



Do youth anxiety measures assess the same construct consistently throughout treatment? Results are...complicated

Jonathan C. Rabner¹ · Thomas M. Olino¹ · Anne Marie Albano² · Golda S. Ginsburg³ · Scott N. Compton⁴ · John Piacentini⁵ · Dara Sakolsky⁶ · Boris Birmaher⁶ · Elizabeth Gosch⁷ · Philip C. Kendall¹

Accepted: 22 February 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Interventionists interpret changes in symptoms as reflecting response to treatment. However, changes in symptom functioning *and* the measurement of the underlying constructs may be reflected in reported change. Longitudinal measurement invariance (LMI) is a statistical approach that assesses the degree to which measures consistently capture the same construct over time. We examined LMI in measures of anxiety severity/symptoms [i.e., Pediatric Anxiety Rating Scale (PARS), Multidimensional Anxiety Scale for Children (MASC), Screen for Child Anxiety and Related Disorders (SCARED)] in anxious youth at baseline and posttreatment. Initial fit was inadequate for 27 of 38 baseline and posttreatment models, but model modifications resulted in acceptable fit. Tests of LMI supported scalar invariance for the PARS and many, but not all, MASC and SCARED subscales. Findings suggest that the PARS, and many MASC and SCARED subscales can accurately be used to measure change over time, however, others may reflect changes in measurement properties.

Keywords Measurement invariance · Anxiety · Treatment · Assessment · Child and adolescent

The consistent measurement of a construct is critical for evaluating change in treatment outcome studies (i.e., the anxiety construct measured at baseline is the same anxiety construct measured at posttreatment). Furthermore, longitudinal research with youth occurs across their development, be it in the short- or long-term. Without demonstrations

that the measurement is consistent across time, it is unclear whether changes reflect true changes, or changes in measurement properties across time. Longitudinal measurement invariance (LMI) is a statistical approach to test this assumption by examining equality of measurement properties over time [1]. For measures of anxiety in youth this assumption is infrequently checked, particularly in the context of treatment.

A review of the literature found three studies that examined LMI in measures of youth anxiety: the Revised Child Anxiety and Depression Scale (RCADS) [2], the Social Anxiety Scale for Adolescents (SAS-A) [3], and the Screen for Child Anxiety and Related Disorders (SCARED) [4]. There was support for all levels of invariance (i.e., configural, metric, and scalar) for the RCADS and the SAS-A, both self-report measures, indicating that the same construct is measured over time [2, 3]; however, there was inconsistent support for different levels of invariance across subscales for SCARED parent- and youth self-reports [4]. All three studies used community samples and naturalistic follow-ups from 2.5 years to 4 years. However, studies have yet to examine LMI in treatment studies of youth anxiety. Evaluating LMI in treatment studies is integral in the determination of whether changes over time are due to changes resulting

✉ Jonathan C. Rabner
jrabner@temple.edu

¹ Department of Psychology and Neuroscience, Temple University, Philadelphia, PA, USA

² Department of Psychiatry, Columbia University, New York City, NY, USA

³ Department of Psychiatry, University of Connecticut School of Medicine, West Hartford, CT, USA

⁴ Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA

⁵ Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA

⁶ Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

⁷ Department of Clinical Psychology, Philadelphia College of Osteopathic Medicine, Philadelphia, PA, USA

from the treatment or potentially influenced by changes in the measurement properties of the measures used (e.g., if items relate differently to the latent anxiety construct before and after treatment, perhaps due to increased psychoeducation or a change in severity, that may result in the observed change in scores).

There are a small number of commonly used measures of anxiety symptomatology in clinical trials. The Pediatric Anxiety Rating Scale (PARS) [5], the Multidimensional Anxiety Scale for Children (MASC) [6], and the Screen for Child Anxiety Related Disorders (SCARED) [7, 8] are frequently used measures in treatment outcome studies (e.g., PARS: [9–11]; MASC: [12–14]; SCARED: [15–17]). The PARS is an Independent Evaluator-rated measure of anxiety severity and impairment based on interviewing both youth and parents, and the MASC and SCARED are measures of anxiety symptoms with both a child- and parent-report. The PARS has a single-factor structure [5], the MASC a four-factor structure [6], and the SCARED a five-factor structure [7]. Despite their use in treatment outcome studies, previous psychometric evaluations are largely cross-sectional or have focused on their ability to detect change in anxiety. However, these analyses assume that the pre- and post-intervention assessments are equivalent and that the change detected is actually a change in the construct measured. The question of whether these measures assess the same anxiety construct consistently throughout treatment (e.g., baseline and post-treatment) has yet to be evaluated.

The present study examined longitudinal measurement invariance of five measures of anxiety severity and symptoms (i.e., total and/or subscale scores for the PARS, MASC parent-report, MASC child-report, SCARED parent-report, SCARED child-report) in a large sample of anxious youth at baseline and posttreatment. A series of models with increasing levels of invariance was estimated. Due to prior findings (e.g., Olino and colleagues [4]), we hypothesized that thresholds or intercepts may change across time and, thus, scalar invariance would not consistently be found (e.g., for the SCARED). This may reflect that youth have changes in thresholds of experienced anxiety needed to endorse higher level severity options.

Methods

Sample

The study included 488 youth, aged 7–17 years ($M = 10.69$, $SD = 2.80$), who participated in the Child-Adolescent Anxiety Multimodal Study (CAMS). The sample was 49.6% female and 25.4% of the sample was characterized as low socioeconomic status. 78.9% of the sample was White,

9.0% was Black, 2.5% was Asian, and 9.6% of the sample identified as a different racial group. All participants met diagnostic criteria for a principal diagnosis of generalized anxiety disorder (GAD), social anxiety disorder (Soc), and/or separation anxiety disorder (Sep). 35.9% ($n = 175$) of participants met criteria for all three diagnoses, 27.7% ($n = 135$) met criteria for Soc and GAD, 8.0% ($n = 39$) met criteria for Sep and GAD, 6.8% ($n = 33$) met criteria for Soc and Sep, 6.8% ($n = 33$) met criteria for GAD, 11.5% ($n = 135$) met criteria for Soc, and 3.3% ($n = 16$) met criteria for Sep. Comorbid diagnoses of lesser severity include attention-deficit hyperactivity disorder (11.9%; $n = 58$), oppositional defiant disorder (9.4%; $n = 46$), obsessive compulsive disorder (8.6%; $n = 42$), and other internalizing disorders (43.6%; $n = 213$). For more details about participants, see Kendall and colleagues [18].

Measures

Pediatric Anxiety Rating Scale (PARS)

The PARS is an Independent Evaluator (IE)-rated assessment of youth anxiety severity and impairment [5]. In the treatment study where this sample originated [19], a 6-item PARS total score was computed by summing six items assessing anxiety severity, frequency, distress, avoidance, and interference during the previous week. PARS item 1 (number of symptoms) is typically not included in the total score. Items were rated on a six-point scale from 0 to 5, with higher scores indicating greater impairment and severity. Historically, a 5-item PARS total score that further excludes the physical symptoms item has also been examined [5]. The 5-item PARS total score demonstrated $r = .97$ inter-rater reliability, $r = .55$ 4-week retest reliability, and strong convergent and divergent construct validity [5]. In the present sample, the 6-item PARS demonstrated $\alpha = 0.72$ internal consistency at baseline and $\alpha = 0.89$ internal consistency at posttreatment.

Multidimensional Anxiety Scale for Children (MASC)

The MASC is a 39-item child- (MASC-C) and parent-report (MASC-P) measure of anxiety symptoms in the prior two weeks [6]. Items were rated on a four-point scale from 0 to 3, with higher scores indicating greater anxiety symptoms. The MASC consists of four subscales: physical symptoms (12 items), social anxiety (9 items), harm avoidance (9 items), and separation anxiety (9 items). The MASC demonstrated good convergent and divergent validity, retest reliability, and diagnostic accuracy [6, 20–23]. In the present sample, internal consistency (α) of the MASC-C subscales and total score ranged from 0.69 (separation anxiety

subscale) to 0.91 (total score) at baseline and 0.72 (separation anxiety subscale) to 0.93 (total score) at posttreatment, and internal consistency of the MASC-P subscales and total score ranged from 0.67 (harm avoidance subscale) to 0.88 (total score) at baseline and 0.73 (separation anxiety subscale) to 0.93 (social anxiety subscale) at posttreatment.

Screen for Child Anxiety Related Disorders (SCARED)

The SCARED is a 41-item child- (SCARED-C) and parent-report (SCARED-P) measure of anxiety symptoms [7, 8]. Items were rated on a three-point scale from 0 to 2, with higher scores indicating a greater presence of anxiety symptoms. Respondents are typically instructed to consider the past three months, however, in the present study, the time frame was reduced to the prior two weeks due to repeated administration. The SCARED consists of five subscales: panic/somatic (13 items), general anxiety (9 items), separation anxiety (8 items), social phobia (7 items), and school phobia (4 items). As CAMS excluded youth who refused to attend school due to anxiety, the school phobia subscale was not examined in the present study. The SCARED demonstrated good retest reliability as well as strong convergent and divergent validity [7, 8, 24]. In the present sample, internal consistency (α) of the SCARED-C subscales and total score ranged from 0.83 (separation anxiety subscale) to 0.94 (total score) at baseline and 0.83 (separation anxiety subscale) to 0.95 (total score) at posttreatment, and internal consistency of the SCARED-P subscales and total score ranged from 0.83 (generalized anxiety subscale) to 0.90 (total score and social phobia subscale) at baseline and 0.83 (separation anxiety subscale) to 0.94 (social phobia subscale) at posttreatment.

Procedures

Institutional review board approval and participant informed consent and assent were obtained. Treatment spanned a 12-week period with assessments completed by the child and parent as well as interviews conducted by IEs at multiple timepoints. Only data from assessments conducted at baseline and posttreatment (i.e., 12 weeks following the start of treatment) were used in the present study. Cognitive-behavioral therapy (CBT; *Coping Cat*) consisted of 14 sessions over 12 weeks with two parent/guardian-only sessions occurring on the same day as the child session. Medication (sertraline) was administered at a dose up to 200 mg per day. For a more complete description of the design, see Compton and colleagues [25].

Data Analysis

All analyses were conducted in R [26], version 3.5.2, using the lavaan package [27] and irr package [28]. For measures with total and subscale scores, single factor models for the full set of items as well as models for individual subscales were estimated to reflect the different ways the measures are used. Models examining the MASC and SCARED were estimated using the weighted least square mean and variance (WLSMV) estimator due to 4-point and 3-point scale categorical responses, respectively. As PARS items contain six response options, items were treated as continuous, and models were estimated using the robust maximum likelihood (MLR) estimator. Thus, for the MASC and SCARED, thresholds were modeled, and intercepts were modeled for the PARS.

Acceptable and good model fit was indicated by a Comparative Fit Index (CFI) score greater than 0.95 and 0.97 and a Root Mean-Square Error of Approximation (RMSEA) score less than 0.08 and 0.05, respectively [29]. A non-significant χ^2 test also indicates good model fit, however, this index has been known to be overly sensitive in large samples. When acceptable model fit was not indicated, model residuals and modification indices were examined to determine whether inclusion in the model of any covariances between variables would improve model fit. This process was repeated until the “revised” model reached adequate fit. For tests of LMI, equivalent “reconciled” models were used where added covariances to the baseline model were also included in the posttreatment model and vice versa.

Subsequently, a series of models with increasing levels of LMI were estimated for all measures. Residual covariances between the same item across time were permitted in each model. Models were specified freely estimating all factor loadings, intercepts, and thresholds and fixing the variance of the latent variable to 1. The following sequence of models was tested: configural invariance, metric (or weak) invariance, and scalar (or strong) invariance. The configural invariance model assigns the same factor structure to both the baseline and posttreatment latent factors (i.e., the same five PARS items are assigned as indicators in both models) with minimal other constraints (i.e., only the first item's intercept and factor loading are constrained to be equal). Next, the metric invariance model examines differences in factor loadings by timepoint by placing equality constraints on the loadings for each observed indicator. Finally, the scalar invariance model examines differences in intercepts (for PARS) or thresholds (for MASC and SCARED) by timepoint by placing equality constraints on the intercepts for each item or thresholds between response options for each item. Put simply, scalar invariance indicates that mean-level comparisons can be conducted. Measurement

invariance was indicated by a change in the $CFI \leq 0.01$ and the $RMSEA \leq 0.015$ [30]. A metric of effect size (d_{MACS}) [31] is also provided for each item from each measure/subscale to aid in interpretation of the degree of invariance. d_{MACS} integrates both factor loadings and intercepts/thresholds into a single effect size metric. Values were interpreted as small (0.4), medium (0.6), and large (0.8) effects in accordance with guidelines for practical importance by Nye and colleagues [32].

When model fit substantively diminished (i.e., decrease in $CFI > 0.01$ or increase in $RMSEA > 0.015$), partial invariance was assessed. Non-invariant items were identified by examining differences in estimates of model parameters. Equality constraints were lifted starting with the parameter with the largest difference and continued until the model achieved adequate fit. When equality constraints on thresholds required lifting, specified thresholds were lifted individually one at a time. Unconstrained parameters remained unconstrained for subsequent measurement invariance models. When lifting equality constraints did not substantially impact model fit, it was deemed that partial invariance was not found.

Finally, as a means of estimating the substantive impact on the item parameters with and without residual covariances, a sensitivity analysis was conducted by estimating intraclass correlations (ICC). ICCs were calculated based on absolute agreement, 2-way mixed effects models and compared factor loadings, thresholds/intercepts, and residual variances. ICCs greater than 0.5, 0.75, and 0.9 indicate moderate, good, and excellent agreement, respectively [33]. Greater agreement would indicate that the addition of residual covariances did not substantively impact model parameters.

Results

Data from the MASC-C, MASC-P, and the SCARED-C was present for all 488 participants at baseline. Baseline data were missing from only one participant on the PARS and from only three participants on the SCARED-P. Due to attrition, data at posttreatment were available for 439 (90.0%) participants on the PARS and the SCARED-C, for 436 (89.3%) participants on the MASC-C and the MASC-P, and for 435 (89.1%) participants on the SCARED-P. Attrition rates differed by treatment condition, with significantly lower rates for participants in the CBT condition (4.3%) than in the medication (17.3%) or placebo (19.7%) groups. The combination group (9.3%) did not differ with any other treatment condition [19]. Results of tests of unidimensionality and LMI are reported separately for the individual measures.

Table 1 Fit statistics for PARS baseline, posttreatment, and longitudinal measurement invariance models

	χ^2 (df)	CFI	RMSEA [90% CI]	Δ CFI	Δ RMSEA
Tests of Unidimensionality					
Baseline Model: Initial	21.59 (9)	0.974	0.054 [0.025, 0.083]		
Baseline Model: Reconciled	7.11 (7)	1.000	0.006 [0.000, 0.056]		
Posttreatment Model: Initial	51.51 (9)	0.971	0.104 [0.078, 0.133]		
Posttreatment Model: Revised ^a	10.26 (7)	0.998	0.033 [0.000, 0.073]		
Posttreatment Model: Reconciled ^a	10.26 (7)	0.998	0.033 [0.000, 0.073]		
Tests of LMI					
Configural Invariance	76.76 (43)	0.984	0.043 [0.027, 0.058]		
Metric Invariance	91.64 (48)	0.979	0.046 [0.031, 0.060]	-0.005	0.003
Scalar Invariance	113.43 (53)	0.971	0.051 [0.038, 0.065]	-0.008	0.005

Note. Initially examined models without residual covariances are labeled “Initial”, when needed, subsequent models with included residual covariances are labeled “Revised”, and final models with equivalent residual covariances at baseline and at posttreatment are labeled “Reconciled”; PARS=Pediatric Anxiety Rating Scale; CFI=comparative fit index; RMSEA=root mean square error of approximation; LMI=longitudinal measurement invariance; All χ^2 values were significant at $p < .05$ except for the revised and reconciled models

^a Denotes that revised and reconciled models are the same model

PARS

Tests of Unidimensionality

The initial baseline model was an excellent fit to the data, however, the initial posttreatment model was not. A revised posttreatment model, including two residual covariance paths, was a good fit to the data. Specific added covariances can be found in Supplemental Materials. Model fit for final reconciled models with equivalent residual covariance paths can be found in Table 1. As noted in the methods section, historically, a 5-item PARS total score has also been used. A similar pattern of fit to the data was found for models used in tests of unidimensionality and LMI, and can be found in the Supplemental Materials.

Table 2 Fit statistics for MASC-C baseline and posttreatment models

	χ^2 (df)	CFI	RMSEA [90% CI]
Physical Symptoms Subscale			
Baseline Model: Initial	95.54 (54)	1.000	0.040 [0.026, 0.053]
Posttreatment Model: Initial	148.19 (54)	0.991	0.064 [0.052, 0.076]
Social Anxiety Subscale			
Baseline Model: Initial	290.67 (27)	0.995	0.142 [0.128, 0.157]
Baseline Model: Revised ^a	68.41 (22)	1.000	0.066 [0.049, 0.084]
Baseline Model: Reconciled ^a	68.41 (22)	1.000	0.066 [0.049, 0.084]
Posttreatment Model: Initial	135.74 (27)	0.997	0.096 [0.081, 0.113]
Posttreatment Model: Revised	91.77 (26)	0.999	0.076 [0.060, 0.094]
Posttreatment Model: Reconciled	61.72 (22)	1.000	0.065 [0.046, 0.084]
Harm Avoidance Subscale			
Baseline Model: Initial	132.91 (27)	0.964	0.090 [0.075, 0.105]
Baseline Model: Revised ^a	67.41 (26)	0.989	0.057 [0.041, 0.074]
Baseline Model: Reconciled ^a	67.41 (26)	0.989	0.057 [0.041, 0.074]
Posttreatment Model: Initial	126.11 (27)	0.984	0.093 [0.077, 0.109]
Posttreatment Model: Revised ^a	68.94 (26)	0.995	0.062 [0.045, 0.080]
Posttreatment Model: Recon- ciled ^a	68.94 (26)	0.995	0.062 [0.045, 0.080]
Separation Anxiety Subscale			
Baseline Model: Initial	85.74 (27)	0.979	0.067 [0.051, 0.083]
Posttreatment Model: Initial	68.96 (27)	0.985	0.060 [0.043, 0.078]

Note. Initially examined models without residual covariances are labeled “Initial”, when needed, subsequent models with included residual covariances are labeled “Revised”, and final models with equivalent residual covariances at baseline and at posttreatment are labeled “Reconciled”; MASC-C=Multidimensional Anxiety Scale for Children – Child-report; CFI=comparative fit index; RMSEA=root mean square error of approximation; All χ^2 values were significant at $p < .05$

^a Denotes that revised and reconciled models are the same model

Tests of LMI

The configural invariance model was an excellent fit to the data and, subsequently, the metric invariance and scalar invariance models were a good fit to the data (see Table 1 for fit statistics). Changes in the CFI and RMSEA between models were within acceptable limits indicating that model fit did not deteriorate with the inclusion of constraints. Thus, it is possible to conclude that the PARS total score has scalar

invariance. All PARS items showed a large effect size difference ($d_{\text{MACS}} > 0.8$).

MASC and SCARED Total Scores

There was an attempt to fit single factor models for the full set of items at baseline and posttreatment, however, all models were a poor fit to the data (see Supplemental Materials). Acceptable fit was not attained despite attempts to add residual covariances. Likewise, an attempt was made to fit models with a second-order latent factor structure where the anxiety measure was specified as a second-order latent factor indicated by its subscales, which in turn were indicated by the items comprising the subscale. Though the baseline SCARED-C model demonstrated acceptable fit [CFI=0.977; RMSEA=0.050 (90% CI=0.046–0.053)], all other models either failed to converge or were a poor fit to the data and did not improve following attempts to add residual covariances.

MASC-C Subscales

Tests of Unidimensionality

For the physical symptoms subscale and the separation anxiety subscale, initial baseline and posttreatment models were an excellent fit to the data. However, for the social anxiety and harm avoidance subscales initial models were a poor fit to the data (see Table 2). Revised social anxiety subscale models, including five residual covariance paths in the baseline model and one residual covariance path in the posttreatment model, were an acceptable fit to the data. Similarly, revised harm avoidance subscale models, each including one residual covariance path, were an acceptable fit to the data. Specific added covariances can be found in Supplemental Materials. Model fit for final reconciled models with equivalent residual covariance paths can be found in Table 2.

Tests of LMI

The configural invariance model and, subsequently, the metric invariance model for all four subscales had good fit (see Table 3 for all fit statistics). Changes in the CFI and RMSEA between these models were within acceptable limits indicating that model fit did not deteriorate with the inclusion of constraints. The scalar invariance model for the physical symptoms subscale and separation anxiety subscale both had excellent fit and changes were within acceptable limits. Good fit was also found for the scalar invariance model for the harm avoidance subscale, however, changes in both the CFI and RMSEA were in excess of acceptable limits.

Table 3 Fit statistics for MASC-C longitudinal measurement invariance models

	χ^2 (df)	CFI	RMSEA [90% CI]	Δ CFI	Δ RMSEA
Physical Symptoms Subscale					
Configural Invariance	318.00 (239)	1.000	0.028 [0.019, 0.036]		
Metric Invariance	329.62 (250)	0.997	0.027 [0.019, 0.035]	-0.003	-0.001
Scalar Invariance	393.05 (285)	0.995	0.030 [0.022, 0.037]	-0.002	0.003
Social Anxiety Subscale					
Configural Invariance	350.48 (115)	0.997	0.069 [0.061, 0.077]		
Metric Invariance	365.91 (127)	0.996	0.064 [0.056, 0.072]	-0.001	-0.005
Scalar Invariance	4501.74 (153)	0.900	0.243 [0.237, 0.250]	-0.096	0.179
Partial Scalar Invariance	3815.68 (143)	0.904	0.245 [0.238, 0.251]	-0.092	0.181
Harm Avoidance Subscale					
Configural Invariance	243.45 (123)	0.990	0.048 [0.039, 0.057]		
Metric Invariance	242.26 (131)	0.988	0.045 [0.036, 0.053]	-0.002	-0.003
Scalar Invariance	380.91 (157)	0.972	0.058 [0.042, 0.058]	-0.016	0.013
Partial Scalar Invariance	323.87 (152)	0.980	0.052 [0.044, 0.059]	-0.008	0.007
Separation Anxiety Subscale					
Configural Invariance	252.35 (125)	0.983	0.049 [0.040, 0.058]		
Metric Invariance	241.15 (133)	0.983	0.044 [0.035, 0.053]	0	-0.005
Scalar Invariance	299.02 (159)	0.977	0.046 [0.038, 0.053]	-0.006	0.002

Note. MASC-C=Multidimensional Anxiety Scale for Children – Child-report; CFI=comparative fit index; RMSEA=root mean square error of approximation; All χ^2 values were significant at $p < .05$

Partial scalar invariance was attained after the freeing of five thresholds (i.e., the first threshold for items 2, 9, 28, and 36 and the second threshold for item 28). Furthermore, the scalar invariance model for the social anxiety subscale demonstrated poor fit. Attempts to free equality constraints on thresholds did not yield a change in fit. Thus, it is possible to conclude that the MASC-C physical symptoms and separation anxiety subscales have scalar invariance, the MASC-C harm avoidance subscale has partial scalar invariance, and the MASC-C social anxiety subscale has metric invariance.

For the physical symptoms subscale, small effect size differences (i.e., $d_{MACS} < 0.4$) were found for 16.7% of items and moderate effect size differences (i.e., $0.5 < d_{MACS} < 0.7$) were found for 75.0% of items. For the social anxiety subscale, moderate effect size differences were found for 66.7% of items and no small effect size differences were found. For the harm avoidance subscale, small effect size differences were found for 22.2% of items and moderate effect size differences were found for 44.4% of items. For the separation anxiety subscale, small effect size differences were found for 33.3% of items and moderate effect size differences were found for 33.3% of items. No large effects (i.e., $d_{MACS} > 0.8$) were found for any item.

MASC-P Subscales

Tests of Unidimensionality

All MASC-P subscales required the addition of residual covariance paths for at least one timepoint. For the physical symptom subscale, the initial posttreatment model was an acceptable fit for the data, however, the initial baseline model was not. A revised baseline model, including two residual covariance paths, was an acceptable fit to the data (specific added covariances can be found in Supplemental Materials). For the remaining models, all initial models were a poor fit to the data (see Table 4). Revised social anxiety subscale models, each including six residual covariance paths, were an acceptable fit to the data. Similarly, revised harm avoidance subscale models, including two residual covariance paths in the baseline model and one residual covariance path in the posttreatment model, were an acceptable fit to the data. Finally, revised separation anxiety subscale models, including three residual covariance paths in the baseline model and one residual covariance path in the posttreatment model, were an acceptable fit to the data. Model fit for all final reconciled models with equivalent residual covariance paths can be found in Table 4.

Table 4 Fit statistics for MASC-P baseline and posttreatment models

	χ^2 (df)	CFI	RMSEA [90% CI]
Physical Symptoms Subscale			
Baseline Model: Initial	284.79 (54)	0.966	0.094 [0.083, 0.105]
Baseline Model: Revised ^a	209.14 (52)	0.979	0.079 [0.068, 0.090]
Baseline Model: Reconciled ^a	209.14 (52)	0.979	0.079 [0.068, 0.090]
Posttreatment Model: Initial	166.13 (54)	0.987	0.070 [0.058, 0.082]
Posttreatment Model: Reconciled	136.25 (52)	0.991	0.062 [0.049, 0.074]
Social Anxiety Subscale			
Baseline Model: Initial	626.71 (27)	0.983	0.214 [0.200, 0.229]
Baseline Model: Revised	67.54 (21)	1.000	0.068 [0.050, 0.086]
Baseline Model: Reconciled	34.60 (20)	1.000	0.039 [0.015, 0.060]
Posttreatment Model: Initial	414.28 (27)	0.991	0.183 [0.168, 0.199]
Posttreatment Model: Revised	55.46 (21)	1.000	0.062 [0.042, 0.082]
Posttreatment Model: Reconciled	50.25 (20)	1.000	0.059 [0.039, 0.080]
Harm Avoidance Subscale			
Baseline Model: Initial	152.66 (27)	0.926	0.098 [0.083, 0.113]
Baseline Model: Revised	91.85 (25)	0.964	0.074 [0.058, 0.091]
Baseline Model: Reconciled	76.07 (24)	0.974	0.067 [0.050, 0.084]
Posttreatment Model: Initial	116.93 (27)	0.977	0.088 [0.072, 0.105]
Posttreatment Model: Revised	62.51 (26)	0.993	0.057 [0.039, 0.076]
Posttreatment Model: Reconciled	61.36 (24)	0.993	0.060 [0.042, 0.079]
Separation Anxiety Subscale			
Baseline Model: Initial	165.68 (27)	0.966	0.103 [0.088, 0.118]
Baseline Model: Revised ^a	91.10 (24)	0.986	0.076 [0.060, 0.093]
Baseline Model: Reconciled ^a	91.10 (24)	0.986	0.076 [0.060, 0.093]
Posttreatment Model: Initial	108.62 (27)	0.979	0.084 [0.068, 0.101]
Posttreatment Model: Revised	75.75 (26)	0.989	0.067 [0.050, 0.085]
Posttreatment Model: Reconciled	62.12 (24)	0.993	0.061 [0.043, 0.080]

Note. Initially examined models without residual covariances are labeled “Initial”, when needed, subsequent models with included residual covariances are labeled “Revised”, and final models with equivalent residual covariances at baseline and at posttreatment are labeled “Reconciled”; MASC-P = Multidimensional Anxiety Scale for Children – Parent-report; CFI = comparative fit index; RMSEA = root mean square error of approximation; All χ^2 values were significant at $p < .05$

^a Denotes that revised and reconciled models are the same model

Tests of LMI

Configural invariance was not found for the MASC-P social anxiety subscale (see Table 5 for all fit statistics). For the remaining three subscales, the configural invariance model and, subsequently, the metric invariance model had good fit. Changes in the CFI and RMSEA between these models were within acceptable limits indicating that model fit did not deteriorate with the inclusion of constraints. The scalar invariance model for the separation anxiety subscale had good fit and, changes were within acceptable limits.

For the physical symptoms subscale, no participants endorsed the highest option for item 18 at posttreatment so only two thresholds were specified in the scalar invariance model. Good fit was found for the scalar invariance model, however, changes in the CFI were in excess of the acceptable limit. Partial scalar invariance was attained after freeing seven thresholds (i.e., all three thresholds for items 1, the third and second threshold for item 31, and the first threshold for items 27 and 20). Finally, the scalar invariance model for the harm avoidance subscale demonstrated poor fit. Partial scalar invariance was attained after freeing eight thresholds (i.e., all three thresholds for items 9 and the first threshold for items 2, 25, 13, 26, and 21). Thus, it is possible to conclude that the MASC-P separation anxiety subscale

has scalar invariance, and the MASC-P harm avoidance and physical symptoms subscales have partial scalar invariance, however, the MASC-P social anxiety subscale did not even have configural invariance.

For the physical symptoms subscale¹, moderate effect size differences (i.e., $0.5 < d_{\text{MACS}} < 0.7$) were found for 16.7% of items, large effect size differences (i.e., $d_{\text{MACS}} > 0.8$) were found in 41.7% of items, and no small effect size differences (i.e., $d_{\text{MACS}} < 0.4$) were found. For the social anxiety subscale, moderate effect size differences were found for 11.1% of items, large effect size differences were found for 88.9% of items, and no small effect size differences were found. For the harm avoidance subscale, small effect size differences were found for 11.1% of items, moderate effect size differences were found for 22.2% of items, and large effect size differences were found for 22.2% of items. For the separation anxiety subscale, small effect size differences were found for 11.1% of items, moderate effect size differences were found for 55.6% of items, and large effect size differences were found for 33.3% of items.

¹ In order to calculate d_{MACS} for this subscale, the highest level of MASC item 18 at baseline was recoded to equal the level below it, to ensure an equivalent number of thresholds with MASC item 18 at posttreatment.

Table 5 Fit statistics for MASC-P longitudinal measurement invariance models

	χ^2 (df)	CFI	RMSEA [90% CI]	Δ CFI	Δ RMSEA
Physical Symptoms Subscale					
Configural Invariance	492.61 (235)	0.979	0.051 [0.045, 0.057]		
Metric Invariance	473.14 (246)	0.978	0.047 [0.040, 0.053]	-0.001	-0.004
Scalar Invariance	671.81 (280)	0.958	0.057 [0.052, 0.063]	-0.020	0.010
Partial Scalar Invariance	586.44 (273)	0.968	0.052 [0.046, 0.058]	-0.010	-0.005
Social Anxiety Subscale					
Configural Invariance	653.83 (111)	0.990	0.107 [0.099, 0.115]		
Harm Avoidance Subscale					
Configural Invariance	248.01 (119)	0.978	0.050 [0.042, 0.059]		
Metric Invariance	249.73 (127)	0.976	0.048 [0.039, 0.056]	-0.002	-0.002
Scalar Invariance	468.74 (156)	0.945	0.061 [0.054, 0.069]	-0.031	0.013
Partial Scalar Invariance	310.50 (145)	0.966	0.052 [0.044, 0.060]	-0.010	0.004
Separation Anxiety Subscale					
Configural Invariance	296.16 (119)	0.985	0.059 [0.051, 0.068]		
Metric Invariance	279.05 (127)	0.985	0.053 [0.045, 0.062]	0	-0.006
Scalar Invariance	351.74 (153)	0.979	0.055 [0.048, 0.063]	-0.006	0.002

Note. MASC-P=Multidimensional Anxiety Scale for Children – Parent-report; CFI=comparative fit index; RMSEA=root mean square error of approximation; All χ^2 values were significant at $p < .05$

SCARED-C Subscales

Tests of Unidimensionality

For the panic/somatic subscale, initial baseline and post-treatment models were an excellent fit to the data. For the

general anxiety subscale, the initial posttreatment model was an adequate fit for the data, however, the initial baseline model was not. A revised baseline model, including two residual covariance paths, was an acceptable fit to the data (specific added covariances can be found in Supplemental Materials). For the remaining subscales, all initial models were a poor fit to the data (see Table 6). Revised separation anxiety subscale models, including two residual covariance paths in the baseline model and one residual covariance path in the posttreatment model, were an acceptable fit to the data. Similarly, revised social phobia subscale models, including one residual covariance path in the baseline model and three residual covariance paths in the posttreatment model, were an acceptable fit to the data. Model fit for all final reconciled models for the above subscales with equivalent residual covariance paths can be found in Table 6.

Tests of LMI

The configural invariance model and, subsequently, the metric invariance model for all subscales had good fit (see Table 7 for all fit statistics). Changes in the CFI and RMSEA between these models were within acceptable limits indicating that model fit did not deteriorate with the inclusion of constraints. The scalar invariance model for the panic/somatic subscale, the separation anxiety subscale, and the social phobia subscale each had excellent fit and, changes were within acceptable limits. However, the scalar invariance model for the general anxiety subscale demonstrated poor fit. Attempts to free equality constraints on thresholds did not yield a change in fit. Thus, it is possible to conclude that the SCARED-C panic/somatic, separation anxiety, and social phobia subscales have scalar invariance and the SCARED-C general anxiety subscale has metric invariance.

For the panic/somatic subscale, small effect size differences (i.e., $d_{MACS} < 0.4$) were found for 15.4% of items, moderate effect size differences (i.e., $0.5 < d_{MACS} < 0.7$) were found for 38.5% of items, and no large effect size differences (i.e., $d_{MACS} > 0.8$) were found. For the general anxiety subscale, moderate effect size differences were found for 33.3% of items, large effect size differences were found for 11.1% of items, and no small effect size differences were found. For the separation anxiety subscale, moderate effect size differences were found for 50.0% of items and no small or large effect size differences were found. For the social phobia subscale, moderate effect size differences were found for all items.

Table 6 Fit statistics for SCARED-C baseline and posttreatment models

	χ^2 (df)	CFI	RMSEA [90% CI]
Panic/Somatic Subscale			
Baseline Model: Initial	118.63 (65)	0.998	0.042 [0.030, 0.054]
Posttreatment Model: Initial	88.44 (65)	1.000	0.029 [0.010, 0.043]
General Anxiety Subscale			
Baseline Model: Initial	210.34 (27)	0.989	0.119 [0.104, 0.134]
Baseline Model: Revised ^a	98.91 (25)	0.997	0.079 [0.063, 0.095]
Baseline Model: Reconciled	98.91 (25)	0.997	0.079 [0.063, 0.095]
Posttreatment Model: Initial	94.12 (27)	0.996	0.076 [0.060, 0.093]
Posttreatment Model: Reconciled	37.99 (25)	1.000	0.035 [0.005, 0.056]
Separation Anxiety Subscale			
Baseline Model: Initial	196.25 (20)	0.972	0.135 [0.118, 0.153]
Baseline Model: Revised ^a	68.50 (18)	0.995	0.076 [0.058, 0.096]
Baseline Model: Reconciled	68.50 (18)	0.995	0.076 [0.058, 0.096]
Posttreatment Model: Initial	80.78 (20)	0.988	0.083 [0.065, 0.103]
Posttreatment Model: Revised	63.40 (19)	0.993	0.073 [0.054, 0.093]
Posttreatment Model: Reconciled	37.66 (18)	0.999	0.050 [0.027, 0.072]
Social Phobia Subscale			
Baseline Model: Initial	93.60 (14)	0.997	0.109 [0.089, 0.130]
Baseline Model: Revised	41.34 (13)	0.999	0.067 [0.045, 0.091]
Baseline Model: Reconciled	14.21 (10)	1.000	0.030 [0.000, 0.062]
Posttreatment Model: Initial	93.99 (14)	0.997	0.115 [0.093, 0.137]
Posttreatment Model: Revised	19.82 (11)	1.000	0.043 [0.004, 0.073]
Posttreatment Model: Reconciled	14.09 (10)	1.000	0.031 [0.000, 0.065]

Note. Initially examined models without residual covariances are labeled “Initial”, when needed, subsequent models with included residual covariances are labeled “Revised”, and final models with equivalent residual covariances at baseline and at posttreatment are labeled “Reconciled”; SCARED-C=Screen for Child Anxiety Related Disorders – Child-report; CFI=comparative fit index; RMSEA=root mean square error of approximation; All χ^2 values were significant at $p < .05$ except for both reconciled social phobia subscale models

^a Denotes that revised and reconciled models are the same model

Table 7 Fit statistics for SCARED-C longitudinal measurement invariance models

	χ^2 (df)	CFI	RMSEA [90% CI]	Δ CFI	Δ RMSEA
Panic/Somatic Subscale					
Configural Invariance	345.70 (285)	1.000	0.023 [0.012, 0.031]		
Metric Invariance	352.30 (297)	0.999	0.021 [0.010, 0.029]	-0.001	-0.002
Scalar Invariance	385.62 (322)	0.998	0.022 [0.012, 0.030]	-0.001	0.001
General Anxiety Subscale					
Configural Invariance	197.55 (121)	0.999	0.039 [0.029, 0.048]		
Metric Invariance	185.01 (129)	1.000	0.032 [0.021, 0.042]	0.001	-0.007
Scalar Invariance	1305.07 (146)	0.930	0.137 [0.130, 0.144]	-0.070	0.105
Partial Scalar Invariance	1054.67 (139)	0.945	0.125 [0.118, 0.132]	-0.055	0.093
Separation Anxiety Subscale					
Configural Invariance	173.34 (91)	0.995	0.046 [0.035, 0.056]		
Metric Invariance	177.05 (98)	0.994	0.043 [0.033, 0.053]	-0.001	-0.003
Scalar Invariance	214.77 (113)	0.992	0.046 [0.036, 0.055]	-0.002	0.003
Social Phobia Subscale					
Configural Invariance	116.12 (61)	1.000	0.046 [0.033, 0.059]		
Metric Invariance	123.44 (67)	0.999	0.044 [0.032, 0.057]	-0.001	-0.002
Scalar Invariance	151.28 (80)	0.999	0.046 [0.047, 0.067]	0	0.002

Note. SCARED-C=Screen for Child Anxiety Related Disorders – Child-report; CFI=comparative fit index; RMSEA=root mean square error of approximation; All χ^2 values were significant at $p < .05$

SCARED-P Subscales

Tests of Unidimensionality

Table 8 Fit statistics for SCARED-P baseline and posttreatment models

	χ^2 (df)	CFI	RMSEA [90% CI]
Panic/Somatic Subscale			
Baseline Model: Initial	176.24 (65)	0.990	0.061 [0.050, 0.072]
Posttreatment Model: Initial	145.07 (65)	0.985	0.054 [0.042, 0.065]
General Anxiety Subscale			
Baseline Model: Initial	205.59 (27)	0.971	0.119 [0.104, 0.134]
Baseline Model: Revised ^a	79.13 (23)	0.993	0.072 [0.055, 0.090]
Baseline Model: Reconciled ^a	79.13 (23)	0.993	0.072 [0.055, 0.090]
Posttreatment Model: Initial	112.96 (27)	0.995	0.086 [0.070, 0.103]
Posttreatment Model: Revised	86.54 (26)	0.997	0.074 [0.057, 0.091]
Posttreatment Model: Reconciled	49.69 (23)	1.000	0.052 [0.032, 0.072]
Separation Anxiety Subscale			
Baseline Model: Initial	455.65 (20)	0.930	0.213 [0.196, 0.230]
Baseline Model: Revised ^a	52.61 (16)	0.997	0.069 [0.049, 0.090]
Baseline Model: Reconciled ^a	52.61 (16)	0.997	0.069 [0.049, 0.090]
Posttreatment Model: Initial	168.36 (20)	0.963	0.131 [0.114, 0.150]
Posttreatment Model: Revised	48.98 (17)	0.998	0.066 [0.045, 0.088]
Posttreatment Model: Reconciled	28.55 (16)	1.000	0.043 [0.014, 0.068]
Social Phobia Subscale			
Baseline Model: Initial	193.39 (14)	0.992	0.164 [0.144, 0.185]
Baseline Model: Revised	44.43 (11)	0.999	0.080 [0.056, 0.105]
Baseline Model: Reconciled	12.48 (9)	1.000	0.028 [0.000, 0.063]
Posttreatment Model: Initial	139.76 (14)	0.998	0.145 [0.123, 0.167]
Posttreatment Model: Revised	21.91 (9)	1.000	0.058 [0.027, 0.089]
Posttreatment Model: Recon- ciled ^a	21.91 (9)	1.000	0.058 [0.027, 0.089]

Note. Initially examined models without residual covariances are labeled “Initial”, when needed, subsequent models with included residual covariances are labeled “Revised”, and final models with equivalent residual covariances at baseline and at posttreatment are labeled “Reconciled”; SCARED-P=Screen for Child Anxiety Related Disorders – Parent-report; CFI=comparative fit index; RMSEA=root mean square error of approximation; All χ^2 values were significant at $p < .05$ except for the reconciled baseline social phobia subscale model

^a Denotes that revised and reconciled models are the same model

For the panic/somatic subscale, initial baseline and post-treatment models were an acceptable fit to the data. For the remaining models, all initial models were a poor fit to the data (see Table 8). Revised general anxiety subscale models, including four residual covariance paths in the baseline model and one residual covariance path in the posttreatment model, were an acceptable fit to the data. Similarly, revised separation anxiety subscale models, including four residual covariance paths in the baseline model and three residual covariance paths in the posttreatment model, were an acceptable fit to the data. For the social phobia subscale, revised models, including three residual covariance paths in the baseline model and five residual covariance paths in the posttreatment model, were an acceptable fit to the data. Model fit for all final reconciled models for the above subscales with equivalent residual covariance paths can be found in Table 8.

Tests of LMI

The configural invariance model for all subscales had good fit (see Table 9 for all fit statistics). For the panic/somatic subscale and the separation anxiety subscale, the metric invariance model and, subsequently, the scalar invariance model each had excellent fit and changes were within acceptable limits. Good fit was also found for the metric invariance model for the general anxiety subscale, however, the scalar invariance model for the general anxiety subscale demonstrated poor fit. Attempts to free equality constraints on thresholds did not yield a change in fit. Finally, good fit was found for the metric invariance model for the social phobia subscale, however, changes in the RMSEA were in excess of acceptable limits. Partial metric invariance was attained after freeing the factor loading for item 39. The resulting partial scalar invariance model demonstrated adequate fit. Thus, it is possible to conclude that the SCARED-P panic/somatic and separation anxiety subscales have scalar invariance, the SCARED-P social phobia subscale has partial scalar invariance, and the SCARED-P general anxiety subscale has metric invariance.

For the panic/somatic subscale, small effect size differences (i.e., $d_{MACS} < 0.4$) were found for 15.4% of items, moderate effect size differences (i.e., $0.5 < d_{MACS} < 0.7$) were found for 53.8% of items, and large effect size differences (i.e., $d_{MACS} > 0.8$) were found for 15.4% of items. For the general anxiety subscale, moderate effect size differences were found for 11.1% of items, large effect size differences were found for 88.9% of items, and no small effect size differences were found. For the separation anxiety subscale, moderate effect size differences were found for 12.5% of items, large effect size differences were found for 50.0% of items, and no small effect size differences were found. For

Table 9 Fit statistics for SCARED-P longitudinal measurement invariance models

	χ^2 (df)	CFI	RMSEA [90% CI]	Δ CFI	Δ RMSEA
Panic/Somatic Subscale					
Configural Invariance	412.92 (285)	0.990	0.033 [0.026, 0.040]		
Metric Invariance	421.33 (297)	0.985	0.032 [0.025, 0.039]	-0.005	-0.001
Scalar Invariance	474.86 (322)	0.980	0.034 [0.027, 0.040]	-0.005	0.002
General Anxiety Subscale					
Configural Invariance	192.04 (117)	0.998	0.039 [0.029, 0.049]		
Metric Invariance	212.13 (125)	0.996	0.041 [0.031, 0.050]	-0.002	0.002
Scalar Invariance	1585.21 (142)	0.872	0.156 [0.149, 0.163]	-0.124	0.115
Partial Scalar Invariance	1334.91 (136)	0.894	0.146 [0.139, 0.153]	-0.102	0.105
Separation Anxiety Subscale					
Configural Invariance	218.37 (87)	0.993	0.060 [0.050, 0.070]		
Metric Invariance	213.93 (94)	0.991	0.055 [0.045, 0.065]	-0.002	-0.005
Scalar Invariance	262.46 (109)	0.988	0.058 [0.049, 0.067]	-0.003	0.003
Social Phobia Subscale					
Configural Invariance	141.05 (59)	1.000	0.057 [0.045, 0.070]		
Metric Invariance	200.04 (65)	0.998	0.070 [0.059, 0.081]	-0.002	0.013
Partial Metric Invariance	176.64 (64)	0.999	0.065 [0.053, 0.076]	-0.001	0.008
Partial Scalar Invariance	233.58 (77)	0.998	0.069 [0.059, 0.080]	-0.001	0.004

Note. SCARED-P=Screen for Child Anxiety Related Disorders – Parent-report; CFI=comparative fit index; RMSEA=root mean square error of approximation; All χ^2 values were significant at $p < .05$

the social phobia subscale, moderate effect size differences were found for 25.0% of items, large effect size differences were found for 50.0% of items, and no small effect size differences were found.

Sensitivity Analysis

As covariances were added to nearly all baseline and post-treatment models, intraclass correlations were calculated for all models comparing the factor loadings, intercepts/thresholds, and residual variances between models with and without covariances. All ICCS were greater than 0.955, indicating excellent agreement between parameters in models with and without covariances. Specific ICCs and model fit statistics for the models without added residual covariances can be found in Supplemental Materials.

Invariance Across Treatment Condition

As the present sample consists of multiple treatment conditions, we explored invariance across treatment condition to ensure that treatment condition did not confound LMI conclusions. Results support measurement invariance across treatment condition at baseline and posttreatment. Fit statistics for scalar invariance models for all measures at each timepoint can be found in Supplemental Materials.

Discussion

The present study examined longitudinal measurement invariance of five measures of anxiety (i.e., PARS, MASC and SCARED parent- and child-reports). Models were assessed with increasing levels of invariance and results present a mixed picture. Scalar invariance, which indicates that valid mean levels comparisons can be conducted [34], was found for the PARS total score and many, but not all, MASC and SCARED subscales (total score models for both the MASC and SCARED were a poor fit to the data and LMI would have had limited validity). Thus, conclusions from prior studies using the PARS are not contaminated by changes in measurement. Most MASC and SCARED subscales are similarly acceptable (e.g., MASC separation anxiety subscale, SCARED panic/somatic and separation anxiety subscales), however, caution is advised for conclusions drawn from longitudinal analyses based on the MASC social anxiety subscale and the SCARED general anxiety subscale. Likewise, caution is advised for longitudinal analysis on the MASC and SCARED total scores until it can be determined whether the total scores are invariant over time.

Results for the SCARED differ slightly from those found in a previous examination. That study found scalar

invariance only in the parent-report general anxiety subscale (we only found metric invariance), partial scalar invariance in the child-report panic/somatic and social anxiety subscales (we found full scalar invariance), and partial metric invariance for the parent-report separation (we found full scalar invariance) [4]. Similar results of metric invariance for the child-report general anxiety subscale and partial scalar invariance in the parent-report social anxiety subscale were found in both studies. The previous report elected to not examine LMI when initial fit at one timepoint was poor (e.g., for the child-report separation anxiety subscale). Had the same approach been used in the present study, LMI would only have been tested for both reports of the panic/somatic subscale as all remaining subscales required the addition of residual covariances due to poor fit. As was concluded in the previous study and replicated here, changes in SCARED scores over time may reflect changes in measurement properties rather than solely changes due to an intervention [4].

For the MASC, all subscales other than the social anxiety subscale showed full or partial scalar invariance. For the child report of the MASC social anxiety subscale, metric invariance was found which indicates equality of factor loadings but not of thresholds. Given this level of invariance, tests of relative standing (e.g., correlations, regression) for these constructs would be valid. Unfortunately, tests of mean-level changes would not be valid based on the lack of support for scalar invariance. However, for the parent-report of the MASC social anxiety subscale, configural invariance was not supported. This indicates that the factor structure of the parent-report MASC social anxiety subscale at baseline and posttreatment are not equivalent and leads to challenges in many modeling contexts.

It is notable that the PARS, which had scalar invariance, is an Independent Evaluator-report based on interviewing both youth and parents, while the MASC and SCARED, which did not consistently demonstrate scalar invariance, are child or parent-report measures. Research on measurement invariance by informant in youth has largely been conducted comparing child- and parent-reports (e.g., Olino and colleagues) [4], and no studies were found that included data from a therapist- or Independent Evaluator-report group. This situation may be a byproduct of the dearth of measures that have both a therapist-report and either a child- or parent-report. Furthermore, this practice is only available in randomized controlled trials or stringent research settings. In the present study, the IEs may have had the strongest basis for evaluating the criterion items, so the support for measurement invariance may reflect that with training, measures may be more stably assessed. Nevertheless, further examination of the present study's discrepant findings are warranted, particularly in the context of the strong body

of literature on informant differences and the benefits of a multi-informant approach [35].

Failure to find unidimensionality in the majority of models merits discussion. Although it is not uncommon to include residual covariances in LMI models of psychological constructs [36], it is noteworthy both the amount of baseline and posttreatment models that required their inclusion (i.e., 27 out of 38 models) and the amount of residual covariances that were required to be added in certain models (i.e., up to six) to reach adequate fit. Conversely, there was excellent agreement between models with and without residual covariances, indicating that the addition of residual covariances did not substantively impact findings. Failure to find unidimensionality has also been found in measures of depression [37]. A lack of unidimensionality indicates that the scale/subscale totals comprise multiple factors rather than one factor and, therefore, may be indicative of multiple constructs rather than the one intended construct. Concerns differentiating between diagnostic criteria for certain anxiety and related disorders in the DSM-5 have been raised [38] and may contribute to the lack of unidimensionality. In the present study, 78.5% of participants met criteria for at least two of the target anxiety diagnoses (i.e., separation anxiety, social anxiety, and generalized anxiety disorders) and 35.9% met for all three diagnoses [18]. Measures of anxiety likely reflect this overlap and contain items that fit criteria for or represent symptoms of multiple disorders rather than a single disorder. For example, the MASC item "The idea of going to camp scares me" that loads onto the separation anxiety subscale could also comprise an element of social anxiety if the fear is related to evaluation at camp rather than (or in addition to) a fear of being away from a loved one. Furthermore, unidimensionality is not required to attain a high value of Cronbach's alpha [39], a commonly used measure of reliability (i.e., internal consistency) in validation studies. Thus, existing measures with high values of Cronbach's alpha and believed to be unidimensional, may not be unidimensional. As measures are developed and assessed, an added focus on unidimensionality rather than simply internal consistency is warranted to ensure that measures do not simply contain items that relate to one another but actually represent the same construct. Future research should also examine this in existing measures.

A further examination of the included residual covariances revealed a potential theme across measures: redundant items. For example, for the MASC social anxiety subscale, multiple residual covariances were added between the item "I feel shy" and other items that are indicative of shyness (i.e., nervous about performing, difficulty asking others to play, worrying about being called on in class). For the SCARED, the generalized anxiety subscale included a residual covariance between items "I am nervous" and "I

am a worrier” which to youth may be synonyms. Likewise, the SCARED social phobia subscale includes items “I don’t like to be with people I don’t know well” and “I feel nervous with people I don’t know well” that may address the same content. Clinicians and researchers should be mindful of this apparent redundancy when selecting measures to assess anxiety in youth. The present findings should also be viewed in light of a recent content analysis of youth anxiety measures which found low overlap such that only 23% of the 42 symptom categories were included in four or more of the seven examined measures (which included the MASC and the SCARED) [40]. As new measures are created and others are refined and updated, it is worth considering this redundancy as an avenue to address the lack of overlap and an effort to more broadly capture differing anxiety presentations.

As the present sample is derived from a treatment outcome study that examined multiple treatments, a possible explanation for these findings (as well as a potential study limitation) is that one of the treatments may have differentially impacted the measurement of anxiety. For instance, it is possible that psychoeducation, an element of the two psychological treatment conditions, may alter ones’ perception and understanding of anxiety differently than the non-therapy treatment conditions. Another consideration was that perhaps medication side-effects not present at baseline impacted the model at posttreatment. Though minimally present overall, significant differences in some adverse events between the medication and CBT conditions were found (e.g., fatigue, restlessness, insomnia) [19]. It is also worth noting that the PARS has an item on physical symptoms and the MASC and SCARED have physical and somatic symptom subscales, respectively. Furthermore, it is possible that treatment drop out may have heightened these effects. Although little drop out occurred, participants in the cognitive-behavioral therapy condition (i.e., Coping cat; 4.3%) were significantly less likely to drop out from treatment than those in the medication (i.e., sertraline; 17.3%) or placebo (19.7%) groups (the combination group, 9.3%, did not significantly differ from other groups with respect to drop out) [19]. If it is believed that medication side effects impacted the endorsement of items, differential participant drop out by treatment condition may exacerbate this effect. Future research is required to examine this further, as well as potential invariance between treatment conditions in general.

The study findings should also be considered in light of limitations. First, as with all examinations of LMI, findings may be exclusive to the characteristics of the present sample (e.g., age range, time interval). Nevertheless, it is important to not take for granted measurement invariance found in the present sample. Next, as noted above, to pursue

measurement invariance models, the majority of models had to be modified with the inclusion of covariances. Although statistically supported, it is possible that the added covariances may not be appropriate or may be providing a “crutch” to the model. Likewise, covariances were added following examination of modification indices, which is not theory driven. Results should be interpreted within this context and future research should replicate these findings utilizing a theory driven approach. Additionally, MASC and SCARED analyses used WLSMV estimation which relies on pairwise data and thus is less adept at handling missing data than MLR or imputation. Finally, the CAMS sample was predominantly white, potentially limiting the generalizability of these findings to more diverse samples. It is possible that when replicated in more diverse samples, conclusions on longitudinal measurement invariance may differ from the present study’s findings, particularly if the measures are not invariant across demographic characteristics. Future studies should assess longitudinal measurement invariance in more diverse samples as well as assess measurement invariance across racial and other factors.

Summary

The present findings, combined with the existing literature, illuminate a complicated picture of whether anxiety measures consistently assess the same construct across treatment. Greater attention to longitudinal measurement invariance is needed when measures are designed and initially validated as researchers need confidence that changes over time are due to treatment effects and not due to changes in the measurement properties. Presently, clinicians and researchers utilizing the MASC or the SCARED to monitor changes in anxiety symptoms may want to use an alternative measure, such as the RCADS, which, in addition to demonstrating measurement of the same construct over time [2], has also been recommended as a consensus measure of youth anxiety [41].

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10578-023-01515-y>.

Author Contribution J.R. and T.O. conducted statistical analyses. J.R. wrote the initial manuscript text with assistance from T.O. and P.K. All authors reviewed the manuscript.

Funding Funding for this study was provided from the US National Institutes of Health awarded to Dr. Olino [R01MH107495].

Data Availability The data reported in this manuscript are publicly available (Registry identification number: NCT00052078. Registry URL: <https://clinicaltrials.gov/ct2/show/NCT00052078>).

Declarations

Ethical Approval All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Competing Interest Mr. Rabner, Dr. Olino, and Dr. Gosch report no potential competing interests. Dr. Kendall has received support from NIMH and NICHD. He has received royalties from the sales of materials related to the treatment of anxiety disorders in youth (e.g., Guilford Press; Workbook Publishing; Gyldendal Norsk; Gyldendal Akademisk). Dr. Albano has received royalties from Oxford University Press for the Anxiety Disorders Interview Schedule, Child and Parent Versions. She has received an Editor's Honorarium from the American Psychological Association. Dr. Ginsburg has received support from NIMH and US Department of Education/Institute of Education Sciences. Dr. Compton has received research support from NIMH, NC GlaxoSmithKline Foundation, Pfizer, and Mursion, Inc. He has served as a consultant for Shire and Mursion, Inc. He has received honoraria from the Nordic Long-Term OCD Treatment Study Research Group and the Centre for Child and Adolescent Mental Health, Eastern and Southern Norway. He has served on the scientific advisory board of Tourette Association of America and Mursion, Inc. He has presented expert testimony for Duke University. Dr. Piacentini has received grant or research support from NIMH, the TLC Foundation for Body-Focused Repetitive Behaviors, the Tourette Association of America, the Pettit Family Foundation, and Pfizer Pharmaceuticals through the Duke University Clinical Research Institute Network. He is a co-author of the Child OCD Impact Scale-Revised (COIS-R), the Child Anxiety Impact Scale-Revised (CAIS-R), the Parent Tic Questionnaire (PTQ), and the Premonitory Urge for Tics Scale (PUTS) assessment tools, all of which are in the public domain therefore no royalties are received. He has received royalties from Guilford Press and Oxford University Press. He has served on the speakers' bureau of the Tourette Association of America, the International Obsessive-Compulsive Disorder Foundation, and the TLC Foundation for Body-Focused Repetitive Behaviors. Dr. Sakolsky has received research support from NIMH. She has received an honorarium from Northwell Health for a child & adolescent lecture at Zucker Hillside Hospital in 2018. Dr. Birmaher has received research support from NIMH. He has or will receive royalties from Random House, Inc., Lippincott Williams and Wilkins, and UpToDate.

References

- Widaman KF, Ferrer E, Conger RD (2010) Factorial Invariance within Longitudinal Structural equation models: measuring the same Construct Across Time. *Child Dev Perspect* 4:10–18. <https://doi.org/10.1111/j.1750-8606.2009.00110.x>
- Mathyssek CM, Olino TM, Hartman CA et al (2013) Does the revised child anxiety and Depression Scale (RCADS) measure anxiety symptoms consistently across adolescence? The TRAILS study. *Int J Methods Psychiatr Res* 22:27–35. <https://doi.org/10.1002/mpr.1380>
- Nelemans SA, Meeus WHJ, Branje SJT et al (2019) Social anxiety scale for adolescents (SAS-A) short form: Longitudinal Measurement Invariance in Two Community samples of Youth. *Assessment* 26:235–248. <https://doi.org/10.1177/1073191116685808>
- Olino TM, Finsaas M, Dougherty LR, Klein DN (2018) Is parent-child disagreement on child anxiety explained by differences in Measurement Properties? An examination of Measurement Invariance Across Informants and Time. *Front Psychol* 9. <https://doi.org/10.3389/fpsyg.2018.01295>
- RUPP Anxiety Study Group (2002) The Pediatric anxiety rating scale (PARS): Development and Psychometric Properties. *J Am Acad Child Adolesc Psychiatry* 41:1061–1069. <https://doi.org/10.1097/00004583-200209000-00006>
- March JS, Parker JDA, Sullivan K et al (1997) The multidimensional anxiety scale for children (MASC): factor structure, reliability, and Validity. *J Am Acad Child Adolesc Psychiatry* 36:554–565. <https://doi.org/10.1097/00004583-199704000-00019>
- Birmaher B, Khetarpal S, Brent D et al (1997) The screen for child anxiety related Emotional Disorders (SCARED): Scale construction and psychometric characteristics. *J Am Acad Child Adolesc Psychiatry* 36:545–553. <https://doi.org/10.1097/00004583-199704000-00018>
- Birmaher B, Brent DA, Chiappetta L et al (1999) Psychometric properties of the screen for child anxiety related Emotional Disorders (SCARED): a replication study. *J Am Acad Child Adolesc Psychiatry* 38:1230–1236. <https://doi.org/10.1097/00004583-199910000-00011>
- Lebowitz ER, Omer H, Hermes H, Scahill L (2014) Parent training for childhood anxiety Disorders: the SPACE Program. *Cogn Behav Pract* 21:456–469. <https://doi.org/10.1016/j.cbpra.2013.10.004>
- Pettit JW, Bechor M, Rey Y et al (2020) A randomized controlled trial of attention Bias Modification Treatment in Youth with treatment-resistant anxiety Disorders. *J Am Acad Child Adolesc Psychiatry* 59:157–165. <https://doi.org/10.1016/j.jaac.2019.02.018>
- Storch EA, Salloum A, King MA et al (2015) A randomized controlled trial in community mental health centers of computer-assisted cognitive behavioral therapy versus treatment as usual for children with anxiety. *Depress Anxiety* 32:843–852. <https://doi.org/10.1002/da.22399>
- Chiu AW, Langer DA, McLeod BD et al (2013) Effectiveness of modular CBT for child anxiety in Elementary Schools. *Sch Psychol Q* 28:141–153. <https://doi.org/10.1037/spq0000017>
- Hancock KM, Swain J, Hainsworth CJ et al (2018) Acceptance and Commitment Therapy versus Cognitive Behavior Therapy for Children with anxiety: outcomes of a Randomized Controlled Trial. *J Clin Child Adolesc Psychol* 47:296–311. <https://doi.org/10.1080/15374416.2015.1110822>
- Wood JJ, Drahota A, Sze K et al (2009) Cognitive behavioral therapy for anxiety in children with autism spectrum disorders: a randomized, controlled trial. *J Child Psychol Psychiatry* 50:224–234. <https://doi.org/10.1111/j.1469-7610.2008.01948.x>
- Cartwright-Hatton S, McNally D, Field AP et al (2011) A new parenting-based group intervention for young anxious children: results of a Randomized Controlled Trial. *J Am Acad Child Adolesc Psychiatry* 50:242–251. <https://doi.org/10.1016/j.jaac.2010.12.015>
- Chu BC, Crocco ST, Esseling P et al (2016) Transdiagnostic group behavioral activation and exposure therapy for youth anxiety and depression: initial randomized controlled trial. *Behav Res Ther* 76:65–75. <https://doi.org/10.1016/j.brat.2015.11.005>
- Kennedy SM, Bilek EL, Ehrenreich-May J (2019) A Randomized Controlled Pilot Trial of the Unified Protocol for Transdiagnostic Treatment of Emotional Disorders in Children. *Behav Modif* 43:330–360. <https://doi.org/10.1177/0145445517753940>
- Kendall PC, Compton SN, Walkup JT et al (2010) Clinical characteristics of anxiety disordered youth. *J Anxiety Disord* 24:360–365. <https://doi.org/10.1016/j.janxdis.2010.01.009>
- Walkup JT, Albano AM, Piacentini J et al (2008) Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *N Engl J Med* 359:2753–2766. <https://doi.org/10.1056/NEJMoa0804633>
- Baldwin JS, Dadds MR (2007) Reliability and validity of parent and child versions of the multidimensional anxiety scale for children in community samples. *J Am Acad Child Adolesc Psychiatry* 46:252–260. <https://doi.org/10.1097/01.chi.0000246065.93200.a1>

21. Rynn MA, Barber JP, Khalid-Khan S et al (2006) The psychometric properties of the MASC in a pediatric psychiatric sample. *J Anxiety Disord* 20:139–157. <https://doi.org/10.1016/j.janxdis.2005.01.004>
22. Villabø M, Gere M, Torgersen S et al (2012) Diagnostic efficiency of the child and parent versions of the multidimensional anxiety scale for children. *J Clin Child Adolesc Psychol* 41:75–85. <https://doi.org/10.1080/15374416.2012.632350>
23. Wei C, Hoff A, Villabø MA et al (2014) Assessing anxiety in Youth with the multidimensional anxiety scale for children. *J Clin Child Adolesc Psychol* 43:566–578. <https://doi.org/10.1080/15374416.2013.814541>
24. Monga S, Birmaher B, Chiappetta L, et al (2000) Screen for child anxiety-related emotional disorders (SCARED): Convergent and divergent validity. *Depress Anxiety* 12:85–91. [https://doi.org/10.1002/1520-6394\(2000\)12:2<85::AID-DA4>3.0.CO;2-2](https://doi.org/10.1002/1520-6394(2000)12:2<85::AID-DA4>3.0.CO;2-2)
25. Compton SN, Walkup JT, Albano AM et al (2010) Child/Adolescent anxiety Multimodal Study (CAMS): rationale, design, and methods. *Child Adolesc Psychiatry Ment Health* 4:1–15. <https://doi.org/10.1186/1753-2000-4-1>
26. R Core Team (2022) R: A language and environment for statistical computing
27. Rosseel Y (2012) lavaan: an R package for structural equation modeling. *J Stat Softw* 48:1–36
28. Gamer M, Lemon J, Fellows I, Singh P (2019) irr: Various Coefficients of Interrater Reliability and Agreement
29. Schermelleh-Engel K, Moosbrugger H, Müller H (2003) Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. 8:23–74
30. Chen FF (2007) Sensitivity of goodness of fit indexes to lack of Measurement Invariance. *Struct Equ Model Multidiscip J* 14:464–504. <https://doi.org/10.1080/10705510701301834>
31. Nye CD, Drasgow F (2011) Effect size indices for analyses of measurement equivalence: understanding the practical importance of differences between groups. *J Appl Psychol* 96:966–980. <https://doi.org/10.1037/a0022955>
32. Nye CD, Bradburn J, Olenick J et al (2019) How big are my Effects? Examining the magnitude of Effect Sizes in Studies of Measurement Equivalence. *Organ Res Methods* 22:678–709. <https://doi.org/10.1177/1094428118761122>
33. Koo TK, Li MY (2016) A Guideline of selecting and reporting Intraclass correlation coefficients for Reliability Research. *J Chiropr Med* 15:155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
34. Widaman KF, Reise SP (1997) Exploring the measurement invariance of psychological instruments: applications in the substance use domain. *The science of prevention: methodological advances from alcohol and substance abuse research*. American Psychological Association, Washington, DC, US, pp 281–324
35. De Los Reyes A, Augenstein TM, Wang M et al (2015) The validity of the Multi-Informant Approach to assessing child and adolescent Mental Health. *Psychol Bull* 141:858–900. <https://doi.org/10.1037/a0038498>
36. Byrne BM, Shavelson RJ, Muthén B (1989) Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychol Bull* 105:456–466. <https://doi.org/10.1037/0033-2909.105.3.456>
37. Fried EI, van Borkulo CD, Epskamp S et al (2016) Measuring depression over time. Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychol Assess* 28:1354–1367. <https://doi.org/10.1037/pas0000275>
38. Chou T, Cornacchio D, Cooper-Vince CE et al (2015) DSM-5 and the Assessment of Childhood anxiety Disorders: meaningful progress, new problems, or Persistent Diagnostic Quagmires? *Psychopathol Rev* 2:30–51. <https://doi.org/10.5127/pr.036214>
39. Ten Berge JMF, Sočan G (2004) The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika* 69:613–625. <https://doi.org/10.1007/BF02289858>
40. Kook M, Clinger JW, Lee E et al (2022) A content analysis of self-report child anxiety measures. <https://doi.org/10.1007/s10578-022-01455-z>. *Child Psychiatry Hum Dev*
41. Krause K, Chung S, Adewuya A, et al (2021) International consensus on a standard set of outcome measures for child and youth anxiety, depression, obsessive-compulsive disorder, and post-traumatic stress disorder. *Lancet Psychiatry* 8:76–86. [https://doi.org/10.1016/S2215-0366\(20\)30356-4](https://doi.org/10.1016/S2215-0366(20)30356-4)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.