnostic utility based on degree of informant agreement within and between diagnostic groups in treatment- and non-treatment-seeking populations.

To achieve these goals, the present study assesses large samples of children who completed the SCARED and underwent a comprehensive psychiatric interview. We aim to describe the SCARED's discriminant validity, diagnostic utility, and informant agreement in both treatment seeking and non-treatment-seeking populations of healthy and anxious youth diagnosed with individually occurring, or comorbid anxiety diagnoses. Given the high lifetime and 3-month prevalence rates of GAD and SAD among children (Costello et al., 2003; Merikangas et al., 2010), and the high costs associated with failing to treat them prior to adulthood (Greenberg et al., 1999), these anxiety disorders are the primary focus of the present study. While lifetime prevalence rates of other anxiety disorders with corresponding SCARED subscales (e.g., separation anxiety and panic disorder) are also high (Merikangas et al., 2010), 3-month prevalence rates vary. For instance, separation anxiety decreases dramatically after 9-10

years of age (Costello et al., 2003). For these reasons, GAD and SAD are of particular interest for this study and the broader program of research conducted at the NIMH, one of the study sites.

The current study addresses several limitations of previous examinations of the SCARED in community samples. Specifically, in this study, all participants were interviewed with the K-SADS-PL, as opposed to only those that scored in the top 25% on the SCARED (DeSousa et al., 2013). Moreover, the current study used the original 41-item version of the SCARED and the corresponding clinical cutoff scores (Birmaher et al., 1999), unlike other studies that used the 66-item SCARED-R which excludes the School Avoidance subscale and includes additional subscales (Muris, Merckelbach, Schmidt , & Mayer, 1998b), or those that used non-standard clinical cutoffs (Canals et al., 2012). Finally, the diagnostic accuracy of the GA and SA subscales as well as the Total scale was assessed. This is an important extension of a prior study, which only utilized the Total scale (Su et al., 2008). Therefore, by interviewing all participants with the K-SADS-PL, using consistent cutoff scores, and conducting ROC analyses of the Total scale, as well as two of the SCARED's subscales, this study provides a uniquely comprehensive assessment of the SCARED.

2.0 General Materials and Methods Across Studies

Child- and parent-report on the SCARED and semi-structured diagnostic interviews to determine current psychopathology (Kiddie-Schedule for Affective Disorders and Schizophrenia-Present and Lifetime; K-SADS-PL) (Kaufman et al., 1997) were obtained during large multi-visit studies at the National Institute of Mental Health (Study 1) and Stony Brook University (Study 2). Anxious youth met full criteria for current GAD and/or SAD. Healthy

children (HC) were free of lifetime psychopathology. All participants were free of major medical or developmental disorders and had at least one English-speaking parent or guardian willing to participate in the study. Institutional review boards approved all procedures for both studies.

2.0.1 Measures

The Screen for Child Anxiety Related Emotional Disorders is a 41-item child- and parent-report questionnaire with 5 subscales designed to screen for symptoms of specific anxiety disorders: Generalized Anxiety Disorder (e.g., "I am nervous"; "My child worries about being as good as other kids"), Social Anxiety Disorder (e.g., "I feel nervous with people I don't know well"; "My child is shy"), Separation Anxiety Disorder (e.g., "I am afraid to be alone in the house"; "My child worries about sleeping alone"), Panic Disorder/Significant Somatic Symptoms (e.g., "When I feel frightened, it is hard to breathe"; "My child gets really frightened for no reason at all"), as well as Significant School Avoidance (e.g., "I get stomachaches at school"; "My child is scared to go to school"). Each item on the questionnaire is rated on a 3-point scale (0=not true or hardly ever true, 1=somewhat true or sometimes true, 2=very true or often true). Items within each subscale are summed to form the 5 subscale scores, which are in turn summed together to form a Total score. Scores are tabulated separately for parent and child informants.

Given our primary focus on independent and comorbid manifestations of GAD and SAD in treatment- and non-treatment-seeking populations, results are restricted to data from the *Total*, *Generalized Anxiety* (*GA*) and *Social Anxiety* (*SA*) subscales and their corresponding diagnostic clinical cutoffs (Birmaher et al., 1999). The Total score is the sum of all 41 items (range = 0-82;

Study 1: α_{child} = 0.937, α_{parent} = 0.964; Study 2: α_{child} = 0.886, α_{parent} = 0.910) and has a clinical cutoff of 25. The GA subscale score is the sum of 9 items (range = 0-18; Study 1: α_{child} = 0.877, α_{parent} = 0.945; Study 2: α_{child} = 0.705, α_{parent} = 0.858) and has a clinical cutoff of 9. The SA subscale score is the sum of 7 items (range = 0-14; Study 1: α_{child} = 0.867, α_{parent} = 0.936; Study 2: α_{child} = 0.728, α_{parent} = 0.886) and has a clinical cutoff of 8. Only participants who completed ≥ 85% of items were included in analyses. Among included participants, missed responses were infrequent (3.6% for child-report, 6.6% for parent report), and were accounted for using multiple imputation.

Children's diagnostic status was ascertained based on both child and parent report during a semi-structured diagnostic interview to determine current psychopathology (Kiddie-Schedule for Affective Disorders and Schizophrenia-Present and Lifetime; K-SADS-PL) (Kaufman et al., 1997). Interviewers were blind to SCARED data.

2.1 Study 1 Participants (Table 1A).

A total of 585 youth completed the study. Healthy and treatment-seeking anxious children (i.e. those with a current primary diagnosis of GAD or SAD) were recruited from the Washington, D.C. metropolitan area to participate in research on the pathophysiology of pediatric anxiety disorders. Anxious children were given the option to obtain Cognitive Behavioral Therapy or pharmacotherapy following participation in the study. Children were 8-17 years of age with an IQ ≥ 70. Anxiety diagnoses were determined by licensed clinicians who administered the K-SADS-PL. Diagnoses of a current anxiety disorder were confirmed during brief consultation with a child psychiatrist (κ for GAD = 0.60; κ for SAD = 0.67). An independent samples t-test showed that HCs were older than anxious children ($t(582)=5.702$, p

< .001) and each anxiety sub-group, as confirmed by a one-way ANOVA, followed by pairwise comparisons ($F(3,583)=10.914$, post-hoc *p's* < .05). A χ^2 analysis determined there were no group differences in sex (*p's* > .05). Socioeconomic status (Hollingshead, 1957) was available for 383 participants. An independent samples t-test showed that HCs (Mean = 76.34 ± 8.73) did not differ from anxious children (Mean = 78.24 ± 8.77) on socioeconomic status ($t(381) = -2.066$, $p > .05$). A one-way ANOVA confirmed there were no differences among anxiety sub-groups ($F(3,382)=2.123$, $p > .05$). Race information was available for 480 participants. Of those participants, 72% were White or Caucasian, 14% were Black or African American, 9% were Multiple Races, 4% were Asian, 0.4% were American Indian or Alaskan Native, and 0.4% were Native Hawaiian and Other Pacific Islander. χ^2 test determined that there were no group differences in race ($p > .05$).

See Supplementary Table S1 for a description of additional comorbid diagnoses.

Studies that use diagnostic interviews to test the SCARED's clinical cutoffs often include unbalanced samples with a minority of anxious patients (e.g., Birmaher et al., 1997; DeSousa et al., 2013; Muris et al., 2001; Su et al., 2008; Weitkamp et al., 2010). Such an imbalance could lead to non-normal distributions and unequal variance between groups, potentially biasing the interpretation of results. To control for this, supplemental analyses that matched HC with anxious youth based on age and sex (heretofore referred to as the Matched sample) were performed to assess discriminant validity and informant agreement. T-tests and χ^2 analyses confirmed samples were well matched with no differences in age or sex (*p's* > .05; see Supplemental Materials for details). Over 90% of participants provided both parent- and child-report data for the SCARED. The remainder of the sample provided either parent- or child-report data.

The majority of participants completed a paper-based version of the SCARED that did not include a timeframe in the questionnaire instructions. While this is consistent with a revised version of the SCARED (Muris, 1998b), it does not correspond with the original version, which specifies responses should reflect experiences in the last 3 months (Birmaher et al., 1999). Beginning in 2013, participants (*N*=93) were given the opportunity to complete a computerized version of the SCARED that instructed participants to report on symptoms present in the past 3 months. MANOVAs confirmed that date of completion did not differentially influence parent- or child-report scores on Total, SA, and GA scales (*F*'s < 2.33; *p*'s > .07). Follow-up independent samples t-tests for each participant group confirmed that parent- and child-report scores did not differ based on date of completion for Total, SA, or GA scales (*t*'s < 1.95; *p*'s > .07). Because instruction timeframe did not appear to influence magnitude of reported symptoms, subsequent analyses pool participants.

2.2 Study 2 Participants (Table 1B).

A total of 331 youth completed the study. Children enrolled in an ongoing longitudinal study on the pathophysiology of risk for pediatric psychopathology were recruited from within 20 miles of Stony Brook University. Children enrolled in the longitudinal study at 3 to 4 years of age and provided data described in the current report at 9 years of age. Anxiety diagnoses were determined by advanced doctoral students or master's-level clinical psychologists, who administered the K-SADS-PL ($\kappa = 0.55$). Administration was supervised and approved by an experienced child psychiatrist and licensed clinical psychologist. No differences in age or sex emerged between any of the groups (i.e. HC, GAD, SAD, GAD+SAD) (*p's* > .05). T-tests and

χ^2 analyses confirmed no differences in age or sex (*p's* > .05) between each group of anxious youth and their corresponding Matched Sample of HCs (see Supplemental Materials for details). Socioeconomic status was available for 305 participants (Hollingshead, 1975). An independent samples t-test showed that HCs (Mean = 45.07 ± 11.21) did not differ from anxious children (Mean = 41.67 ± 9.86) on socioeconomic status ($t(303) = 1.647$, $p > .05$). A one-way ANOVA confirmed there were no differences among anxiety sub-groups ($F(3,304)=1.307$, $p > .05$). Race information was available for 331 participants. Of those participants, 91% were White or Caucasian, 6% were Black or African American, 3% were Asian, and 0.3% were American Indian or Alaskan Native. A χ^2 test determined that there were no group differences in race ($p > .05$). Over 98% of participants provided both parent- and child-report data for the SCARED. The remainder of the sample provided either parent- or child-report data. Parents reported on their child's symptoms within the past 3 months. Children were not given a timeframe on which to base their reports, but completed the SCARED as part of a battery of questionnaires focused on current experiences and symptoms.

2.3 Analytic Plan

The same analytic plan was utilized for both studies.

2.3.1 Discriminant validity.

One-way ANOVAs tested whether child- and parent-report scores on Total, GA, and SA subscales differed across diagnostic groups (HC, GAD, SAD, GAD+SAD). Discriminant validity will be confirmed if: 1) each anxious group (GAD, SAD, GAD+SAD) has higher Total, GA, and SA scores than the HC group; 2) the GAD and SAD groups differ such that GAD have higher GA scores than SAD, and SAD have higher SA scores than GAD; and 3) the GAD+SAD group has higher GA scores than the SAD group, and SA scores than the GAD group. Significant

group differences were interrogated with Bonferroni pairwise comparisons. Because six comparisons were conducted for each scale, the p-value was adjusted for multiple comparisons for each scale (critical *p*-value = 0.05/6 = .008). When Shapiro-Wilk tests identified non-normally distributed data (see Table 2), results were confirmed with Kruskal-Wallis tests. Brown-Forsythe tests revealed unequal variance for Total, GA, and SA scores across groups in both studies. However, Dunnett's T3 post-hoc tests largely confirmed findings (see supplementary materials for further details).

2.3.2 Diagnostic Utility.

One-sample t-tests assessed whether child- and parent-report scores on Total, GA, and SA subscales differed from clinical cutoffs for each diagnostic group (HC, GAD, SAD, GAD+SAD). Diagnostic utility will be confirmed if 1) HC scores are below the clinical cutoff on each scale; 2) each patient group (GAD, SAD, GAD+SAD) exceeds the clinical cutoff for the Total score (≥ 25); 2) the GAD and SAD groups exceed the clinical cutoff for GA (≥ 9) and SA (≥ 8) subscales, respectively; and 3) the GAD+SAD group exceeds the clinical cutoff for both the GA and SA subscales. Diagnostic utility was further assessed with receiver operating characteristic (ROC) analyses to quantify the sensitivity (true positive rate) and specificity (true negative rate) of the established clinical cutoff scores in each sample. Sensitivity indicates the probability that a child's SCARED score will meet or exceed the clinical cutoff when the child is diagnosed on the K-SADS-PL, and the specificity indicates the probability that a child's score will fall below the clinical cutoff when the child is not diagnosed on the K-SADS-PL. Positive and negative predictive values and overall correct classification were also quantified. Positive Predictive Values (PPV) and Negative Predictive Values (NPV) reflect rates of true positive and negative scores. Specifically, the PPV and NPV are the number of true observations (according

to the K-SADS-PL) divided by the total number of positive or negative observations, according to SCARED cutoff scores. In this case, the PPV was calculated as the percent of participants that met or exceeded the SCARED's clinical cutoff, and had a diagnosis on the K-SADS-PL. Similarly, the NPV was calculated as the percent of participants that fell below the SCARED's clinical cutoff and did not have a diagnosis on the K-SADS-PL. The Overall Correct Classification (OCC) is the sum of the number of true positives and true negatives, divided by the overall number of participants. This gives the percent of participants correctly classified by the SCARED as either having a true diagnosis or not having a diagnosis.

2.3.3 Informant Agreement.

Bivariate Pearson correlations assessed the association between child- and parent report scores on the Total, GA, and SA subscales across each study's full sample, and within each diagnostic group (HC, GAD, SAD, GAD+SAD). Non-parametric Spearman Rho analyses largely confirmed findings (see supplementary materials for further details). Group differences in the magnitude and direction (i.e. whether the child-report was higher than parent-report or vice versa) of child-parent discrepancies on Total, GA, and SA subscales were assessed with a Group (HC, GAD, SAD, GAD+SAD) X Informant (child, parent) repeated measures ANOVA. Significant interactions were interrogated with Bonferroni-corrected pairwise comparisons. Because six comparisons were conducted for each scale, the p-value was adjusted for multiple comparisons for each scale (critical p-value = 0.05/6 = .008). Kruskal-Wallis tests on parent-child difference scores, and follow-up analyses with Wilcoxon Signed Rank tests, confirmed unequal variance did not bias findings.

3. Results

3.1. Study 1: Treatment-seeking Sample

Results derived from comparisons between anxious youth and the Matched sample were largely consistent with those from the full sample, thus details are reported in supplementary materials.

3.1.1. Discriminant Validity (Table 2, Figure 1)

Discriminant validity for the SCARED was largely supported and consistent across child- and parent-report. One-way ANOVAs showed that the groups (HC, GAD, SAD, GAD+SAD) differed for Total child ($F(3, 554) = 149.5$, $p < .001$, $\eta^2 = .45$), and parent ($F(3, 530) = 374.9$, $p < .001$, $\eta^2 = .68$) scores, GA child ($F(3, 554) = 134.0$, $p < .001$, $\eta^2 = .42$), and parent ($F(3, 530) = 438.2$, $p < .001$, $\eta^2 = .71$) scores, and SA child ($F(3, 554) = 74.3$, $p < .001$, $\eta^2 = .29$), and parent ($F(3, 530) = 250.9$, $p < .001$, $\eta^2 = .59$) scores. Results from post-hoc pair-wise comparisons between group scores on each subscale are depicted in Figure 1 (see Table S2 for mean difference scores and corresponding p-values for Bonferroni pairwise comparisons). HCs scored consistently below all three anxious groups; GAD+SAD scored above GAD (parent-report only) and SAD on the Total scale; GAD+SAD and GAD scored above SAD on the GA subscale; and GAD+SAD and SAD scored above GAD on the SA subscale (see supplementary materials for further details).

3.1.2. Diagnostic Utility

3.1.2.1. Clinical Cutoffs (Figure 1). Figure 1 depicts clinical cutoffs (grey dotted line) for each scale. For parent- and child-report, HCs fell well below the cutoff for Total, GA, and SA subscales (*p's* < .001; see supplementary Table S3 for t-scores).

Among GAD, scores for both child- and parent-report were above the clinical cutoff for the Total scale ($p < .001$ and $p < .05$, respectively). For the GA subscale, parent-report was above the clinical cutoff ($p < .01$), whereas child-report scores did not differ from the clinical cutoff ($p > .05$). For the SA subscale, both informant scores were below the clinical cutoff (*p's* < .001).

Among SAD, parent-report was above the clinical cutoff for the Total scale ($p < .05$), whereas child-report scores did not differ from the clinical cutoff ($p > .05$). For the GA subscale, child-report was below the clinical cutoff ($p < .001$), whereas parent-report scores did not differ from the clinical cutoff ($p > .05$). For the SA subscale, only parent-report was above the clinical cutoff ($p < .001$), whereas child-report scores did not differ from the clinical cutoff ($p > .05$).

Among GAD+SAD, scores for both child- and parent-report were above the clinical cutoff for the Total (*p's* < .001) and SA scales ($p < .05$ and $p < .001$, respectively). For the GA subscale, parent-report was above the clinical cutoff ($p < .001$), whereas child-report scores did not differ from the clinical cutoff ($p > .05$).

3.1.2.2. Receiver Operating Characteristic Analysis (Diagnostic Sensitivity and Specificity) (Table 3). Child- and parent-report scores successfully discriminated the presence/absence of any anxiety disorder for the Total scale (AUC = 0.903, 0.981, respectively, *p's* < .001), presence/absence of GAD with the GA subscale (AUC = 0.860, 0.927, *p's* < .001),

and presence/absence of SAD with the SA subscale (AUC = 0.833, 0.941, *p's* < .001). Table 3 describes the sensitivity and specificity based on the SCARED's clinical cutoff scores for Total score and the GA and SA subscales. While all reports show high specificity (>84%) across informant and scale/subscale, they come at the cost of lower sensitivity (<80%). In line with these findings, PPV ranges from 53.19% to 98.56%, while NPVs and OCCs are consistently high, ranging from 80.30% to 94.40% and 80.72% to 89.64%, respectively.

3.1.3. Informant Agreement.

3.1.3.1. Correlations (Table 4). When collapsing across the full sample, informant reports were highly correlated for each scale (r's > .50). More modest correlations emerged among HCs and when collapsing across anxiety sub-groups. There was less consistency within each anxiety disorder subtype. There was no relation between informant scores for the Total scale, or among the SAD group for any scale. However, significant agreement emerged for the GAD group on the SA subscale, and the GAD+SAD group on the GA subscale. Non-parametric Spearman Rho analyses largely confirmed findings (see supplementary materials Table S5).

3.1.3.2. Group Differences in Informant Report (Table 2, Figure 1). Repeated measures ANOVAs revealed an Informant (child, parent) x Group (HC, GAD, SAD, GAD+SAD) interaction for Total (F (3, 497)= 15.79, p < .001, η^2 = .087), GA (F (3, 497)= 32.017, p < .001, η^2 = .116), and SA scores (F (3, 497)= 19.50, p < .001, η^2 = .093). Figure 1 depicts significant child-parent discrepancies in post-hoc analyses within each group (see supplementary Table S6 for within-group mean difference scores and corresponding p-values for t-tests). Pairwise comparisons demonstrate HC child-report was higher than parent-report on all three scales (*p's* <

.001). GAD+SAD, GAD, and SAD child-report was lower for GA scores than parent-report (*p*'s < .01,). SAD child-report was lower for the SA scores than parent-report ($p < .01$)

3.2. Study 2: Non-Treatment Seeking Sample

Results derived from comparisons between anxious youth and the Matched sample were largely consistent with those from the full sample, thus details are reported in supplementary materials.

3.2.1. Discriminant Validity (Table 2, Figure 2)

Discriminant validity for the SCARED was largely supported for parent-, but not child-report data. One-way ANOVAs showed groups (HC, GAD, SAD, GAD+SAD) differed for Total child ($F(3, 330) = 3.70$, $p < .05$, $\eta^2 = .033$) and parent ($F(3, 330) = 72.98$, $p < .001$, $\eta^2 = .402$) scores, GA parent ($F(3, 330) = 444850$, $p < .001$, $\eta^2 = .309$) scores, and SA child ($F(3, 330) = 4.95$, $p < .01$, $\eta^2 = .043$) and parent ($F(3, 330) = 37.75$, $p < .001$, $\eta^2 = .258$) scores, but not GA child scores ($F(3, 330) = 2.35$, $p > .05$, $\eta^2 = .021$). Results from post-hoc pair-wise comparisons between group scores on each subscale are depicted in Figure 2 (Figure 1; see Table S2 for mean difference scores and corresponding p-values for Bonferroni pairwise comparisons). For parent-report, HCs scored consistently below all three anxious groups on the Total scale. GAD+SAD scored above GAD and SAD on the Total scale; GAD+SAD and GAD scored above SAD on the GA subscale; and GAD+SAD and SAD scored above GAD on the SA subscale. For child-report, HCs from the full or matched sample scored below SAD on the SA subscale (see supplementary materials for further details).

3.2.2. Diagnostic Utility

3.2.2.1. Clinical Cutoffs (Figure 2). Figure 2 depicts clinical cutoffs (dotted line) for each scale. For parent- and child-report, HCs fell well below the cutoff for Total, GA, and SA subscales (*p's* < .001; see supplementary Table S3 for t-scores).

Among GAD, scores for both informants failed to differ from the clinical cutoff for the Total scale (*p*'s > .05). For the GA subscale, parent-report failed to differ from the clinical cutoff *p* > .05), whereas child-report scores were below the clinical cutoff (*p* < .001). For the SA subscale, both informant scores were below the clinical cutoff (*p's* < .01).

Among SAD, parent-report scores were below the clinical cutoff on the Total scale (*p* < .01), whereas child-report failed to differ from the clinical cutoff (*p* > .05). For the GA subscale, both parent- and child-report scores were below the clinical cutoff (*p's* < .001). For the SA subscale, both informant scores failed to differ from the clinical cutoff (*p's* > .05).

Among GAD+SAD, scores for both informants failed to differ from the clinical cutoff for Total, SA, or GA scales (*p's* > .05).

3.2.2.2. Receiver Operating Characteristic Analysis (Diagnostic Sensitivity and Specificity) for child, parent, and average child-parent report scores (Table 3). Child- and parent-report scores successfully discriminated the presence/absence of any anxiety disorder for the Total scale (AUC = 0.657, 0.924, respectively, *p's* < .01), the presence/absence of SAD with the SA subscale (AUC = 0.758, 0.944, *p's* < .001). However, only the parent-report (AUC = 0.931, *p's* < .001), not the child-report score (AUC = .582, *p* > .05), successfully discriminated the presence/absence of GAD on the GA subscale. Table 3 describes the sensitivity and specificity based on the SCARED's clinical cutoff scores for Total score and the GA and SA subscales. While all reports shows high specificity (>76%) across informant and scale/subscale,

they come at the cost of lower sensitivity (<71%). In line with these findings, PPV ranges from 3.85% to 92.31 %, while NPVs and OCCs are consistently high, ranging from 91.53 % to 98.34% and 72.21% to 95.44 %, respectively.

3.2.3. Informant Agreement in the SBU subsample.

3.2.3.1. Correlations (Table 4). When collapsing across the full sample, informant reports were significant, but modest for each scale. Weaker relations emerged for each scale among HCs. Correlations were substantially higher, but largely non-significant given the relatively small sample size, when collapsing across anxiety sub-groups and within each anxiety disorder subtype. Non-parametric Spearman Rho analyses largely confirmed findings (see supplementary materials Table S5).

3.2.3.2 Group Differences in Informant Report (Table 2, Figure 2). Repeated measures ANOVAs revealed an Informant (child, parent) x Group (HC, GAD, SAD, GAD+SAD) interaction for Total ($F(3, 325) = 9.708$, $p < .001$, $\eta^2 = 0.082$), GA ($F(3, 325) = 17.326$, $p < .001$, $\eta^2 = 0.138$), and SA scores (F (3, 325)= 6.340, $p < .001$, $\eta^2 = .055$). Figure 2 depicts significant child-parent discrepancies in post-hoc analyses within each group (see supplementary Table S6 for within-group mean difference scores). Overall, HC child-report was higher than parent-report on all three scales. The same pattern emerged for SAD on the Total scale. GAD and GAD+SAD child-reports were lower than parent-report on the GA subscale. No other differences were noted for the SA subscale.

4. Discussion

The present study aimed to test discriminant validity, diagnostic utility, and informant agreement of the SCARED in a large sample of treatment- and non-treatment-seeking youth who were well characterized with diagnostic interviews as exhibiting GAD, SAD, comorbid GAD+SAD, or as free of psychopathology (HC). Results support discriminant validity and diagnostic utility among treatment-seeking youth. However, data suggest careful consideration is needed when utilizing the SCARED in a non-treatment-seeking community sample and when choosing to utilize child- or parent-report scores.

4.1. Discriminant validity.

The majority of analyses suggest adequate discriminant validity between HC's and patients for Total, GA, and SA subscales of the SCARED. For the treatment-seeking sample, both child and parent reports of anxiety were significantly lower for HCs than anxious participants. This is consistent with prior studies that utilize diagnostic interviews to test the SCARED in treatment-seeking anxious patients and youth without psychopathology (Bodden et al., 2009) and those with non-anxiety-related disorders (Birmaher et al., 1999; Hariz et al., 2013; Monga et al., 2000; Muris et al., 2004). However, results for child-report in the non-treatment-seeking sample were less compelling. Specifically, child-reports of anxiety did not discriminate between HCs and SAD or GAD on the Total scale and GA subscale, nor between HCs and comorbid GAD+SAD on the SA subscale.

Few studies have assessed discriminant validity in anxious patients diagnosed with the disorders measured by the SCARED's specific subscales, or among patients with comorbid anxiety diagnoses. Across all analyses and scales, and largely irrespective of informant, youth

comorbid for GAD+SAD had more anxiety symptoms than other patients—as indexed by higher Total, GA, and SA scores (Figs 1-2, purple lines). Thus, youth with comorbid diagnoses had more severe levels of anxiety than individuals with a single diagnosis. This replicates prior work in patients tested with clinical interviews where severity was measured with questionnaires (Hofmeijer-Sevink et al., 2012). Support for discriminant validity between patient subtypes was also obtained across the majority of analyses among the treatment-seeking sample. GA scores were higher for GAD and GAD+SAD than SAD alone, while SA scores were higher for SAD and GAD+SAD than GAD alone. A similar pattern emerged for parent-report among the community sample. However, little evidence for discriminant validity was obtained for child-report in the community sample. Taken together, this suggests that in a treatment-seeking sample, the SCARED is able to discriminate between children with anxiety disorders and those free of any psychopathology, and between sub-types of anxious patients based on the subscale that is utilized.

In a sample drawn from the community, child-report on the SCARED may be more heterogeneous and reflect a continuum of anxiety symptoms rather than the categorical expression of clinical characteristics. Yet, data from other studies demonstrate that in community settings, child-report on the SCARED has good convergent validity with other measures of anxiety (e.g., Essau, Muris, & Ederer, 2002; Muris et al., 1998a; Muris, Schmidt, & Merckelbach, 2000; Su et al., 2008). Moreover, other facets of the SCARED's psychometric properties including test-retest reliability (e.g., Boyd, Ginsburg, Lambert, Cooley, & Campbell, 2003; Muris et al., 1999; Su et al., 2008) and internal consistency (e.g., Boyd et al., 2003; Essau, Anastassiou-Hadjicharalambous, & Munoz, 2013; Isolan, Salum, Osowski, Amaro, & Manfro, 2011; Su et al., 2008; Weitkamp et al., 2010) are well established with child-report in community

samples. These properties have been confirmed across numerous countries and various languages (e.g., Crocetti, Hale, Fermani, Raaijmakers, & Meeus, 2009; Essau et al., 2013; Isolan et al., 2011; Su et al., 2008; Vigil-Colet et al., 2009). Thus, child-report data in community settings should not be discounted. Instead, researchers and clinicians are simply urged to consider whether child-report data discriminate between clinically meaningful subtypes of anxiety diagnoses in community samples or if such measures may better reflect a continuum of symptoms.

4.2. Diagnostic utility.

The present data found mixed support for the diagnostic utility of the SCARED. In both studies, HCs consistently had scores below Total, GA, and SA clinical cutoffs, regardless of informant. Results were less consistent for anxious youth: among those in the treatment-seeking sample, parent-report scores exceeded clinical cutoffs for Total, GA and SA subscales among GAD and GAD+SAD samples. There is less support for the SCARED's diagnostic utility in the community sample, where no patient group fell above clinical cutoffs for any of the three scales; however, this could be a result of the relatively small number of patients in each diagnostic group.

ROC analyses corroborate this pattern of findings. In the treatment-seeking sample, across informants, scores successfully discriminated the presence/absence of any anxiety disorder, GAD, and SAD on the Total, GA, and SA subscales, respectively. While specificity was uniformly high across informants and scales (> 85%), sensitivity ranged from 57% detection rate for child-report on the GA subscale to 75% detection rate for parent-report on the GA and SA subscales. This pattern of results is consistent with other ROC analyses of the SCARED which identified high specificity (0.89-0.92) at the cost of low sensitivity (0.44) (e.g., Gardner et

al., 2007). Rates dropped substantially in the community sample. While specificity remained relatively high (> 76%), sensitivity was under 50% for child- and parent-report for Total and GA subscales, reaching a maximum of nearly 67% detection rate for parent-report on the SA subscale.

One of the key motivations for developing the SCARED was to address the issue of under-diagnosis and under-treatment of anxious youth who may receive less clinical attention due to an absence of behavioral problems (Birmaher et al., 1999). Thus, a tool used to screen for anxiety disorders with specific clinical cutoffs should have appropriate levels of sensitivity, to ensure that youth with the disorder do not go undiagnosed or untreated, and specificity, to ensure diagnoses and treatment are not provided unnecessarily to non-anxious individuals. Given the high long-term cost of failing to treat an anxious child before they develop long term, difficult to treat symptoms (Wittchen et al., 2011), a screening tool for pediatric anxiety may choose to sacrifice *specificity* to increase *sensitivity*. Thus, more inclusive clinical cutoffs may be beneficial. Indeed, researchers have suggested clinical cutoffs may benefit from downward adjustments to more appropriately characterize both treatment-seeking and community samples (Canals et al., 2012; DeSousa et al., 2013).

4.3. Informant agreement.

For the HC group, child-reported anxiety symptoms were higher than parent-report (Figures 1-2, green lines). This suggests that children free of clinically significant levels of anxiety report more symptoms than their parents. Among anxious participants a different pattern of results emerged such that parental reports of anxiety were largely equal to or higher than children's reports.

Within the treatment-seeking sample, correlations between parent and child report were relatively high when collapsed across all groups (i.e., HC and anxious youth). However, as with prior reports on child-parent agreement (Birmaher et al., 1999; Birmaher et al., 1997; Cosi et al., 2010; Muris et al., 1999), when considered separately, correlations among HCs and anxious youth were relatively modest. Few prior studies have had a sufficient sample size to assess child-parent agreement in youth with specific anxiety diagnoses. We were able to assess such agreement separately for GAD, SAD, and GAD+SAD comorbid samples, identifying distinct patterns of results. Specifically, while child-parent correlations for GAD and GAD+SAD patients were modest, strikingly little agreement was observed among patients with SAD.

Among the community sample, correlations between parent and child report were modest when collapsed across all groups (i.e., HC and anxious youth). This may have been driven by the low level of agreement among the HC's, as modest correlations were observed across all patients, and within each diagnostic group. Indeed, previous studies of non-diagnosed community samples have also found that children report higher scores than their parents (Cosi et al., 2010; Muris et al., 1999; Su et al., 2008; Wren et al., 2007; Wren, Bridge, & Birmaher, 2004), further suggesting a difference in how children and parents generally report on the child's symptoms of anxiety. While the data does not necessarily support the proposition that children are responding more *accurately* than their parents (as indicated by the ROC analyses), it is possible that children are, in certain cases, more knowledgeable about their own feelings and symptoms of anxiety as opposed to the symptoms externally observable to their parents. Therefore, discrepancies in child- and parent-report could be related to communication between the child and his/her parent(s).

Given informant discrepancy, generating a composite score that reflects an average of child- and parent-report could help minimize potential disagreement across informants. However, the context in which the tool is being used must be considered carefully. If used as a clinical screening tool, parental report on the SCARED may be of greater utility than child report. This reflects the fact that parental reports match more closely with clinical cutoffs and tend to have higher sensitivity and specificity for accurately discriminating healthy and anxious patients. The bias toward diagnostic utility for parent-report may, to some degree, be indicative of the fact that diagnostic clinical interviews tend to weigh parental reports more heavily than child self-report when discrepant information is provided – particularly among young children (Grills & Ollendick, 2003). It is also possible that parents may be more inclined to view anxiety as a categorical trait that is present or absent in their child, which could be reflected by scores that are either very low (0-2) or very high (>40); they may also have a bias, such that they exhibit a "halo" effect where they do not distinguish between items. Children, on the other hand, may have unbiased reports of their experience with each symptom, without regard or perhaps knowledge of the dichotomous nature of DSM-based psychiatric diagnoses.

Recent research also suggests that informant disagreement may, to some extent, reflect a discrepancy in cognitive abilities between parents and their children, as well as possible concerns for social desirability (De Los Reyes, et al., 2015). For example, an 8-year-old child may have difficulty remembering whether they experienced certain symptoms (e.g., getting headaches at school), or may not report symptoms that he/she considers embarrassing (e.g., difficulty sleeping away from home). Their parent may be better able to remember instances when and where the child reported symptoms, and could be less inhibited about reporting embarrassing symptoms. Moreover, as suggested by De Los Reyes et al. (2015), when discrepancies occur, clinicians tend

to defer to the informant that sought treatment in the first place. That is, parents may be more valid informants because they often initiate their child's treatment, which in turn causes clinicians to rely more heavily on their diagnostic interviews. While this may shed light on discrepancies observed in Study 1's treatment-seeking sample, it is an unlikely explanation for discrepancies observed in Study 2's non-treatment-seeking community sample. Therefore, while parent-child agreement on the SCARED is moderately low, such findings are quite common and could be reflect a number of factors other than measurement error.

The SCARED is often used as a dimensional measure of anxiety severity in neuroimaging and behavioral research of in- and outpatient treatment-seeking (e.g., Guyer et al., 2008; Ladouceur, Dahl, Birmaher, Axelson, & Ryan, 2006) and community samples (e.g., Kessel, Kujawa, Hajcak Proudfit, & Klein, 2015). It has also been used to track improvement of anxiety symptoms during the course of treatment (Birmaher et al., 2003). Given the limited sample size of anxious youth in Study 2, we cannot conclude that the SCARED is best suited for one population over another. Moreover, the community sample was, on average, younger than the treatment-seeking sample. Thus, differences across samples could partially reflect differences in age. Instead, we believe that these results suggest that child- and parent-report supply different information about children's anxiety symptoms. In practice, clinicians could use differences in child and parent reporting to their benefit by referring children whose parents report higher scores for further screening for anxiety disorders, or to track improvement during treatment. However, in contexts where there is greater interest in the child's subjective experience of anxiety, regardless of whether this experience reflects the presence or absence of clinically significant expression of symptoms, self-report measures may be of greater utility. This is line with recent conceptualizations of informant discrepancy, including the Operations

Triad Model, which suggests that multiple informants contribute unique perspectives and therefore distinct and meaningful information (De Los Reyes, Thomas, Goodman, & Kundey, 2013).

4.4. Limitations and Conclusions

Despite its strengths, the results from the present work must be considered in light of its weaknesses. First, differences in sample characteristics prevented the direct comparison of SCARED data for treatment- and non-treatment seeking samples. This critical next step is needed to establish or confirm that clinical cutoff scores have acceptable specificity and sensitivity in both community and treatment-seeking samples. One set of cutoffs may not be sufficient for such disparate samples. Instead, separate cutoffs may be necessary to appropriately accommodate both groups.

Second, issues related to age, sex, and ethnicity are beyond the scope of the current study. Previous research suggests that such factors may differentially impact discriminant validity, diagnostic utility, and informant agreement on the SCARED (e.g., Boyd et al., 2003; Crocetti et al., 2009; Hale et al., 2011); however little is known about the extent to which these relations vary across community compared to treatment-seeking samples. In the present study, the majority of participants at both sites were Caucasian. Targeted recruitment is needed in future studies to obtain a more diverse sample in both community and treatment-seeking populations.

Third, although the treatment-seeking sample in Study 1 is the largest to date to test the SCARED in conjunction with diagnostic interviews, the community sample of anxiety-disordered youth in Study 2 was relatively small. Given the limited number of anxious youth in the community sample (N=33), our ability to generalize the findings from this group is somewhat limited. Studies with small samples have a lower likelihood of identifying true effects given their

relatively low power. Large community-based studies that administer the SCARED in conjunction with a clinical interview, such as the K-SADS-PL, are needed to yield a larger number of anxious youth. Studies that directly compare the scores on the SCARED in well-matched treatment-seeking and community populations will likely need to oversample from the community to achieve this goal.

Fourth, because only participants with either a primary anxiety disorder or no psychiatric disorders were assessed, we were not able to determine discriminant validity between youth with anxiety disorders and youth with other psychiatric diagnoses (e.g., depressive disorders, disruptive disorders). However, previous studies, have demonstrated the SCARED's discriminant validity between such populations (Birmaher et al., 1999; Birmaher et al., 1997; Dirks et al., 2014; Monga et al., 2000; Muris et al., 2004; Su et al., 2008; Weitkamp et al., 2010).

Fifth, we cannot entirely rule out the possibility that inconsistencies in the timeframe described in the SCARED's instructions may have contributed to discrepancies in child and parent reports. However, analyses in Study 1 suggest that participants who were or were not instructed to complete the SCARED based on the past 3 months, reported similar levels of symptoms. This consistency may reflect the relatively high 3-month prevalence rates of GAD and SAD (Costello et al., 2003), and/or the fact that research assistants told participants they would fill out a battery of questionnaires about how they had been feeling recently, unless instructions explicitly noted otherwise. A similar method for administering questionnaires was employed in Study 2. Here, parents, but not children, received instructions to complete the SCARED based on their experience in the past 3 months. Because of how the battery of questionnaires was administered, and also because participants have been part of an ongoing longitudinal study since age 3, responses likely reflect current, rather than lifetime symptoms.

While there is no way to confirm this, data from Study 1 suggest that timeframe instructions may not, in this context, have a large influence on how participants respond. Given the various contexts in which the SCARED is administered, it is important that researchers carefully consider, and report, the timeframe they instruct participants to use when completing the SCARED.

Finally, the present study tested parent- and child-report scores to determine if one or the other better informs discriminant validity, diagnostic utility, and the extent to which there was overall informant agreement. An alternative approach would be to test relative patterns of child-parent agreement. For instance, if two children have the same score, but one is below the parent's rating, and the other is above the parent's rating, how do these relative differences impact discriminant validity and diagnostic utility? This strategy necessitates a larger sample size than available in the current work.

This study and others have demonstrated various aspects of the psychometric properties of the SCARED. We believe the SCARED is a valuable tool, but that like many tools, there may more or less productive ways to utilize it across a wide variety of studies. Thus, clinicians and researchers should carefully consider the context in which they plan to utilize the SCARED. Such considerations must include the sample they aim to study (community vs. clinical), the informant reports they are able to obtain (parent vs. child or both), and whether this combination of factors will enable them to best use the SCARED as a screening assessment tool, or as a measure of symptom severity.

Acknowledgments

This work was supported by the National Alliance for Research in Schizophrenia and Affective

Disorders (NARSAD) Young Investigator Award: Ellen Schapiro & Gerald Axelbaum Investigator (to J. M. Jarcho), by the Richard J. Wyatt Memorial Fellowship Award for Translational Research (to J. M. Jarcho), by National Institute of Mental Health Intramural Program (ZIA MH002781; to D. S. Pine) and NIMH Grant R01MH069942 (to D. N. Klein) and was conducted under NIH Clinical Study Protocol 01-M-0192 (clinicaltrials.gov, NCT 00018057. Funding agencies had no involvement in the study design, collection, analysis and interpretation of the data, the writing of this report, or the decision to submit this article for publication. The authors would like to acknowledge David Luckenbaugh for his consultation on the data analyses and Adina Heckelman for her assistance in data verification.

References

American Psychiatric Association, & American Psychiatric Association. (1994). Diagnostic and statistical manual (DSM-IV). *Washington, DC: American Psychiatric Association*.

Birmaher, B., Axelson, D. A., Monk, K., Kalas, C., Clark, D. B., Ehmann, M., ... & Brent, D. A. (2003). Fluoxetine for the treatment of childhood anxiety disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(4), 415-423. doi: 10.1097/01.CHI.0000037049.04952.9F

Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): a replication study. *J Am Acad Child Adolesc Psychiatry, 38*(10), 1230-1236. doi: 10.1097/00004583-199910000-00011

Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., & Neer, S. M. (1997). The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. *J Am Acad Child Adolesc Psychiatry, 36*(4), 545-553. doi: 10.1097/00004583-199704000-00018

Bodden, D. H. M., Bogels, S. M., & Muris, P. (2009). The diagnostic utility of the Screen for Child Anxiety Related Emotional Disorders-71 (SCARED-71). *Behaviour Research and Therapy, 47*(5), 418-425. doi: 10.1016/j.brat.2009.01.015

Boyd, R. C., Ginsburg, G. S., Lambert, S. F., Cooley, M. R., & Campbell, K. D. M. (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 and Adolescent Psychiatry, 42*(10), 1188-1196. doi: 10.1097/01.chi.0000082033.88099.8d

Canals, J., Hernandez-Martinez, C., Cosi, S., & Domenech, E. (2012). Examination of a cutoff score for the Screen for Child Anxiety Related Emotional Disorders (SCARED) in a non-clinical Spanish population. *Journal of Anxiety Disorders, 26*(8), 785-791. doi: 10.1016/j.janxdis.2012.07.008

Cosi, S., Canals, J., Hernandez-Martinez, C., & Vigil-Colet, A. (2010). Parent-child agreement in SCARED and its relationship to anxiety symptoms. *Journal of Anxiety Disorders, 24*(1), 129-133. doi: 10.1016/j.janxdis.2009.09.008

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of general psychiatry*, *60*(8), 837-844. doi:10.1001/archpsyc.60.8.837

Crocetti, E., Hale, W. W., Fermani, A., Raaijmakers, Q., & Meeus, W. (2009). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED) in the general Italian adolescent population: A validation and a comparison between Italy and The Netherlands. *Journal of Anxiety Disorders, 23*(6), 824-829. doi: 10.1016/j.janxdis.2009.04.003

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*(4), 858-900. doi: 10.1037/a0038498

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology, 9*, 123-149. doi: 10.1146/annurev-clinpsy-050212-185617

DeSousa, D. A., Salum, G. A., Isolan, L. R., & Manfro, G. G. (2013). Sensitivity and Specificity of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A Community-Based Study. *Child Psychiatry & Human Development, 44*(3), 391-399. doi: 10.1007/s10578-012-0333-y

Dirks, M. A., Weersing, V. R., Warnick, E., Gonzalez, A., Alton, M., Dauser, C., . . . Woolston, J. (2014). Parent and youth report of youth anxiety: evidence for measurement invariance. *Journal of Child Psychology and Psychiatry, 55*(3), 284-291.

Essau, C. A., Anastassiou-Hadjicharalambous, X., & Munoz, L. C. (2013). Psychometric Properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED) in Cypriot Children and Adolescents. *European Journal of Psychological Assessment, 29*(1), 19-27. doi: 10.1027/1015-5759/a000116

Essau, C. A., Muris, P., & Ederer, E. M. (2002). Reliability and validity of the Spence Children's Anxiety Scale and the Screen for Child Anxiety Related Emotional Disorders in German children. *Journal of Behavior Therapy and Experimental Psychiatry, 33*(1), 1-18. doi: 10.1016/s0005-7916(02)00005-8

Gardner, W., Lucas, A., Kolko, D. J., & Campo, J. V. (2007). Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*(5), 611-618. doi: 10.1097/chi.0b013e318032384b

Greenberg, P. E., Sisitsky, T., Kessler, R. C., Finkelstein, S. N., Berndt, E. R., Davidson, J. R., & Fyer, A. J. (1999). The economic burden of anxiety disorders in the 1990s. *The Journal of clinical psychiatry, 60*(7), 427-435. doi: 10.4088/JCP.v60n0702

Grills, A. E., & Ollendick, T. H. (2003). Multiple informant agreement and the anxiety disorders interview schedule for parents and children. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*(1), 30-40. doi: 10.1097/01.chi.0000024899.60748.ee

Guyer, A. E., Lau, J. Y., McClure-Tone, E. B., Parrish, J., Shiffrin, N. D., Reynolds, R. C., ... & Ernst, M. (2008). Amygdala and ventrolateral prefrontal cortex function during anticipated peer evaluation in pediatric social anxiety. *Archives of general psychiatry, 65*(11), 1303-1312. doi: 10.1001/archpsyc.65.11.1303

Hale, W. W., Crocetti, E., Raaijmakers, Q. A. W., & Meeus, W. H. J. (2011). A meta-analysis of the cross-cultural psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED). *Journal of Child Psychology and Psychiatry, 52*(1), 80-90. doi: 10.1111/j.1469-7610.2010.02285.x

Hariz, N., Bawab, S., Atwi, M., Tavitian, L., Zeinoun, P., Khani, M., . . . Maalouf, F. T. (2013). Reliability and validity of the Arabic Screen for Child Anxiety Related Emotional Disorders (SCARED) in a clinical sample. *Psychiatry Research, 209*(2), 222-228. doi: 10.1016/j.psychres.2012.12.002

Hofmeijer-Sevink, M. K., Batelaan, N. M., van Megen, H., Penninx, B. W., Cath, D. C., van den Hout, M. A., & van Balkom, A. (2012). Clinical relevance of comorbidity in anxiety disorders: A report from the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Affective Disorders, 137*(1-3), 106-112. doi: 10.1016/j.jad.2011.12.008

Hollingshead, A. B. (1957). Two-Factor Index of Social Position. New Haven, Conn.: Yale University Department of Sociology.

Hollingshead, A. B. (1975). Four factor index of social status.

Isolan, L., Salum, G. A., Osowski, A. T., Amaro, E., & Manfro, G. G. (2011). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED) in

Brazilian children and adolescents. *Journal of Anxiety Disorders, 25*(5), 741-748. doi: 10.1016/j.janxdis.2011.03.015

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children – Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. [Article] *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 980-988.

Kessel, E. M., Kujawa, A., Hajcak Proudfit, G., & Klein, D. N. (2015). Neural reactivity to monetary rewards and losses differentiates social from generalized anxiety in children. *Journal of Child Psychology and Psychiatry*, *56*(7), 792-800. doi: 10.1111/jcpp.12355

Klein, R. G. (1991). Parent-child agreement in clinical-assessment of anxiety and other psychopathology-a review. *Journal of Anxiety Disorders, 5*(2), 187-198. doi: 10.1016/0887-6185(91)90028-r

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Lowe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine, 146*(5), 317-325.

Ladouceur, C. D., Dahl, R. E., Birmaher, B., Axelson, D. A., & Ryan, N. D. (2006). Increased error-related negativity (ERN) in childhood anxiety disorders: ERP and source localization. *Journal of Child Psychology and Psychiatry*, *47*(10), 1073-1082. doi: 10.1111/j.1469-7610.2006.01654.x

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., 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. doi: 10.1016/j.jaac.2010.05.017

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. doi: 10.1002/1520-6394(2000)12:2<85::aid-da4>3.0.co;2-2

Muris, P., Dreessen, L., Bogels, S., Weckx, M., & van Melick, M. (2004). A questionnaire for screening a broad range of DSM-defined anxiety disorder symptoms in clinically referred children and adolescents. *Journal of Child Psychology and Psychiatry*, *45*(4), 813-820. doi: 10.1111/j.1469-7610.2004.00274.x

Muris, P., Merckelbach, H., Kindt, M., Bogels, S., Dreessen, L., van Dorp, C., . . . Snieder, N. (2001). The utility of Screen for Child Anxiety Related Emotional Disorders (SCARED) as a tool for identifying children at high risk for prevalent anxiety disorders. *Anxiety Stress and Coping, 14*(3), 265-283. doi: 10.1080/10615800108248357

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. doi: 10.1016/S0005-7967(01)00056-0

Muris, P., Merckelbach, H., Mayer, B., van Brakel, A., Thissen, S., Moulaert, V., & Gadet, B. (1998a). The Screen for Child Anxiety Related Emotional Disorders (SCARED) and traditional childhood anxiety measures. *Journal of Behavior Therapy and Experimental Psychiatry, 29*(4), 327-339. doi: 10.1016/s0005-7916(98)00023-8

Muris, P., Merckelbach, H., Schmidt, H., & Mayer, B. (1998b). 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. doi: 10.1016/S0191-8869(98)00130-5

Muris, P., Merckelbach, H., van Brakel, A., & Mayer, B. (1999). The revised version of the screen for child anxiety related emotional disorders (SCARED-R): Further evidence for its reliability and validity. *Anxiety Stress and Coping, 12*(4), 411-425. doi: 10.1080/10615809908249319

Muris, P., Schmidt, H., & Merckelbach, H. (2000). Correlations among two self-report questionnaires for measuring DSM-defined anxiety disorder symptoms in children: the Screen for Child Anxiety Related Emotional Disorders and the Spence Children's Anxiety Scale. *Personality and Individual Differences, 28*(2), 333-346. doi: 10.1016/s0191-8869(99)00102-6

Muris, P., & Steerneman, P. (2001). The Revised version of the Screen for child Anxiety Related Emotional Disorders (SCARED-R): First evidence for its reliability and validity in a clinical sample. *British Journal of Clinical Psychology, 40*, 35-44. doi: 10.1348/014466501163463

Stein, M. B., & Kean, Y. M. (2000). Disability and quality of life in social phobia: Epidemiologic findings. *American Journal of Psychiatry, 157*(10), 1606-1613. doi: 10.1176/appi.ajp.157.10.1606

Stevanovic, D. (2012). Childhood depression and anxiety disorders in Serbia: a psychometric study of four screening questionnaires. [Letter]. *Epidemiology and Psychiatric Sciences, 21*(1), 111-116. doi: 10.1017/s2045796011000655

Su, L. Y., Wang, K., Fan, F., Su, Y., & Gao, X. P. (2008). Reliability and validity of the screen for child anxiety related emotional disorders (SCARED) in Chinese children. *Journal of Anxiety Disorders, 22*(4), 612-621. doi: 10.1016/j.janxdis.2007.05.011

Vigil-Colet, A., Canals, J., Cosi, S., Lorenzo-Seva, U., Ferrando, P. J., Hernandez-Martinez, C., Domenech, E. (2009). The factorial structure of the 41-item version of the Screen for Child Anxiety Related Emotional Disorders (SCARED) in a Spanish population of 8 to 12 years-old. *International Journal of Clinical and Health Psychology, 9*(2), 313-327.

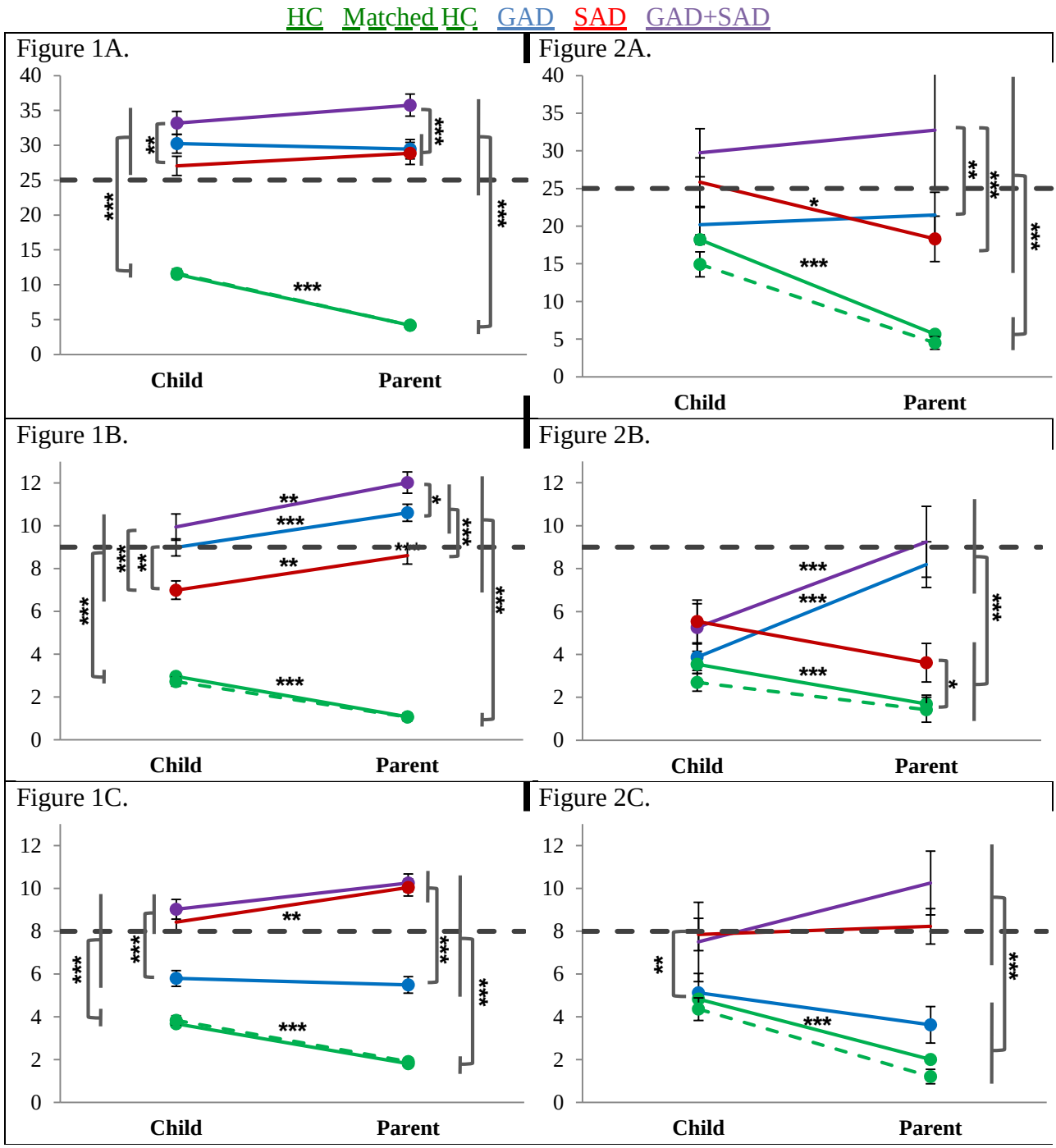
Weitkamp, K., Romer, G., Rosenthal, S., Wiegand-Grefe, S., & Daniels, J. (2010). German Screen for Child Anxiety Related Emotional Disorders (SCARED): Reliability, Validity, and Cross-Informant Agreement in a Clinical Sample. *Child and Adolescent Psychiatry and Mental Health, 4*(19), 1-8. doi: 10.1186/1753-2000-4-19

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jonsson, B., . . . Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology, 21*(9), 655-679. doi: 10.1016/j.euroneuro.2011.07.018

Wren, F. J., Berg, E. A., Heiden, L. A., Kinnamon, C. J., Ohlson, L. A., Bridge, J. A., . . . Bernal, M. P. (2007). Childhood anxiety in a diverse primary care population: Parent-child reports, ethnicity and SCARED factor structure. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*(3), 332-340. doi: 10.1097/chi.0b013e31802f1267

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

Figure 1 & 2. Child- and parent-report Total (A), GA (B), and SA (C) scores for HCs and Patient groups with between- and within-group differences amongst the treatment-seeking NIMH (Figure 1) and non-treatment-seeking SBU (Figure 2) samples



*** $p < .001$, ** $p < .01$, * $p < .05$. Left/Right gray brackets and asterisks reflect between-group Bonferroni post-hoc tests. Asterisks above each line reflect significant within-group differences in pairwise comparisons for child- and parent-report. Circles at the end of the lines reflect significant differences (i.e. greater than or less than) from the clinical cutoff. In Figure 2C, left brackets indicate significant differences between the HC/Matched HC group and the SAD group.

Table 1. Demographics

	N	% Female	Age M ± SD
A. Study 1: Treatment-Seeking Sample	**585**	**51**	**12.68 ± 2.78**
Healthy	364	51	13.18 ± 2.70
Anxious	221	53	11.86 ± 2.70
GAD	103	55	11.75 ± 2.60
SAD	66	48	11.93 ± 2.84
GAD+SAD	52	54	12.69 ± 2.78
Matched Healthy	221	53	12.24 ± 2.59
B. Study 2: Non-Treatment-Seeking Sample	**331**	**49**	**9.17 ± 0.39**
Healthy	298	50	9.17 ± 0.40
Anxious	33	42	9.17 ± 0.32
GAD	16	44	9.14 ± 0.21
SAD	13	54	9.22 ± 0.43
GAD+SAD	4	25	9.13 ± 0.20
Matched Healthy	33	42	9.21 ± 0.35

Table 2. Means and standard deviations of child- and parent-report SCARED scores

	SCARED-Child			ED-Parent		
	E	**GA Subscale M±SE**	**SA Subscale M±SE**	**Total M±SE**	**GA Subscale M±SE**	**SA Subscale M±SE**
A. Study 1: Treatment-Seeking Sample						
Healthy	11.48±0.46	2.96±0.15	3.67±0.16	4.19±0.29	1.07±0.1	1.81±0.13
Anxious	29.98±0.86	8.61±0.29	7.35±0.26	30.70±0.90	10.33±0.29	7.94±0.29
GAD	30.23±1.35	8.99±0.4†	5.79±0.37	29.44±1.39†	10.61±0.4	5.49±0.38
SAD	27.19±11.29†	6.99±0.53†	8.42±0.43†	28.84±1.58†	8.61±0.57†	10.04±0.4
GAD+SAD	32.95±11.61†	9.94±0.61	9.02±0.46	35.76±1.57	12.02±0.5†	10.25±0.42
B. Study 2: Non-Treatment-Seeking Sample						
Healthy	18.21±0.62	3.54±0.17	4.84±0.18	5.66±0.3	1.69±0.13	2.01±0.15
Anxious	23.58±1.80†	4.70±0.52	6.49±0.60†	21.61±2.15	6.52±0.76†	6.24±0.71†
GAD	20.19±2.31†	3.88±0.62	5.13±0.9	21.5±3†	8.19±1.07†	3.63±0.85
SAD	25.86±3.22†	5.54±1.00†	7.85±0.75†	18.21±3	3.62±0.9	8.23±0.83†
GAD+SAD	29.75±3.20†	5.25±1.11†	7.5±1.85†	32.75±7.6†	9.25±1.65†	10.25±1.49†

† Normally distributed according to Shapiro-Wilk test ($p > .05$). All other distributions are non-normal.

Table 3. Sensitivity, specificity, PPV, NPV, and OCC of cutoff scores for child- and parent-report on Total, GA, and SA subscales

	SCARED-Child			SCARED-Parent		
	Total Score (≥25)	**GA Subscale Score (≥9)**	**SA Subscale Score (≥8)**	**Total Score (≥25)**	**GA Subscale Score (≥9)**	**SA Subscale Score (≥8)**
A. Study 1: Treatment-Seeking Sample						
Sensitivity	64.44%	54.97%	64.66%	65.24%	77.55%	79.28%
Specificity	92.33%	91.34%	84.97%	99.38%	90.89%	92.38%
PPV	84.26%	70.34%	53.19%	98.56%	76.51%	73.33%
NPV	80.30%	84.44%	90.10%	81.38%	91.36%	94.40%
OCC	81.48%	81.44%	80.72%	85.88%	87.19%	89.64%
B. Study 2: Non-Treatment-Seeking Sample						
Sensitivity	36.36%	5.00%	64.71%	36.36%	55.00%	70.59%
Specificity	76.17%	91.96%	78.98%	99.66%	98.06%	95.19%
PPV	14.46%	3.85%	14.29%	92.31%	64.71%	44.44%
NPV	91.53%	93.77%	97.64%	93.35%	97.12%	98.34%
OCC	72.21%	86.71%	78.25%	93.31%	95.44%	93.92%

PPV = Positive Predictive Value, NPV = Negative Predictive Value, OCC = Overall Correct Classification

Table 4. Bivariate Pearson correlation coefficients of SCARED-Child and SCARED-Parent Report

	Total Score	**GA Subscale**	**SA Subscale**
A. Study 1: Treatment-Seeking Sample			
Full Sample	.624***	.622***	.550***
Healthy Youth	.266***	.180**	.303***
Anxious Youth	.149*	.237**	.356***
GAD	.158	.185	.271**
SAD	-.059	.003	-.031
GAD+SAD	.233	.426**	.213
B. Study 2: Non-Treatment-Seeking Sample			
Full Sample	.182**	.117*	.217***
Healthy Youth	.025	.036	.121*
Anxious Youth	.515**	.235	.408*
GAD	.372	.462	.270
SAD	.705*	.457	.025
GAD+SAD	.839	.398	.710

*** $p < .001$, ** $p < .01$, *$p < .05$