

Psychological Inflexibility in Childhood and Adolescence: Development and Evaluation of the Avoidance and Fusion Questionnaire for Youth

Laurie A. Greco
University of Missouri—St. Louis

Warren Lambert
Vanderbilt University

Ruth A. Baer
University of Kentucky

The authors describe the development and validation of the Avoidance and Fusion Questionnaire for Youth (AFQ-Y), a child-report measure of psychological inflexibility engendered by high levels of cognitive fusion and experiential avoidance. Consistent with the theory underlying acceptance and commitment therapy (ACT), items converged into a 17-item scale (AFQ-Y) and an 8-item short form (AFQ-Y8). A multimethod psychometric approach provides preliminary support for the reliability and validity of the AFQ-Y and AFQ-Y8. In 5 substudies, 3 samples (total $N = 1369$) were used to establish (a) item comprehension ($n = 181$), (b) initial item selection ($n = 513$), (c) final item reduction and development of a short form for research ($n = 346$), (d) comprehensive psychometric evaluation of the AFQ-Y and AFQ-Y8 ($n = 329$), and (e) convergent and construct validity for both versions of the AFQ-Y. Overall, results suggest that the AFQ-Y and AFQ-Y8 may be useful child-report measures of core ACT processes.

Keywords: acceptance and commitment therapy (ACT), experiential avoidance, cognitive fusion, children and adolescents

Supplemental materials: <http://dx.doi.org/10.1037/1040.3590.20.2.93.sup>

The mental health field is witnessing rapid scientific advancement of a new generation of behavior and cognitive therapies, such as acceptance and commitment therapy (ACT; Hayes, Strosahl, & Wilson, 1999), dialectical behavior therapy (DBT; Linehan, 1993), and mindfulness-based cognitive therapy (MBCT; Segal, Williams, & Teasdale, 2001). These approaches differ from standard cognitive behavior therapy (CBT) on philosophical, theoretical, and clinical grounds (see Hayes, 2004), though only a few primary clinical differences are summarized here. Perhaps the most notable difference is in the conceptualization and treatment of private events such as thoughts, feelings, memories, and physical bodily sensations. Rather than targeting and attempting to alter the content, frequency, and/or form of private events directly, as in traditional cognitive therapy models, acceptance-based therapies seek to alter the context and function of internal phenomena so as to diminish their behavioral impact.

ACT in particular is a functional contextual behavior therapy that seeks to promote psychological flexibility, or “the ability to

contact the present moment more fully as a conscious human being and to change or persist in behavior when doing so serves valued ends” (Hayes, Luoma, Bond, Masuda, & Lillis, 2006, p. 7). From an ACT perspective, symptom alleviation is not explicitly sought and, if it occurs at all, is viewed as a welcomed and often transitory by-product of therapy. Central goals of ACT are to enhance psychological flexibility and values-based living, regardless of thoughts or feelings in any particular moment. The efficacy and effectiveness of acceptance-based approaches such as ACT have been documented across a wide range of adult clinical populations (see Hayes et al., 2006). As acceptance-oriented approaches continue to gain momentum and are adapted for youth (see Greco & Hayes, 2008), it is essential to develop child instruments that assess clinically relevant outcomes and processes.

Clinically Relevant Processes in ACT

ACT’s empirically grounded clinical model suggests that human suffering is exacerbated by psychological inflexibility that is produced by two interrelated processes: cognitive fusion and experiential avoidance (or more broadly referred to as experiential control). *Cognitive fusion* refers to entanglement with the content of private events. Rather than noticing the ongoing process of thinking and feeling, fusion involves an attachment to the content of private events and a response to this content as if it were literally true (Luoma & Hayes, 2003). When thoughts and feelings are mistaken as accurate representations of reality instead of experienced as transient internal phenomena, unproductive avoidance and control efforts may ensue. In this way, cognitive fusion gives

Laurie A. Greco, Department of Psychology, University of Missouri—St. Louis; Warren Lambert, Kennedy Center and Peabody College, Vanderbilt University; Ruth A. Baer, Department of Psychology, University of Kentucky.

We thank Amy Murrell, Lisa Coyne, Sarah Leek, and Erin Barnett for their feedback and contributions to this work.

Correspondence concerning this article should be addressed to Laurie A. Greco, Department of Psychology, University of Missouri—St. Louis, 325 Stadler Hall, One University Boulevard, St. Louis, MO 63121. E-mail: grecoL@umsl.edu

rise to experiential avoidance, or the unwillingness to experience certain private events and attempts to avoid, manage, alter, or otherwise control their frequency, form, or situational sensitivity (Hayes & Gifford, 1997).

Experiential avoidance falls on the opposing end of psychological acceptance—the openness or willingness to experience private events fully, as they are, without struggle or defense (Hayes et al., 1999). Experiential avoidance is a ubiquitous process learned early in life and is reinforced by the social-verbal community throughout the lifespan (e.g., Greco, Blackledge, Coyne, & Enreheich, 2005; Greco & Eifert, 2004). Once learned, efforts at cognitive and emotional control appear highly resistant to change, due in large part to culturally sanctioned assumptions that painful thoughts and feelings are somehow bad and should be regulated or controlled at all costs. Human beings are unique in their tendency to persist in ineffective and even clearly harmful types of avoidance behavior due to the immediate, short-term relief these actions produce. In the long-term, however, chronic and inflexible efforts to run from our insides may lead to paradoxical increases in emotional pain and come at significant personal costs.

Despite recent empirical advances in adult populations, little is known about the nature and role of psychological inflexibility in childhood. A major obstacle to research with young people has been the absence of developmentally appropriate instruments to assess processes such as cognitive fusion and experiential avoidance. As an initial step toward filling this conceptual and empirical gap, we developed the Avoidance and Fusion Questionnaire for Youth (AFQ-Y). The AFQ-Y asks respondents to rate how true each item is for them, on a 5 point rating scale (0 = *not at all true*; 4 = *very true*). Items are based on ACT's model of human suffering, and they were generated to represent a theoretically cohesive conceptualization of psychological inflexibility fostered by cognitive fusion (e.g., "My thoughts and feelings mess up my life"; "The bad things I think about myself must be true") and experiential avoidance (e.g., "I push away thoughts and feelings that I don't like"; "I stop doing things that are important to me whenever I feel bad").

A paper and pencil measure will provide a convenient and cost-effective way of assessing ACT processes in population-based studies. Although self-reports of emotion are viewed with caution in very young children, evidence suggests that older children and adolescents are reliable reporters of internal experiences (Loeber, Green, & Lahey, 1990). Children's reports of private events also contain valuable information not available to other informants and may more accurately reflect personal experiences across situations (Rohrbeck, Azar, & Wagner, 1991). Furthermore, many types of experiential avoidance are internal in nature and may be virtually undetectable to outside observers. (Common examples include cognitive and emotional regulation strategies such as cognitive disputation and restructuring as well as relaxation, imagery, and self-talk exercises used to alter or manage the content, frequency, and/or situational sensitivity of aversive private events). For these reasons, development of a self-report measure seemed an appropriate starting point for studying the nature and role of psychological inflexibility in children.

This article has five parts: (a) Study 1 ($n = 181$) describes the development of an initial pool of 50 items with feedback from experts and children; (b) Study 2 ($n = 513$) reduces the initial item pool from 50 items to 25 items through theory-driven exploratory

factor analyses; (c) Study 3 ($n = 346$) includes classical test theory, Rasch modeling, and expert feedback in a learning sample to reduce the number of items on the AFQ-Y (from 25 to 17 items) and to devise a short form for research (AFQ-Y8); (d) Study 4 ($n = 329$) evaluates the final, 17-item AFQ-Y and the AFQ-Y8 in a cross-validation sample with multiple criteria from classical test theory, confirmatory factor analysis, and Rasch modeling; and (e) Study 5 ($N = 513 + 675$) presents convergent and construct validity coefficients for both versions of the AFQ-Y. Evidence of discriminant validity (Campbell & Fiske, 1959; Fiske & Campbell, 1992) is also evaluated by examining the AFQ-Y's distinctness from related constructs such as acceptance, mindfulness, and thought suppression.

Study 1: Item Development

Item development was guided by ACT's empirically based model of human suffering. Items were generated by doctoral and master's level psychologists with advanced training in ACT and child clinical psychology.¹ Item content was modeled after the Acceptance and Action Questionnaire (AAQ; Hayes et al., 2004), the most widely used measure of psychological flexibility for adults. Negatively worded items on the AFQ-Y were thought to reflect psychological inflexibility produced by cognitive fusion and experiential avoidance, and positively worded items were thought to reflect psychological flexibility engendered by acceptance- and mindfulness-based processes.

An initial pool of items was evaluated by four independent raters with expertise in ACT. Feedback regarding item clarity, developmental appropriateness, and theoretical coherence guided modifications to the initial item set. Expert reviewers suggested simplifying instructions further and rephrasing items that may be confusing for youths. Items with overlapping content were replaced with behavioral correlates or common indicators of cognitive fusion and experiential avoidance in childhood (e.g., "I say things to make me sound cool"; "I play videogames or use the internet to take my mind off things"; "I cannot be a good friend when I feel upset").

Versions of this measure were administered to 9 children who were 8 years to 14 years old ($M = 11.34$ years), 4 of whom were participating in outpatient psychotherapy for emotional disorders. Children were asked to put items into their own words and to comment on the measure's comprehensibility. Minor wording changes were made, and items were sent back to reviewers for final approval. The resulting 50 items were administered to 181 sixth-grade through eighth-grade children ($M = 12.69$ years, $SD = 1.98$ years; 56% girls, 44% boys; 87% Caucasian, 11% African American, 3% other). To investigate comprehensibility of items, children were asked to circle any confusing items or words. Items rated as confusing by over 2% of the total sample were reworded or replaced (total = 4 items). Overall, results of this pilot work suggested good comprehension of items.

¹ Items on the AFQ-Y were developed by Laurie A. Greco, Amy Murrell, and Lisa Coyne.

Study 2: Initial Item Reduction

In Study 2, we used exploratory factor analysis to remove items inconsistent with ACT theory, which posits that psychological inflexibility is a single construct resulting from two overlapping processes: cognitive fusion and experiential avoidance.

Method

General Procedures

Participants in Studies 2 through 5 were recruited from four public schools in middle Tennessee, with consent and assent procedures approved by the Vanderbilt University Institutional Review Board. All data collection took place at the child's respective school during school hours. Group administration procedures were used, with children participating separately by class (classroom participation rates ranged from 73% to 94%, $M = 86\%$). For each study, an undergraduate or graduate research assistant administered measures in an empty classroom, lunchroom, or library and was available to answer questions before and after each session.

Participants

Study 2 participants were 513 youth (53% girls, 47% boys) in grades 5–10 (age: $M = 12.43$ years; $SD = 2.14$ years). This sample was 80% Caucasian, 13% African American, 2% Hispanic, 2% Asian American, 1% Native American, and 2% other or unknown ethnicities. Students completed the 50-item version of the AFQ-Y as well as other measures assessing behavioral health outcomes (described in Study 5).

Data Analyses

We conducted an exploratory factor analysis of the initial pool of 50 items, using principal factors analysis with squared multiple correlations as the initial communality estimates. The goal of this analysis was to choose items consistent with ACT theory by checking for off-theory multidimensionality in the pool of item candidates (Floyd & Widaman, 1995). We extracted correlated factors using oblique promax rotation and considered items with rotated loadings $>.40$ to represent each factor.

Results and Discussion

We began by inspecting traditional principal component scree plots (Cattell, 1966). The first unrotated component had an eigenvalue of 9.63 and accounted for 19.26% of the variance. The second and third factors had eigenvalues of 4.31 and 2.58, accounting for 8.61% and 5.16% of the variance, respectively. The content of items on Factor 1 reflected two interrelated processes that characterize psychological inflexibility: cognitive fusion and experiential avoidance. Factor 2 contained the ability to engage in desired behavior when feeling bad; Factor 3 contained the belief that it is acceptable to experience negative thoughts and feelings. Notably, Factors 2 and 3 differed from Factor 1 more in their scoring direction than in their content. For example, "I stop doing things that are important to me when my stomach hurts" loaded on Factor 1, whereas, "I do things that are important to me even when

I have a headache or stomachache" loaded on Factor 2. Similarly, "Feeling scared or sad is bad" loaded on Factor 1, whereas "It's OK for me to feel scared or sad" loaded on Factor 3.

Although we first expected items with related content to load onto a single factor, it is common for positive and negative items to load onto separate factors (Enos, 2001; Finney, 2001). Marsh (1996) has shown that inclusion of both positively and negatively worded items in self-report instruments can lead to the appearance of distinct factors that are artifacts of scoring direction. Marsh also presents evidence that this artifact is more pronounced in children and adolescents. The content of Factor 1 represents core clinical processes in ACT; as such, we dropped items on Factors 2 and 3 and proceeded to develop Factor 1 as a single-factor measure. This result is consistent with ACT theory, which suggests that the processes underlying psychological inflexibility are functionally related such that excessive cognitive entanglement (fusion) leads to unworkable experiential control and concomitant behavioral ineffectiveness. Table 1 presents Factor 1 loadings after removal of Factor 2 and Factor 3 items.

Study 3: Additional Item Reduction and Development of the AFQ-Y8

In Study 3, we used a multimethod statistical approach to achieve two main goals: (a) shorten the AFQ-Y by eliminating items with questionable psychometric properties and overlapping item content and (b) develop a short form of the AFQ-Y as an efficient screening tool for group-based research.

Method

Participants and Procedures

Study 3 and Study 4 participants were 675 children. Prior to analyses, the full sample was randomly divided to form a learning sample (Study 3; $n = 346$) and a cross-validation sample (Study 4; $n = 329$). Participants in the learning sample were 346 youths (60% girls, 40% boys) who had a mean age of 12.63 years ($SD = 1.74$ years). The sample was 83.6% Caucasian, 10.5% African American, 2% Hispanic, 0.9% Native American, 0.9% Asian American, and 2% other or unknown ethnicities. Study 3 participants completed the 25-item AFQ-Y and measures assessing behavioral health outcomes (described in Study 5).

Preliminary Analyses

Prior to analyzing data in the learning sample (Study 3) and the cross-validation sample (Study 4), we assessed missing data. For the full sample ($N = 675$), the 25-item AFQ-Ys were, on average, 99% complete, and 82% of the protocols had no missing values at all. The worst case individual AFQ-Y with the most missing data was 80% complete. To simplify later analyses, we used single imputation with the expectation maximization (EM; Little, R. J. A., & Rubin, 1987) algorithm to fill in the 1% missing item scores based on each child's nonmissing responses. The average item mean and standard deviation were nearly identical before imputation ($M = 1.21$; $SD = 0.70$) and after imputation ($M = 1.22$; $SD = 0.70$).

Table 1
Factor Loadings For The 25-Item Version of The Avoidance and Fusion Questionnaire For Youth

Items	Loading
Feeling scared or sad is bad.	.51
My life won't be good until I feel happy. ^{a, b}	.51
I need to keep very busy.	.52
My thoughts and feelings mess up my life. ^{a, b}	.47
I must feel good before doing important things in my life.	.53
If I feel sad or afraid, something must be wrong with me. ^a	.61
The bad things I think about myself must be true. ^{a, b}	.57
I don't try out new things if I'm afraid of messing up. ^a	.48
It's not OK to think about bad things that have happened.	.50
I stop doing things that are important to me when my stomach hurts.	.45
I must get rid of my worries and fears so I can have a good life. ^a	.68
I do all I can to make sure I don't look dumb in front of other people. ^a	.58
I try hard to erase hurtful memories from my mind. ^a	.58
I wish I could wave a magic wand to make all my sadness go away. ^a	.59
If my heart beats fast, there must be something wrong with me. ^{a, b}	.54
I push away thoughts and feelings that I don't like. ^a	.59
I stop doing things that are important to me whenever I feel bad. ^{a, b}	.54
I stay away from people and places that make me feel bad or sad.	.52
If I breathe quickly, something must be wrong.	.59
I try hard to stop feeling nervous.	.53
I do worse in school when I have thoughts that make me feel sad. ^{a, b}	.50
I can't be a good friend when I feel upset. ^{a, b}	.48
I can't stand to feel pain or hurt in my body. ^a	.53
I am afraid of my feelings. ^{a, b}	.59
I say things to make me sound cool. ^a	.41

Note. $N = 513$. Results of exploratory factor analysis yielded a one-factor solution accounting for 31.4% of the total variance.

^a Items retained on the final AFQ-Y.

^b Items retained on the AFQ-Y8.

Data Analyses

To select items for the final AFQ-Y, we combined expert feedback and multiple criteria from classical test theory and Rasch (1980) modeling with WINSTEPS (Linacre, 2006b). These models provide complementary descriptions of item effectiveness that can converge to show a weight of evidence for identifying stronger and weaker items on a test.

Classical test theory. According to classical test theory, good items should have high item-total correlations to give the test adequate alpha reliability (Cronbach, 1951; Cronbach & Shavelson, 2004). In the first step of the item selection process, we dropped items with item-total correlations below .40.

Rasch (1980) modeling. Rasch (1980) rating scale modeling was used in conjunction with expert feedback to identify additional items for elimination. Rasch modeling is a one-parameter member of the item response theory family (Embretson, 1996;

Embretson & Reise, 2000) that offers many practical tools for test construction (Bond & Fox, 2001; Linacre & Wright, 2006, 2007). The Rasch model was originally developed for right-wrong test items; however, extensions made by Andrich (1978, 1988) provide the rating scale model we used. This rating scale model treats Likert scale responses as ordered polytomous responses (Wright & Masters, 1982). Items that fit the Rasch rating scale model have acceptable infit and outfit mean squares indices of misfit. The infit mean square measures unexpected responses with close-targeted items, meaning items with trait levels close to the person's level. The outfit mean square, or outlier fit, measures the item's model fit when the item and person levels are markedly different (Linacre & Wright, 1999). Popular criteria favor fit indices that lie between 0.5 and 1.5 (Linacre, 2006a) or 0.7 and 1.3 (Bond & Fox, 2001).

The Rasch (1980) rating scale model was used to evaluate both test items and respondents in order to construct a single latent trait measured on an equal-interval scale (i.e., items that fit a rational measurement model). Rasch modeling also provided information about whether items target the high clinical end of the AFQ-Y. Items were deleted if they had (a) measure scores on the Rasch rating scale model nearly identical to measure scores of other items (i.e., differences < 0.20) and (b) overlapping item content based on expert opinion.

Development of a short form. The final goal of Study 3 was to develop a unidimensional short form to provide an efficient screening tool for group-based research. Items were selected from the final 17-item AFQ-Y with the same EFA method described in Study 2 (principal factors with squared multiple correlations on the diagonal). Items with loadings $> .50$ in the learning sample were retained on the short form.

Results and Discussion

All 25 items were screened on the multiple criteria described above. Four items had item-total correlations below the .40 cutoff and were therefore dropped. Four other items had measure scores on the Rasch (1980) rating scale model that were nearly identical to measure scores of other items (i.e., differences < 0.20 on a scale on which items range from 44 to 54). Items in these pairs had differing content such as, "I say things to make me sound cool" ($\theta = 50.60$), and "feeling scared or sad is bad" ($\theta = 50.63$). Expert raters examined item pairs and deleted items with redundant content. For example, "feeling scared or sad is bad" had overlapping content with "if I feel sad or afraid, something must be wrong with me." This item ("feeling scared or sad is bad") had an additional weakness as the only item that had infit-outfit mean squares outside the desired range of 0.7 to 1.3 (infit mean square = 1.32; outfit mean square = 2.05). Thus, in total, 8 items were deleted: 4 due to low item-total correlation and 4 due to redundancy (nearly identical Rasch measure scores and overlapping item content). This process shortened the AFQ-Y from 25 items to the final 17-item measure.

After identifying and deleting 8 items, EFA was used on the 17-item scale to identify items for a short-form. Eight items with loadings > 0.50 in Factor 1 were retained in the learning sample to form the AFQ-Y8. Items retained on the final version of the

AFQ-Y and the AFQ-Y8 are indicated in Table 1, and both measures are available online.²

Study 4: Psychometric Evaluation of the AFQ-Y and AFQ-Y8

Study 4 used multiple methods and a cross-validation sample to evaluate psychometric properties of the AFQ-Y and AFQ-Y8. Both versions were evaluated with confirmatory factor analysis, classical test theory, and Rasch modeling.

Method

Participants

As noted above, in Study 4, we used a cross-validation sample consisting of 329 randomly chosen youths (64% girls, 36% boys) who had a mean age of 12.70 years ($SD = 1.56$ years). The sample was 78.5% Caucasian, 14.5% African American, 3.4% Hispanic, 1.5% Native American, 0.3% Asian American, and 1.8% other or unknown ethnicities.

Data Analyses

Preliminary analysis. Before confirmatory analysis, we checked for normality. Of the AFQ-Y's 17 items scored on a 0 to 4 scale, 15 had a mode of zero and 2 had a mode of one. Moreover, there was significant positive multivariate kurtosis (Mardia's normalized $z = 33$, $p < .0001$). This marked flooring of items may be a consequence of the use of school samples in which the clinical trait of interest may not be as prevalent as in clinical samples.

Confirmatory factor analysis. We ran confirmatory tests of the factorial validity of the AFQ-Y and AFQ-Y8 to see how well the hypothesized single factor model fit the data. Single-factor measurement models were estimated with MPLUS 4.2, with Satorra and Bentler (2001) scaled estimates to reduce the impact of nonnormality. We used three criteria to evaluate model fit (Muthén & Muthén, 2003, p. 38): (a) statistically significant misfit ($p > .05$), (b) comparative fit index (CFI) $\geq .96$ (Bentler, 1990; Hu & Bentler, 1999), and (c) root mean square error of approximation (RMSEA) $\leq .05$ (Bentler & Wu, 1993; Steiger, 1990, 2000).

The nonnormality of the AFQ-Y items with this school sample suggested item parceling as a possible technical workaround. According to Bandalos (2002), serious departures from normality may cause erroneous rejection of valid CFA models as much as 100% of the time. Using known distributions, Bandalos increased the normality by aggregating items into groups called *item parcels* or *testlets*, whose score was the mean of several items. The central limit theorem suggests that aggregated means will become more normal as more items are added. Bandalos found that item parcels improved the sensitivity of CFA to recognize known models, especially when item parcels included four or more items. Although parceling is a traditional psychometric method (Cattell & Burdsal, 1975), there are arguments both for and against its use (e.g., Bandalos, 2002; Little, Cunningham, Shahar, & Widaman, 2002; Plummer, 2001).

We calculated three sets of item parcels, using two different formulas to make sure results were consistent. All three parceled analyses of the AFQ-Y used mean scores of 4, 4, 4, and 5 items, rounded to the integer. The first method matched items with high

kurtosis to items with low kurtosis in quartiles, each parcel containing a Quartile 1, a Quartile 2, a Quartile 3, and a Quartile 4 item. Two additional aggregations sorted items into four groups by a random number.

Classical test theory. Cronbach's alpha internal consistency and item-total correlations were calculated to see whether the AFQ-Y and AFQ-Y8 were reliable in traditional terms. Cronbach's alpha of .80 may be considered acceptable (Clark & Watson, 1995), particularly for research evaluating groups (Nunnally & Bernstein, 1994). For individual client evaluation in clinical settings, however, a reliability of at least .90 to .95 is the desirable standard (Nunnally & Bernstein, 1994, p. 264).

Rasch modeling. Rasch rating scale modeling was used to evaluate item parameters (infit and outfit) and the reliabilities of the whole test (item reliability and person reliability). Measure scores were scaled to have a mean of 50 ($SD = 10$), units familiar to clinicians.

Results

Dimensionality of the AFQ-Y and AFQ-Y8

We first ran a CFA with the 17 individual items of the AFQ-Y, with results suggesting inadequate fit for a one-factor measurement model in the cross-validation sample. We next conducted a CFA to evaluate the AFQ-Y8 in the cross-validation sample. As shown in Table 2, fit indices for a one-factor model were satisfactory, thus suggesting that the AFQ-Y8 fits the theoretical one-factor model. As mentioned earlier, there were problems with item normality in this school sample (e.g., 15 of 17 items having a mode of zero). Item parcel CFAs appear in Table 2, rows 3–5. Parceled CFAs showed good fit to a single-factor model according to 8 of 9 estimates. We conclude that the 17-item AFQ-Y is arguably unidimensional and the AFQ-Y8 is unidimensional.

Comprehensive Item and Scale Characteristics

Items on the AFQ-Y had adequate internal consistency reliability ($\alpha = .90$), consistent medium to high item-total correlations (.47–.67), consistent medium to high standardized loadings on a one-factor CFA model (.50–.71), and adequate person separation reliability (.88) and item separation reliability (.97). Rasch misfit statistics (infit and outfit mean squares) were good, ranging between 0.70 and 1.30. The item with the worst fit was "My life won't be good until I feel happy," with infit = 1.12 and outfit = 1.30.

The AFQ-Y8 had very similar characteristics: item-total correlations = .48 to .64; CFA standardized loadings = .52 to .74; Rasch misfit statistics = 1.22 to 1.15; and item-separation reliability = .93. With fewer items, however, the AFQ-Y8 had lower internal consistency ($\alpha = .83$) and person separation reliability (.75). This suggests that the AFQ-Y8 may be more appropriate for group-based research than for the clinical evaluation of individuals.

² Both versions of the AFQ-Y are available online at the supplemental materials web address given above and at www.contextualpsychology.org

Table 2
Confirmatory Fit Indices for a One-Factor Model in Cross-Validation Sample

Row	Scale	Misfit ($p > .05$)	CFI ($> .96$)	RMSEA ($< .05$)
1	AFQ-Y (17 items)	<u>.001</u>	<u>.90</u>	<u>.06</u>
2	AFQ-Y8 (short form)	.12	.99	.034
3	Kurtosis parcel	.71	1.00	.00
4	Random Parcel 1	.70	1.00	.00
5	Random Parcel 2	.07	.99	<u>.07</u>

Note. $N = 329$. Cutoffs for acceptable values are shown in parentheses. Underlined values in row 1 and row 5 do not meet criteria used to evaluate model fit. AFQ-Y = Avoidance and Fusion Questionnaire for Youth; AFQ-Y8 = Avoidance and Fusion Questionnaire for Youth eight-item short form.

Item Targeting

A feature of the Rasch model is its ability to compare both people and items on the same latent trait scale. In Figure 1, AFQ-Y Rasch measure scores appear on the x -axis. The upper histogram (gray) shows the distribution of measure scores for the 17 items, and the lower histogram (black) shows the distribution of measure scores for the 329 children in the cross-validation sample. The high placement of the gray histogram suggests that the AFQ-Y items target the upper end of the school distribution. This targeting is consistent with the use of the AFQ-Y, which is to identify clinical problems marked by heightened cognitive fusion and experiential avoidance rather than to identify children who are more psychologically flexible than average. On the left of the black histogram appears a block of low-end outliers who endorsed none or nearly none of the items.

Discussion

Study 4 evaluated the AFQ-Y and the AFQ-Y8 in a cross-validation sample with CFA, classical test theory, and Rasch modeling. In most respects, the weight of evidence from these four measurement models indicates adequate psychometric properties for the AFQ-Y. In a largely positive picture, two limitations were found: The 17-item AFQ-Y has arguable unidimensionality, as the raw scores did not fit the single-factor CFA model. Although item parcel CFAs did generally fit, the technique has arguable validity. The 17-item AFQ-Y, however, is reliable enough for individual patient assessment.

The AFQ-Y8, on the other hand, may have reliability too low for individual patient assessment. However, the AFQ-Y8 is unidimensional and may therefore be preferred for group-based research.

Items on the AFQ-Y appear to target the upper end of the school distribution, suggesting this scale may be most useful in identifying youths with elevated levels of psychological inflexibility. Results also indicated a group of outliers who endorsed none or nearly none of the items. Although not investigated in this study, it is possible that youth scoring at the extreme low end may not truly demonstrate excellent flexibility; instead, they may have avoided the task altogether by answering *not at all true* to most or

all items. Such a response bias may reflect a possible limitation of including only negatively worded items on the AFQ-Y.

Study 5: Normative Data and Validity Testing of the AFQ-Y and AFQ-Y8

In Study 5, we report normative data and validity coefficients for the AFQ-Y and AFQ-Y8. Based on ACT's model of human suffering and evidence from the adult clinical literature, we expected scores to correlate positively with adverse outcomes such as somatic complaints, internalizing symptoms, and problem behavior. In contrast, we expected scores to correlate negatively with favorable outcomes such as quality of life, social skills, and academic competence. We predicted positive associations between the AFQ-Y and more specific types of cognitive avoidance (i.e., thought suppression) and negative associations between the AFQ-Y and processes thought to produce psychological flexibility (i.e., acceptance and mindfulness). Finally, we examined whether the AFQ-Y measures anything unique by examining partial correlations after removing the effects of acceptance, mindfulness, and thought suppression.

Method

Participants and Procedures

Participants were the school-based samples described in Study 2 ($N = 513$) and Studies 3 and 4 ($N = 675$). For ease of reading, we refer to participants in Study 2 as Sample A and participants in Studies 3 and 4 as Sample B. For maximum precision, results use full samples, ignoring the learning-validation division in Sample B. Child and teacher reports were used in both samples.

Measures of Symptoms and Functioning

Multidimensional Anxiety Scale for Children (MASC; March, 1997). Children and adolescents in Sample A completed the MASC, a 36-item instrument that screens for various types of

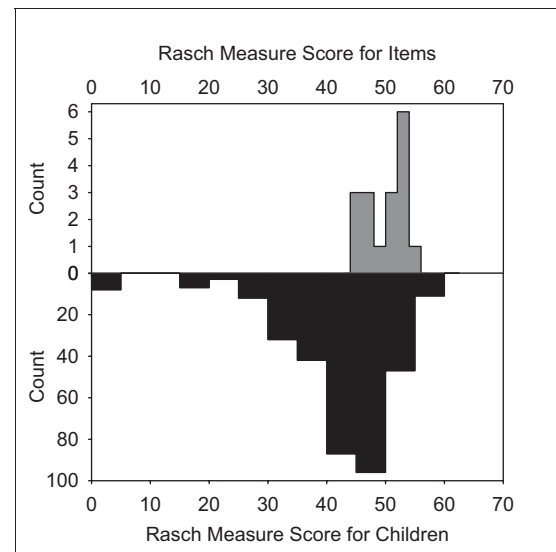


Figure 1. Measure scores for items and for children.

anxiety symptoms experienced by youth. Subscales on the MASC can be combined to yield a total anxiety score that has good concurrent and predictive validity (e.g., March, Parker, Sullivan, Stallings, & Conners, 1997). In this study, total scores were used as an index of child anxiety ($\alpha = .91$; $M = 37.73$, $SD = 18.40$).

Children's Somatization Inventory-Short Form (CSI-SF; Walker & Garber, 2001). Children and adolescents in Samples A and B completed the CSI-SF, an 18-item measure assessing a range of physical-somatic symptoms (e.g., headache, stomachache, nausea, fatigue, dizziness). The CSI-SF has good concurrent and predictive validity and discriminates between children with and children without recurrent pain. Mean scores on the CSI-SF were 18.47 ($SD = 9.24$; $\alpha = .86$) in Sample A and 16.38 ($SD = 12.04$; $\alpha = .89$) in Sample B.

Youth Quality of Life-Revised (YQOL-R; Patrick, Edwards, & Topolski, 2002). Children and adolescents in Samples A and B completed the YQOL-R, a 41-item instrument that assesses life quality and well-being across multiple domains, including self fulfillment, peer relationships, family, and school. The YQOL-R total score has good convergent and discriminant validity (Patrick et al., 2002). Total raw scores on the YQOL-R were used as an overall index of life quality and well-being in Sample A ($M = 289.39$; $SD = 78.94$) and Sample B ($M = 301.45$; $SD = 74.33$). Cronbach's alphas were .95 and .96 in Samples A and B respectively.

Symptoms and Functioning Scale (SFS; Bickman, 2006). Children and adolescents in Sample B completed the SFS, a 33-item child report measure of internalizing symptoms, such as anxiety and depression, and externalizing behavior, such as hyperactivity and conduct problems. The SFS total score had good internal consistency and concurrent validity in a sample of 616 children (Bickman, 2006). Coefficient alphas for Sample B were .89 (internalizing) and .91 (externalizing), with scores ranging from 35 to 150 ($M = 74.29$; $SD = 21.81$).

Social Skills Rating System-Teacher Form (SSRS-TF; Gresham & Elliot, 1990). Teachers completed the SSRS-TF for youths in Samples A and B. The SSRS consists of three subscales: (a) Social Skills, with 30 items assessing cooperation, assertion, and self-control; (b) Problem Behavior, with 18 items assessing hyperactivity, oppositional behavior, and internalizing symptoms; and (c) Academic Competence, with 9 items reflecting children's performance in several academic areas. Research suggests good 2 month test-retest reliability and criterion validity for all three SSRS subscales (Gresham & Elliot, 1990). Respective mean scores and standard deviations for Samples A and B were 44.22 ($SD = 12.27$) and 47.30 ($SD = 12.18$) on the Social Skills scale; 8.24 ($SD = 6.01$) and 7.14 ($SD = 7.23$) on the Problem Behavior scale; and 24.12 ($SD = 7.64$) and 32.94 ($SD = 8.96$) on the Academic Competence scale. In Samples A and B, coefficient alphas on all three subscales ranged from .90 to .97.

Measures Related to ACT Processes

Child Acceptance and Mindfulness Measure (CAMM; Greco & Baer, 2006). Children and adolescents in Sample B completed the CAMM, a 25-item measure of mindfulness that assesses the extent to which youth observe internal experiences (e.g., "I pay close attention to my thoughts"), act with awareness (e.g., "I walk from class to class without noticing what I'm doing"; reverse

scored), and accept internal experiences without judgment (e.g., "I get upset with myself for having certain thoughts"; reverse scored). Research on the CAMM suggests acceptable internal consistency of items and good concurrent validity (Greco, 2005). A total acceptance-mindfulness score can be generated by reverse scoring negatively worded items and summing the item total, yielding a possible range in scores from 0 to 100 (higher scores indicate higher levels of acceptance and mindfulness). In Sample B, scores on the CAMM ranged from 0 to 96 ($M = 47.82$; $SD = 14.95$), with a Cronbach's alpha of .87.

White Bear Suppression Inventory (WBSI; Wegner & Zanakos, 1994). Children and adolescents in Sample B completed the WBSI, a 15-item measure of thought suppression and control (e.g., "There are things that I try not to think about;" "I have thoughts that I cannot stop"). The WBSI has been used to measure thought suppression in school-aged community samples (e.g., Laugesen, Dugas, & Bukowski, 2003) and has been found to have good internal consistency and concurrent validity in both youth and adult samples (Laugesen et al., 2003; Muris, Merckelbach & Horselenberg, 1996). In Sample B, the mean score on the WBSI was 49.96 ($SD = 11.28$), with a Cronbach's alpha of .88.

Data Analyses

In Samples A and B, bivariate correlations were used to investigate relations between the AFQ-Y and clinically relevant measures. Because we conducted multiple significance tests on these correlations, a Hochberg (1998) correction was used for each series of tests to control for familywise false discovery rate. Effect sizes for correlations significant at the adjusted $p < .05$ level were evaluated as small (.10), medium (.30), or large (.50) (Cohen, 1992).

In Sample B, partial correlations were computed between the AFQ-Y and behavioral health outcomes, removing the impact of related processes measured by the CAMM and WBSI. If a measure of psychological inflexibility is to be useful, it must measure something above and beyond the effects of related processes. If the

Table 3
Means and Standard Deviations on the Avoidance and Fusion Questionnaire for Youth (AFQ-Y) and the Eight-Item Short Form (AFQ-Y8)

Demographic	AFQ-Y		AFQ-Y8	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Sex				
Girls ($n = 685$)	23.55	13.15	9.19	6.41
Boys ($n = 503$)	20.51	12.76	7.62	6.20
Race				
Caucasian ($n = 948$)	21.14	12.91	8.19	6.32
African American ($n = 156$)	24.76	14.15	9.45	6.69
Other ($n = 84$)	25.62	14.56	10.03	8.12
Grade				
Grades 5–6 ($n = 515$)	23.88	13.65	8.93	6.86
Grades 7–8 ($n = 446$)	21.47	12.58	8.08	6.17
Grades 9–10 ($n = 227$)	21.04	13.01	8.06	6.08

Note. There were no significant differences between Samples A and B; thus, means and standard deviations are presented for the combined sample ($N = 1,188$).

Table 4

Correlations Between the Avoidance and Fusion Questionnaire For Youth (AFQ-Y and AFQ-Y8) and Behavioral Health Outcomes

Measure	AFQ-Y		AFQ-Y8	
	Sample A	Sample B	Sample A	Sample B
Child report: Symptoms and functioning				
MASC	.58***	—	.56***	—
CSI	.37***	.45***	.42***	.39***
SFS	—	.64***	—	.63***
YQOL-R	-.30***	-.39***	-.29***	-.43***
Child report: Acceptance, mindfulness, cognitive avoidance				
CAMM	—	-.53***	—	-.44***
WBSI	—	.53***	—	.46***
Teacher report: SSRS				
Social skills	-.08	-.13**	-.04	-.18**
Behavior problems	.11*	.11*	.08	.14**
Academic competence	-.19**	-.15**	-.11*	-.17**

Note. For AFQ-Y Sample A, $N = 513$; for AFQ-Y Sample B, $N = 675$; for AFQ-Y8 Sample A, $N = 513$; and for AFQ-Y8 Sample B, $N = 675$. AFQ-Y8 = Avoidance and Fusion Questionnaire for Youth eight-item short form; MASC = Multidimensional Anxiety Scale for Children; CSI = Child Somatization Inventory; SFS = Symptoms and Functioning; YQOL-R = Youth Quality of Life Inventory-Revised; CAMM = Child Acceptance and Mindfulness Measure; WBSI = White Bear Suppression Inventor; SSRS = Social Skills Rating System.

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

AFQ-Y does not make a unique contribution to important child outcomes then it may be necessary to revise either the measure or how we conceptualize the construct (Hayes et al., 2004).

Results and Discussion

Total scores on the AFQ-Y and the AFQ-Y8 were derived by summing responses, yielding total possible scores of 68 and 32, respectively. Means and standard deviations for the combined Samples A and B are summarized in Table 3. Detailed instructions for scoring and a lookup table for percentile ranks and raw-to-Rasch score transformations are available online at the supplemental materials web address given above and at www.contextualpsychology.org

Scores on the AFQ-Y and AFQ-Y8 correlated significantly in the expected directions with measures of symptoms and functioning. As shown in Table 4, both versions of the AFQ-Y correlated positively with child-reported anxiety, somatic complaints, and problem behavior (medium to large effect sizes) and negatively with overall quality of life (medium to large effect sizes). Scores on the AFQ-Y and AFQ-Y8 correlated significantly (though less consistently) with teacher-rated behavior problems, academic competence, and social skills; however, effect sizes for teacher ratings were uniformly small and should therefore be interpreted with caution.

Finally, both versions of the AFQ-Y correlated as expected with conceptually overlapping processes, providing initial support for the construct validity of this measure (medium to large effect sizes). After removing the effects of the CAMM and WBSI,³ both versions of the AFQ-Y correlated significantly in the expected directions with all child measures (mostly medium to large effects) and teacher measures (small effects). Results of partial correlations are presented in Table 5.

In summary, results support the convergent and construct validity of the AFQ-Y and AFQ-Y8. In addition, both versions of the AFQ-Y appear to measure a unique process, as correlations with

children's symptoms and functioning remained significant after controlling for the effects of related constructs (i.e., acceptance, mindfulness, and thought suppression).

General Discussion

Our study is the first to describe the development and validation of a theoretically grounded measure of psychological inflexibility for youth. The AFQ-Y and AFQ-Y8 appear to be psychometrically sound instruments that can be administered quickly and inexpensively. Strengths of this research include the use of cross-validation methods and a theoretically cohesive framework to guide item development. In addition, we used a multimethod statistical approach (e.g., classical test theory, Rasch modeling, and confirmatory analyses) to refine and validate our measure in multiple samples.

Limitations of this work should be considered when interpreting results and planning for future research. The majority of participants were Caucasian youths residing in largely middle class to lower-middle class neighborhoods. Thus, psychometric properties of the AFQ-Y and AFQ-Y8 should be investigated in different geographic regions with youth who have diverse cultural and socioeconomic backgrounds (parenthetically, the AFQ-Y has been translated into several languages and is being evaluated in the Netherlands, Spain, Korea, Sweden, Ireland, Italy, Russia, and Australia). The usefulness of the AFQ-Y and AFQ-Y8 should also be examined in other age groups, particularly among children under 10 years of age. In addition, future research should examine both versions of the AFQ-Y in clinical populations to establish clinical cutoff scores and to explore predictive usefulness in treat-

³ Scores on the Child Acceptance and Mindfulness Measure (CAMM) and the White Bear Suppression Inventory (WBSI) correlated significantly in the expected directions with all child report measures (small to large effects) and with most of the teacher report measures (small effects).

Table 5
Partial Correlations for the Avoidance and Fusion
Questionnaire for Youth (AFQ-Y and AFQ-Y8) After Removing
Variance From the CAMM and WBSI

Measure	AFQ-Y		AFQ-Y8	
	CAMM	WBSI	CAMM	WBSI
CSI	.34***	.33***	.38***	.37***
SFS	.49***	.50***	.52***	.54***
YQOL-R	-.18*	-.22**	-.22**	-.34***
CAMM	—	-.32***	—	-.35***
WBSI	.33***	—	.37***	—
SSRS-SS	-.08	-.14*	-.11*	-.18*
SSRS-PB	.15*	.16*	.15*	.18*
SSRS-AC	-.11*	-.19**	-.08	-.19*

Note. For Sample B, $N = 675$. CSI = Child Somatization Inventory; SFS = Symptoms and Functioning questionnaire; YQOL-R = Youth Quality of Life Inventory-Revised; CAMM = Child Acceptance and Mindfulness Measure; WBSI = White Bear Suppression Inventory; SSRS = Social Skills Rating System; SS = Social Skills; PB = Problem Behaviors; AC = Academic Competence.

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

ment outcome studies. Examination of minimal important difference (Hays & Woolley, 2000) will also be useful, given the measure's intended use for clinical outcome research.

Findings are further limited by our primary reliance on child report. Although teacher informants were included, marginal relations were found between scores on the AFQ-Y and teacher reports in comparison with medium and large effects found for child reports, possibly an indication of shared method variance among child-report measures. The relatively small correlations between the AFQ-Y and teacher reports may also reflect the very nature of cognitive fusion and experiential avoidance—attempts to escape from painful private experiences are often internal and may not be apparent to outside observers. It is possible, for example, to display adaptive behavior outwardly while avoiding painful private experiences covertly with cognitive and emotional control strategies. It will be useful to complement subjective reports with more objective methodologies such as behavioral observations and neurobiological assessment. Such objective measures may be especially useful with very young children who lack the verbal abilities needed to detect, label, and report both the content of internal experiences and reactions to this content. Finally, measures of parent and teacher fusion and experiential avoidance may provide valuable information regarding the social context of ACT-relevant processes.

Development and validation of child measures is an initial step toward expanding the new generation of cognitive and behavior therapies to younger populations. These approaches continue to be adapted for a wide range of child clinical populations (Greco & Hayes, 2008); thus, there is a critical need to develop measures that assess other clinically relevant constructs such as mindfulness, self-compassion, and values-consistent behavior. We hope that the availability of the AFQ-Y and AFQ-Y8 will stimulate work in this area and will serve as an impetus for future research examining the nature and role of psychological inflexibility across child and adolescent populations.

References

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–573.
- Andrich, D. (1988). *Rasch models for measurement*. Newbury Park: Sage.
- Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, 9, 78–102.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246.
- Bentler, P. M., & Wu, E. J. C. (1993). *EQS/Windows user's guide*. Los Angeles: BMDP Statistical Software.
- Bickman, L. (2006). Peabody Treatment Progress Battery manual [Electronic version]. Retrieved September 1, 2006, from <http://peabody.vanderbilt.edu/ptpb/>
- Bond, T. G., & Fox, C. M. (2001). Applying the Rasch model: Fundamental measurement in the human sciences. Mahwah, NJ: Erlbaum.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245–276.
- Cattell, R. B., & Burdsal, C. A. (1975). The radial parcel double factoring design: A solution to the item-vs-parcel controversy. *Multivariate Behavioral Research*, 10, 165–179.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309–319.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64, 391–418.
- Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, 8, 341–349.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Enos, M. M. (2001). *The impact of negation in survey research*. Unpublished doctoral dissertation, University of Chicago.
- Finney, S. J. (2001). *A comparison of the psychometric properties of negatively and positively worded questionnaire items*. Unpublished doctoral dissertation, University of Nebraska, Lincoln.
- Fiske, D. W., & Campbell, D. T. (1992). Citations do not solve problems. *Psychological Bulletin*, 112, 393–395.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7, 286–299.
- Greco, L. A. (2005, November). *Acceptance, mindfulness, and related processes in childhood: Measurement issues, clinical relevance, and future directions*. Paper presented at the Association for Behavior and Cognitive Therapies, Washington, DC.
- Greco, L. A., & Baer, R. A. (2006). *Child Acceptance and Mindfulness Measure (CAMM)*. Unpublished measure, University of Missouri—St. Louis.
- Greco, L. A., Blackledge, J. T., Coyne, L. W., & Ehrenreich, J. (2005). Integrating acceptance and mindfulness into treatments for child and adolescent anxiety disorders: Acceptance and Commitment Therapy (ACT) as an example. In S. M. Orsillo & L. Roemer (Eds.), *Acceptance and mindfulness-based approaches to anxiety: Conceptualization and treatment* (pp. 301–324). New York: Springer.
- Greco, L. A., & Eifert, G. H. (2004). Treating parent-adolescent conflict: Is acceptance the missing link for an integrative family therapy? *Cognitive and Behavioral Practice*, 11, 305–314.
- Greco, L. A. & Hayes, S. C. (Eds.). (2008). *Acceptance and mindfulness*

- treatments for children and adolescents: A practitioner's guide. Oakland, CA: New Harbinger.
- Gresham, F. M., & Elliot, S. N. (1990). *Social Skills Rating System*. Circle Pines, MN: American Guidance Service.
- Hayes, S. C. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavior therapy. *Behavior Therapy*, 35, 639–665.
- Hayes, S. C., & Gifford, E. V. (1997). The trouble with language: Experiential avoidance, rules, and the nature of private events. *Psychological Science*, 8, 170–175.
- Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model processes and outcomes. *Behaviour Research and Therapy*, 44, 1–25.
- Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. New York: Guilford Press.
- Hayes, S. C., Strosahl, K. D., Wilson, K. G., Bissett, R. T., Pistorello, J., Toarmino, D., et al., (2004). Measuring experiential avoidance: A preliminary test of a working model. *The Psychological Record*, 54, 553–578.
- Hays, R. D., & Woolley, J. M. (2000). The concept of clinical meaningful difference in health-related quality of life research. *Pharmacoeconomics*, 18, 419–423.
- Hochberg, Y. (1998). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75, 800–802.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Laugesen, N., Dugas, M. J., & Bukowski, W. M. (2003). Understanding adolescent worry: The application of a cognitive model. *Journal of Abnormal Child Psychology*, 31, 55–64.
- Linacre, J. M. (2006a). *Misfit diagnosis: Infit outfit mean-square standardized*. Retrieved June 1, 2006, from <http://www.winsteps.com/winman/diagnosingmisfit.htm>
- Linacre, J. M. (2006b). Rasch analysis of rank-ordered data. *Journal of Applied Measurement*, 7, 129–139.
- Linacre, J. M., & Wright, B. D. (1999). *WINSTEPS Rasch-model computer program* (Version 2.88) [Computer software]. Chicago: MESA Press.
- Linacre, J. M., & Wright, B. D. (2006). *WINSTEPS Rasch-model computer program* (Version 3.47) [Computer software]. Chicago: MESA Press.
- Linacre, J. M., & Wright, B. D. (2007). *WINSTEPS Rasch-model computer program* (Version 3.47) [Computer software]. Chicago: MESA Press.
- Linehan, M. M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. New York: Guilford.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151–173.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Loeber, R., Green, S. M., & Lahey, B. B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants of childhood pathology. *Journal of Clinical Child Psychology*, 19, 136–143.
- Luoma, J., & Hayes, S. C. (2003). Cognitive defusion. In W. T. Donohue, J. E. Fisher, & S. C. Hayes (Eds.), *Empirically supported techniques for cognitive behavior therapy: A step by step guide for clinicians*. New York: Wiley.
- March, J. S. (1997). *The Multidimensional Anxiety Scale for Children manual*. New York: Multi-health Systems.
- March, J. S., Parker, J. D. A., Sullivan, K., Stallings, P., Conners, C. K. (1997). The Multidimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. *Journal of American Academy of Child and Adolescent Psychiatry*, 36, 554–565.
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifacts? *Journal of Personality and Social Psychology*, 70, 810–819.
- Muris, P., Merckelbach, H., & Horselenberg, R. (1996). Individual differences in thought suppression. The White Bear Suppression Inventory: Factor structure, reliability, validity, and correlates. *Behaviour Research and Therapy*, 34, 501–513.
- Muthén, L. K., & Muthén, B. O. (2003). *Traditional latent variable modeling using Mplus*. Los Angeles: Author.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Patrick, D. L., Edwards, T. C., & Topolski, T. D. (2002). Adolescent quality of life, part II: Initial validation of a new instrument. *Journal of Adolescence*, 25, 287–300.
- Plummer, B. A. (2001). To parcel or not to parcel: The effects of item parceling in confirmatory factor analysis. *Dissertation Abstracts International*, 61(10), 5617B.
- Rasch, G. (1980). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests* (Expanded ed.). Chicago: University of Chicago Press.
- Rohrbeck, C. A., Azar, S. T., & Wagner, P. E. (1991). Child Self-Control Rating Scale: Validation of a child self-report measure. *Journal of Clinical Child Psychology*, 20, 179–183.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507–514.
- Segal, Z. V., Williams, J. M. G., & Teasdale, J. T. (2001). *Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse*. New York: Guilford.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180.
- Steiger, J. H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduck and Glaser. *Structural Equation Modeling*, 7, 149–162.
- Walker, L. S., & Garber, J. (2001). *Manual for the Children's Somatization Inventory*. Nashville, TN: Vanderbilt University, Department of Pediatrics.
- Wegner, D. M., & Zanakos, S. (1994). Chronic thought suppression. *Journal of Personality*, 62, 615–640.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: Mesa Press.

Received January 19, 2006

Revision received January 15, 2008

Accepted February 5, 2008 ■