

Michael, K. D., & Merrell, K. W. (1998). The reliability of children's self-reported internalizing symptoms over brief- to medium- length time intervals. **Journal of the American Academy of Child and Adolescent Psychiatry**, 37:2, pp.194-201. Feb. 1998. (ISSN: 0890-8567) Lippincott, Williams & Wilkins

Keywords: reliability | internalizing symptoms | self-report | children.

## **The reliability of children's self-reported internalizing symptoms over brief- to medium- length time intervals**

**Kurt D. Michael and Kenneth W. Merrell**

### **ABSTRACT**

**Objective:** To determine whether children between the ages of 8 and 12 years are able to reliably report internalizing symptoms over short to medium-length time intervals as measured by an objective self-report instrument of internalizing symptoms. **Method:** The Internalizing Symptoms Scale for Children (ISSC) was group-administered initially to 131 children and at subsequent intervals of 2 weeks, 4 weeks, and 12 weeks. **Results:** Pearson product-moment correlations for the ISSC total scores of the participants were computed across the various retest intervals. At 2 weeks, the correlation was .84. At 4 weeks, the correlation was .76. After 12 weeks, the correlation was .74. **Conclusions:** These data indicate that children between 8 and 12 years old can reliably report their experience over short to medium-length intervals. These findings provide strong support for the ISSC as a research and clinical instrument for the assessment of internalizing symptoms in children between 8 and 12 years of age, which may ultimately prove beneficial in the identification and treatment of childhood internalizing disorders. Limitations and recommendations for future research are discussed.

### **ARTICLE**

Recent attempts to create empirically sound taxonomies of child psychopathology have yielded two broad dimensions of emotional and behavioral problems, namely, internalizing and externalizing disorders (Achenbach, 1985). Internalizing disorders are a constellation of inner-directed or overcontrolled expressions of distress, whereas externalizing disorders are defined as outerdirected or undercontrolled behavioral problems such as

aggression, impulsivity, hyperactivity, and delinquency (Achenbach and McConaughy, 1992). Internalizing disorders encompass a wide variety of problems including depression, anxiety, social withdrawal, and somatic complaints (Reynolds, 1990). These problems have been found to be interrelated clinically and have been shown to be strongly associated in factor-analytic studies (Ollendick and King, 1994). Prevalence rates of internalizing disorders in children vary depending on the particular disorder under investigation and the diagnostic criteria used. However, prevalence estimates for particular childhood internalizing disorders have ranged from 2.0% for depression to 8.9% for anxiety in normal samples (Anderson et al., 1987; Costello, 1989).

Internalizing problems in children may have negative effects on their self-esteem (Merrell, 1994), academic performance (Quay and La Greca, 1986), physical health (Walker and Greene, 1989), social competence (Fischer et al., 1984), and future psychological adjustment (Kovacs, 1985). In light of the negative outcomes associated with internalizing problems, researchers and clinicians have stressed the importance of early identification and treatment of internalizing symptoms in children (Reynolds, 1992). Traditionally, the evaluation of childhood psychological disorders has relied on the verbal or written reports of parents, teachers, and other significant figures in the child's environment. However, because internalizing disorders are, in great measure, subjective perceptions of internal distress, they are often not readily or reliably identified by external observers (Flanery, 1990). Direct behavioral observations and behavioral checklists often yield discrepancies among child, parent, and teacher observations and reports (Achenbach et al., 1987), and outside observers often underestimate the intensity and breadth of a child's emotional experience (Kurdek and Berg, 1987). In addition, external evaluations of a child's internal state are subject to significant observer bias (Edelbrock et al., 1986). As a result, several authors have emphasized the value of eliciting the child's perspective through self report assessment as part of a multimethod, multisource evaluation protocol (La Greca, 1990). Subsequently, several self-report methods have been developed to assess internalizing problems in children, ranging from clinical and structured interviews to objective paper and-

pencil inventories, with the latter method being the specific focus of this study.

Whereas structured or semistructured interviews are often used to determine whether children's self-reported symptoms reach a diagnostic threshold, objective self-report instruments are typically used to assess the degree to which children endorse clinically significant symptomatology relevant to a particular problem area. Kazdin (1988) described this phenomenon as the difference between categorical (diagnostic threshold) and dimensional (degree of symptomatology) assessment practices. Several excellent objective self-report instruments have been designed to assess specific internalizing problems in children. The most prominent of these measures include the Children's Depression Inventory (CDI) (Kovacs, 1992), the Reynolds Child Depression Scale (RCDS) (Reynolds, 1989), the Revised Children's Manifest Anxiety Scale (RCMAS) (Reynolds and Richmond, 1985), and the State-Trait Anxiety Inventory for Children (STAIC) (Spielberger, 1973). In addition, the Youth Self-Report (YSR) (Achenbach, 1991) is designed to measure common internalizing constructs such as depression and anxiety as well as other internalizing subcomponents such as withdrawal and somatic complaints. Overall, these instruments possess adequate to excellent psychometric properties and they have been well researched for use in research and clinical settings.

These instruments tend to use a common response format, in that the children are presented with a series of statements regarding the presence or absence of specific symptoms, and then they rate how true these statements are for them or how often they occur. The authors of most objective self-report assessment devices have established cutoff scores to operationalize clinically relevant levels of symptomatology for their instruments (Reynolds, 1989). A number of metrics have been used to indicate clinical cutoff points for various self-report instruments including raw scores, T scores, and percentile ranks. A response set that is 1.0 to 2.0 standard deviations above the mean on self-report inventories is often considered to be a good indication of clinically relevant self-reported symptomatology, assuming that certain assumptions regarding the sample (e.g., normally distributed, clinical versus nonclinical, random sample)

have been considered (Merrell, 1994). Although the establishment of clinical cutoff scores on self-report measures is not equivalent to a formal diagnosis (Reynolds, 1989), it often provides valuable information about the severity of certain symptomatology in comparison to a particular reference group as long as the base rates for the criterion (e.g., depression, anxiety) are taken into account (Finn and Kamphuis, 1995).

Although objective self-report measures are important in the assessment of internalizing problems in children, these instruments have limitations as well. Some of these limitations include biased response styles (i.e., social desirability, faking good, faking bad) (Borg and Gall, 1989), children's ability to understand and report their emotions (Clarizio, 1984), children's ability to reliably report subjective states of internal distress (Edelbrock et al., 1985), and whether the instrument has an age-appropriate reading level (Prout and Chizik, 1988). Despite these limitations, several authors have stressed the importance of self-report when evaluating internalizing symptoms in children (La Greca, 1990; Saylor et al., 1984). Objective self-report tests in particular appear to have gained widespread acceptance among clinicians and researchers.

For an objective self-report measure to be useful to both clinicians and researchers, it must be valid and reliable. The specific question addressed by this study is whether children are able to reliably report subjective internal states as measured by an objective self-report instrument of internalizing symptoms over time. Test-retest reliability, as applied to psychometric instruments, refers to the stability of the measuring device over time as well as the temporal stability of the underlying construct (Cronbach, 1960). Test-retest reliability coefficients are correlations between initial and subsequent administrations of the same measure to the same sample at different time intervals. Anastasi (1988) suggested that test-retest reliability is an integral part of the test development process and "shows the extent to which scores on a test can be generalized over different occasions" (p. 117).

Whereas there is general agreement among psychologists that scores on measures of intelligence, interest,

and aptitude should be highly stable over time (i.e., retest coefficients in the .80s and .90s), it is much less clear whether tests that measure personality or psychopathology should be held to the same standard (Graham, 1993). It is important to differentiate between the error variance of scores (fluctuations due to chance factors) and true variance, or the actual fluctuations in the construct( ~) being measured. However, this differentiation is often easier said than done. Flanery (1990) suggested that classic psychometric theory is built on the assumption that the constructs being measured are "trait-like" and relatively stable. Thus, a measure with low test-retest reliability is often judged to be a poor test. However, if a particular construct varies naturally over time, an accurate measurement of such an "unstable" construct will reveal relatively low test-retest reliability coefficients (Anastasi, 1988).

Edelbrock and colleagues (1985) argued that there is no absolute test-retest reliability standard for a child's social-emotional self-report instrument. However, on objective self-report measures of internalizing constructs such as the RCDS and the CDI, Reynolds (1989) suggested that moderately high test-retest reliability coefficients (.70s and higher) over a period of several weeks are adequate because many internalizing symptoms (e.g., depression) fluctuate naturally over time. Kovacs (1992) suggested that "one would not expect a depressive syndrome to remain uniformly stable over months . . . thus, for a symptom oriented instrument, a two-week test-retest interval may be the most appropriate" (p. 37). Furthermore, when assessing children, the evaluator must remain cognizant of possible developmental changes (e.g., intellectual, social, perceptual, affective) that could lower stability estimates of internalizing disorders (Flanery, 1990). For example, certain early childhood fears and anxieties tend to abate as the child gets older, which has been conceptualized as a normal developmental process (Campbell, 1986). In light of these considerations, most researchers recommend brief testing intervals between 2 and 4 weeks for social-emotional self-report tests (Anastasi, 1988; Borg and Gall, 1989; Kovacs, 1992).

As previously mentioned, some researchers have

questioned whether young children (i.e., preadolescent) can reliably report subjective internal states and emotions (Costello, 1986). For example, Edelbrock and colleagues (1985) reported that children younger than the age of 10 did not give reliable self-reports (average  $r = .43$ ). Boyle and colleagues (1993) reported similar findings in a group of children between 6 and 11 years old (average  $r$  for externalizing disorders = .32; average  $r$  for internalizing disorders = .06). In both studies, the authors reported consistent age-related increases in the reliability of self-reported symptomatology over 1- to 3-week intervals in children 10 to 12 years of age and older.

While these findings are relevant, it is important to note that both studies incorporated structured interview formats (e.g., Diagnostic Interview Schedule for Children; Costello et al., 1984) as opposed to objective self-report measures such as the CDI or the RCMAS. As mentioned above, most structured interview schedules assess whether a particular diagnosis is either present or absent; thus the reliability coefficients are based on a particular diagnostic threshold and whether the child reports symptomatology consistently (i.e., diagnostic status) across various intervals. In contrast, while many objective self-report measures establish clinical cutoff scores, reliability coefficients are typically calculated by comparing the total scores for the various intervals. As a result, the attenuated reliability coefficients reported for young children using structured interview formats may be attributed to analyses based on discrete variables (diagnostic status) as compared to total score correlations (continuous variables) on objective self-report measures.

In contrast, the ability of younger children (<10) to reliably report their experience over short to medium length intervals on objective self-report measures of internalizing constructs has been established in several studies including the RCDS (Reynolds and Graves, 1989), the CDI (Finch et al., 1987), the RCMAS (Reynolds, 1981; Wisniewski et al., 1987), and the STAIC (Spielberger, 1973). Refer to Table 1 for further clarification.

In view of the need to understand better the ability

of young children to reliably report their own perceptions of social-emotional functioning, as well as the

**TABLE 1**  
Test-Retest Reliability Coefficients for the Various Self-Report Instruments

| Instrument | Sample       | Author                        | Interval  | <i>r</i> |
|------------|--------------|-------------------------------|-----------|----------|
| CDI        | Normal       | Finch et al., 1987            | 2 weeks   | .82      |
|            | Normal       | Finch et al., 1987            | 4 weeks   | .66      |
|            | Referred     | Weiss et al., 1991            | 16 weeks  | .54      |
| RCDS       | Normal       | Breen, 1987                   | 2 weeks   | .85      |
|            | Normal       | Reynolds and Graves, 1989     | 4 weeks   | .82      |
| RCMAS      | Normal       | Wisniewski et al., 1987       | 1-5 weeks | .88-.77  |
|            | Normal       | Reynolds, 1981                | 9 months  | .68      |
| STAIC      | Normal-Trait | Speilberger, 1973             | 6 weeks   | .65-.71  |
|            | Normal-State | Spielberger, 1973             | 6 weeks   | .31-.41  |
| YSR        | Normal       | Achenbach and Edelbrock, 1987 | 1 week    | .81      |
|            | Referred     | Achenbach and Edelbrock, 1987 | 6 months  | .69      |

*Note:* CDI = Children's Depression Inventory; RCDS = Reynolds Child Depression Scale; RCMAS = Revised Children's Manifest Anxiety Scale; STAIC = State-Trait Anxiety Inventory for Children; YSR = Youth Self-Report.

need for continuing research and clarification of the broad-band internalizing construct in children, the present investigation was conducted. The specific purpose of this investigation was to evaluate the stability of children's self-reported internalizing symptomatology using an objective self-report instrument at several short to medium-length time intervals.

## METHOD

### Participants and Procedure

Participants for this study consisted of middle to late elementary school-age children between the ages of 8 and 12 years. The accessible sample consisted of all third-, fourth-, and fifth-grade students ( $N = 199$ ) from an elementary school in an urban area in the intermountain west. The final sample consisted of 131 children (65.82%; 66 boys, 65 girls), and there was no evidence that the participants differed significantly from the nonparticipants. The children were between the ages of 8 and 12, with a mean age of 9.42 years. The students were from grades 3 through 5, with a mean grade of 3.94. The sample was 86% Caucasian, 8% Hispanic, 4% Asian, 1% African-American, and <1% Pacific Islander or Native American. Of this sample, 36% of the students received free or reduced-price lunch based on low family income status.

Parents of each of the 199 potential child participants were sent a

letter describing the study, with an attachment to complete and return to the investigator indicating whether they did or did not give their consent for their child to participate in the study. Of the 199 consent forms sent out for review, 173 (86.93%) were returned to the investigator. The parent(s) of 144 children (72.3696) gave informed consent for their children to participate in the study. Individual child subjects were also given the opportunity to decline participation in the study if they desired, even if their parent(s) had given consent for their participation. Child subjects who did not participate in the study were not penalized in any way, and they were given an alternative activity (e.g., homework, reading, drawing) to work on during administration of the Internalizing Symptoms Scale for children (ISSC).

The ISSC was initially group-administered to 144 children at an elementary school in the intermountain west. The children completed the ISSC on their own after the directions were read verbatim. The ISSC took approximately 10 to 15 minutes to administer. The confidentiality of the child subjects was protected by assigning each participant an identification number. Subsequent to the initial administration, the ISSC was readministered to the same sample of children at intervals of 2 weeks, 4 weeks, and 12 weeks.

During the course of the study, data obtained from 13 children were not included in the analysis because of incomplete ISSC protocols (more than 3 of the 48 items missing), illness, or an absence during any one of the four ISSC administrations. The final sample consisted of 131 children who were present for all of the administrations. Only those children ( $n = 131$ ) who completed the ISSC during all four administrations were included in the statistical analysis procedures. Missing data (i.e., unanswered items on individual protocols 53) were dealt with by incorporating item mean substitutions based on the norms of the national standardization sample.

## **Instrumentation**

The ISSC is a 48-item self-report instrument, designed to assess internalizing symptomatology in children (Merrell and Walters, in press). The rationale for developing the ISSC was based on the apparent dearth of a comprehensive instrument that assesses the broad band of internalizing problems in children. The endorsement format of the ISSC is based on a 4-point Likert scale (values of 0 through 3). Respondents can either disavow or endorse symptoms by checking boxes corresponding to "never true," "rarely true," "sometimes true," and "often true." A total score is obtained for the

ISSC based on a sum of all 48 item scores, with the positively worded items reverse-scored so that greater values always indicate more internalizing distress.

The ISSC normative sample included more than 2,200 cases from each of the four major geographic regions of the United States. The normative sample has been stratified to represent the population, both geographically and ethnically. An internal consistency coefficient of .91 was reported for the ISSC total score (Merrell and Walters, in press).

In a factor-analytic study of the ISSC, a two-factor solution was indicated (Merrell et al., 1997b), using factor loading cutoffs of .30 or higher, and including confirmatory factor analysis. The first factor, Negative Affect/General Distress, contains items that indicate the presence of specific internalizing symptoms or emotional distress. This factor included 35 items. The second factor, Positive Affect, contains items that denote the absence of internalizing symptoms or the presence of positive affect and cognitions incompatible with emotional distress. This factor included 17 items. Four of the items cross-load on both factors. These findings are consistent with the work of several researchers who have suggested that positive and negative expressions of affectivity are independent components that make unique contributions to the etiology and prevention of internalizing disorders such as depression and anxiety (e.g., Clark et al., 1990). Several sample ISSC items are listed by factor in Table 2.

In a recent convergent validity study (Merrell et al., 1997a), the ISSC was compared with three instruments that purport to measure constructs within the internalizing domain including the CDI, the RCMAS, and the Internalizing broad-band score from the YSR. Convergent validity coefficients were obtained by computing the Pearson product-moment correlations between the various instruments. The correlation between the ISSC total score and the CDI total score was .75, indicating that the two instruments measure strongly related, but slightly different constructs. The correlation between the total scores of the ISSC and the RCMAS was .78, also an indication of a moderately strong relationship between the two instruments. The correlation between the ISSC total score and the Internalizing broad-band score on the YSR was .86, which was the strongest relationship found in the study. These results are evidence that the ISSC is a broad-band measure of internalizing symptoms (e.g., depression, anxiety) when compared with other well-researched instruments that purport to measure internalizing constructs. Other validity studies that have been conducted on the ISSC indicate that it is sensitive to theoretically based differences among

various educational and clinical groups (Merrell and Dobmeyer, 1996; Merrell et al., 1996), providing evidence of construct validity.

## RESULTS

Descriptive statistics for the ISSC total and factor scores at each testing interval are presented in Table 3. The mean ISSC total score for the sample of 131 children was 53.65 (SD = 19.51) during the initial administration and were 49.56 (SD = 22.37), 47.83 (SD = 21.88), and 48.07 (SD = 21.20) during the subsequent intervals of 2, 4, and 12 weeks, respectively. Test-retest reliability coefficients were calculated by computing Pearson product-moment correlations

**TABLE 2**  
Sample ISSC Items, Listed by Factor

| Factor 1: Negative Affect/General Distress | Factor 2: Positive Affect         |
|--|-----------------------------------|
| I am shy                                   | I feel cheerful                   |
| I worry about things                       | I feel important                  |
| I have bad dreams                          | I have lots of energy             |
| I worry that I will hurt someone           | I do things as well as other kids |
| I have trouble sleeping                    | I like the way I look             |
| Lots of things scare me                    | I do well in school               |
| When there is a problem, it is my fault    | I feel happy                      |
| It is hard for me to breathe               | I like myself                     |

*Note:* ISSC = Internalizing Symptoms Scale for Children.

between the ISSC scores obtained during- the initial, 2-week, 4-week, and 12-week administrations of the ISSC. ISSC reliability coefficients for 2-, 4-, and 12-week retest intervals were .84, .76, and .74, respectively. The test-retest reliability coefficients for the ISSC total scores are presented in Table 4.

Similarly, test-retest reliability coefficients for the two factor scores on the ISSC were computed on the basis of Pearson product-moment correlations between the

initial and subsequent factor scores on the ISSC. Correlation coefficients for the first factor score (Negative Affect/General Distress) were .81 at 2 weeks, .73 at 4 weeks, and .70 at 12 weeks. Test-retest reliability coefficients for the second factor score (Positive Affect) were .79 at 2 weeks, .79 at 4 weeks, and .72 at 12 weeks. These results are summarized in Table 4.

An analysis of variance with repeated measures was conducted to determine the stability of ISSC scores across the administrations. The results indicated a significant difference among the mean scores across the four intervals ( $F[3,390] = 12.31, p < .001$ ). An examination of the means (Table 3) indicates that there is a general attenuation of symptom endorsement after the initial administration of the ISSC.

**TABLE 3**  
Means and Standard Deviations for the ISSC Total Score and Factor Scores for Each Interval

| Interval | Total Score |       | Factor 1 |       | Factor 2 |      |
|----------|-------------|-------|----------|-------|----------|------|
|          | Mean        | SD    | Mean     | SD    | Mean     | SD   |
| Initial  | 53.65       | 19.51 | 45.86    | 16.09 | 11.27    | 8.14 |
| 2-week   | 49.56       | 22.37 | 42.39    | 18.83 | 10.29    | 8.37 |
| 4-week   | 47.83       | 21.88 | 40.83    | 18.70 | 10.16    | 8.19 |
| 12-week  | 48.07       | 21.20 | 41.32    | 18.51 | 10.00    | 7.53 |

*Note:* ISSC = Internalizing Symptoms Scale for Children.

**TABLE 4**  
Test-Retest Reliability for ISSC Total Score and Factor Scores for Each Interval

| Administration | Total Score | Factor 1 | Factor 2 |
|----------------|-------------|----------|----------|
| Initial        | —           | —        | —        |
| 2 weeks        | .84         | .81      | .79      |
| 4 weeks        | .76         | .73      | .79      |
| 12 weeks       | .74         | .70      | .72      |

*Note:* ISSC = Internalizing Symptoms Scale for Children.

## DISCUSSION

The primary purpose of this investigation was the determination of the degree to which preadolescent children are able to reliably report internalizing symptoms over short to medium-length time intervals. Overall, the results of this investigation indicate that children between the ages of 8 and 12 years are able to give consistent self-reports of subjective internal states as measured by the ISSC. The 2-week test-retest reliability coefficient for the ISSC total score was high (.84). After 4 weeks, the test-retest reliability coefficient for the ISSC total score was moderate (.76) and decreased only slightly after 12 weeks to .74. The attenuated reliability coefficients reported for the longer retest intervals (4, 12 weeks) are consistent with the findings from previous studies, which provide support for the notion that a number of internalizing mood states fluctuate naturally over time. Nonetheless, the magnitude of the reliability coefficients for the ISSC over short to medium-length intervals is strong enough to provide empirical support for the instrument as a reliable measure of internalizing symptoms in children between 8 and 12 years old.

Consistent with the ISSC total score test-retest correlations, the test-retest reliability coefficients for each of the factor scores were moderate to high. Correlations for the ISSC factor 1 (Negative Affect/General Distress) score ranged from .81 at 2 weeks to .70 after 12 weeks. Similarly, test-retest reliability coefficients for the ISSC factor 2 score were moderately stable over time, ranging from .79 at 2 weeks to .72 after 12 weeks.

In general, the test-retest reliability coefficients reported for the ISSC are consistent with, and in some cases superior to, the findings from studies investigating the reliability of other objective self-report instruments of internalizing constructs as described previously. The test-retest reliability of the ISSC was very stable after short intervals and moderately stable after medium length intervals. These data strongly support the ISSC as a psychometrically sound assessment device.

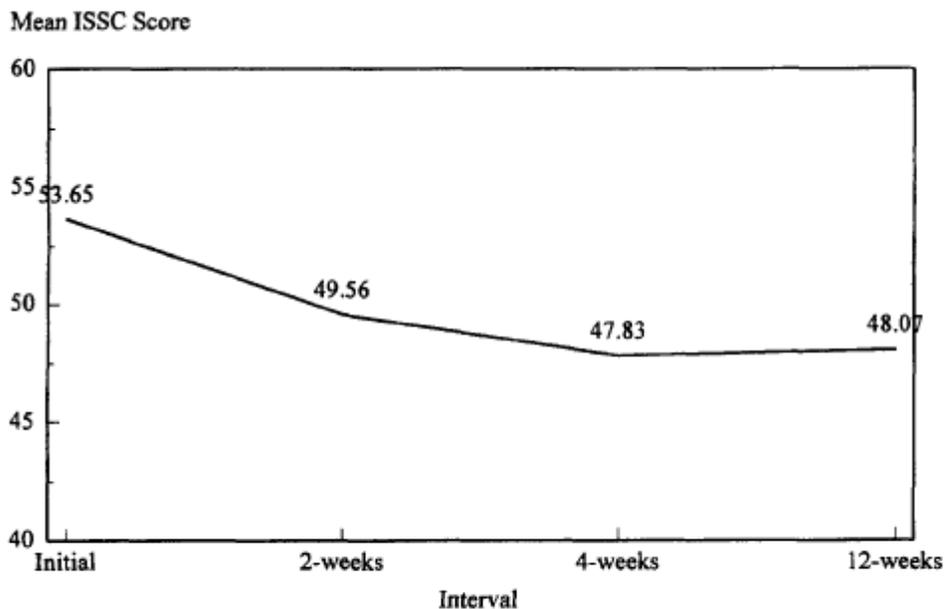
For the ISSC national normative group, the mean total score was 53.95 for the standardization sample of more than 2,200 children. Similarly, the mean ISSC

score for the sample in this study was 53.65 during the initial assessment. However, after each of the three subsequent administrations of the ISSC, the mean scores dropped an average of 5.26 points during time 2 (49.56), time 3 (47.83), and time 4 (48.07).

While a repeated-measures analysis of variance revealed a significant difference among the means for each of the four intervals, a practical analysis of these data indicates that there is general attenuation of symptom endorsement after the initial administration of the ISSC (Fig. 1). This finding is consistent with data reported from other test-retest reliability studies of instruments that purport to measure internalizing constructs (e.g., Finch et al., 1987; Reynolds and Graves, 1989). Possible interpretations of the attenuated mean ISSC scores during subsequent administrations include an over endorsement of internalizing symptomatology by distressed children upon initial testing (Reynolds, 1986), an expected variation in reported symptomatology due to natural fluctuations in mood over time (Kovacs, 1992), and/or a better understanding of the assessment task during subsequent intervals.

Combined with the findings obtained from other studies, the results of this investigation provide a mixture of support and concern regarding the reliability of self-reported internalizing symptoms in children between the ages of 8 and 12 years. The children in this study came from a normal rather than a clinical population. Thus, generalizations about the temporal stability of self-reported internalizing symptoms over short to medium intervals in a clinical population based on the data obtained in this study may be limited. Furthermore, the sample for this investigation was relatively homogeneous in terms of ethnicity. It is unknown how these results might generalize to more ethnically diverse samples.

Future investigations might examine the reliability of self-reported internalizing symptoms in clinical and more ethnically diverse populations. Also, it would be interesting to obtain cross-informant data in children's internalizing symptoms by correlating self-ratings of children and ratings by their parents on a parallel version of the ISSC. In addition, in light of equivocal



**Fig. 1** Attenuation of mean Internalizing Symptoms Scale for Children (ISSC) scores across intervals after initial administration.

nature of the studies that either support or refute the ability of young children (i.e., below the age of 10) to reliably report their internal experience, future investigations might systematically compare those self-report instruments that reportedly have high test-retest reliability (objective self-report) and those that do not (structured interviews).

### **Clinical Implications**

There are three main implications for clinical practice based on the findings from this investigation. First, it appears that preadolescent children between the ages of 8 and 12 years are able to reliably report their experience over short to medium-length intervals on an objective self-report instrument that purports to measure the broad domain of internalizing symptomatology. Second, it appears that a child's endorsement of internalizing symptomatology attenuates over time and repeated administrations on objective self-report instruments. These data might have implications for how clinicians and researchers choose to monitor and measure symptomatology during experimental trials and treatment

regimens. Third, these findings provide strong support for the ISSC as a research and clinical tool for the assessment of internalizing symptoms in children between the ages of 8 and 12 years, which may ultimately prove beneficial in the identification and treatment of childhood internalizing disorders. In certain situations, it may be advantageous to use the ISSC with or instead of other self-report measures such as the CDI, RCDS, RCMAS, or STAIC. For example, a child client may exhibit a broad range of internalizing symptoms that a syndrome-specific instrument might fail to address fully. Also, the ISSC is currently the only self-report measure for children designed to assess positive and negative affectivity, which is increasingly being considered as essential in understanding internalizing disorders (Clark et al., 1990).

## REFERENCES

Achenbach TM (1985), *Assessment and Taxonomy of Child and Adolescent Psychopathology*. Beverly Hills, CA: Sage

Achenbach TM (1991), *Manual for the Youth Self-Report and Profile*. Burlington: University of Vermont Department of Psychiatry

Achenbach TM, Edelbrock C (1987), *Manual for the Youth Self-Report and Profile*. Burlington: University of Vermont Department of Psychiatry

Achenbach TM, McConaughy SH (1992), Taxonomy of internalizing disorders of childhood and adolescence. In: *Internalizing Disorders in Children and Adolescents*, Reynolds WM, ed. New York: Wiley, pp 19-60

Achenbach TM, McConaughy SH, Howell CT (1987), Child and adolescent behavioral and emotional problems: implications of cross-informant correlations for situation specificity. *Psychol Bull* 101:213-232

Anastasi A (1988), *Psychological Testing*, 6th ed. New York: Macmillan  
Anderson JC, Williams S, McGee R, Silva PA (1987), DSM-III disorders in preadolescent children. *Arch Gen Psychiatry* 44:69-76

Borg WR, Gall MD (1989), *Educational Research: An Introduction*, 5th ed. New York: Longman

Boyle MH, Offord DR, Racine Y et al. (1993), Evaluation of the Diagnostic

Interview for Children and Adolescents for use in general population samples. *J Abnorm Child Psychol* 21:663-681

Breen T (1987), Changing family structures: an intervention study in the schools. Doctoral dissertation, University of Wisconsin-Madison

Campbell SB (1986), Developmental issues in childhood anxiety. In: *Anxiety Disorders of Childhood*, Gittelman R, ed. New York: Guilford, pp 24-57

Clarizio H (1984), Childhood depression: diagnostic considerations. *Psychol Sch* 21: 181-197

Clark D, Beck A, Stewart B (1990), Cognitive specificity and positive-negative affectivity: complementary or contradictory views of anxiety and depression? *J Abnorm Psychol* 99: 148-155

Costello AJ (1986), Assessment and diagnosis of affective disorders in children. *J Child Psychol Psychiatry* 27:565-574

Costello AJ, Edelbrock LS, Dulcan MK, Kalas R, Klaric SH (1984), Report on the NIMH Diagnostic Interview Schedule for Children (DIS-C). Washington, DC: National Institute of Mental Health

Costello EJ (1989), Developments in child psychiatric epidemiology. *J Am Acad Child Adolesc Psychiatry* 28:836-841

Cronbach LJ (1960), *Essentials of Psychological Testing*, 2nd ed. New York: Harper & Row

Edelbrock CS, Costello AJ, Dulcan MK, Conover NC (1985), Age differences in the reliability of the psychiatric interview of the child. *Child Dev* 56:265-275

Edelbrock CS, Costello AJ, Dulcan MK, Conover NC, Kalas R (1986), Parent-child agreement on child psychiatric symptoms assessed via structured interview. *J Child Psychol Psychiatry* 27: 181-190

Finch AJ, Saylor CF, Edwards GL, McIntosh JA (1987), Children's Depression Inventory: reliability over repeated administrations. *J Clin Child Psychol* 16:339-341

Finn SE, Kamphuis JH (1995), What a clinician needs to know about base rates. In: *Clinical Personality Assessment: Practical Approaches*, Butcher JN, ed. New York: Oxford University Press, pp 224-235

Fischer M, Rolf JE, Hasazi JE, Cummings L (1984), Follow-up of a

preschool epidemiological sample: cross-age continuities and predictions of later adjustment with internalizing and externalizing dimensions of behavior. *Child Dev* 55: 137-150

Flanery RC (1990), Methodological and psychometric considerations in child reports. In: *Through the Eyes of a Child: Obtaining Self-Reports From Children and Adolescents*, La Greca AM, ed. Boston: Allyn & Bacon, pp 57-82

Graham JR (1993), *MMPI-2: Assessing Personality and Psychopathology*, 2nd ed. New York: Oxford University Press

Kazdin AE (1988), The diagnosis of childhood disorders: assessment issues and strategies. *Behav Assess* 10:67-94

Kovacs M (1985), The natural history and course of depressive disorders in children. *Psychiatr Ann* 15:387-389

Kovacs M (1992), *Children's Depression Inventory*, New York: Multi-Health Systems

Kurdek LA, Berg B (1987), Children's Beliefs About Parental Divorce Scale: characteristics and concurrent validity. *J Consult Clin Psychol* 55:712-718

La Greca AM (1990), Issues and perspectives on the child assessment process. In: *Through the Eyes of a Child: Obtaining Self-Reports From Children and Adolescents*, La Greca AM, ed. Boston: Allyn & Bacon, pp 3-17

Merrell KW (1994), *Assessment of Behavioral, Emotional, and Social Problems: Direct and Objective Methods for Use With Children and Adolescents*. White Plains, NY: Longman

Merrell KW, Anderson KE, Michael KD (1997a), Convergent validity of the Internalizing Symptoms Scale for Children with three self-report measures of internalizing problems. *J Psychoeduc Assess* 15:56-66

Merrell KW, Crowley SL, Walters AS (1997b), Development and factor structure of a self-report measure for assessing internalizing symptoms of elementary-age children. *Psychol Sch* 34: 197-210

Merrell KW, Dobmeyer AC (1996), An evaluation of gender differences in self-reported internalizing symptoms of elementary-age children. *J Psychoeduc Assess* 14: 196-207

Merrell KW, Gill SJ, McFarland H, McFarland T (1996), Internalizing social-emotional characteristics of gifted elementary-age students: a comparative

study. *PsycholSch* 33: 185-191

Merrell KW, Walters AS (in press), *Internalizing Symptoms Scale for Children*. Austin, TX: PRO-ED

Ollendick TH, King NJ (1994), Diagnosis, assessment, and treatment of internalizing problems in children: the role of longitudinal data. *J Consult Clin Psychol* 62:9 18-927

Prout HT, Chizik R (1988), Readability of child and adolescent self-report measures. *J Consult Clin Psychol* 56:152-154

Quay HC, La Greca AM (1986), Disorders of anxiety, withdrawal, and dysphoria. In: *Psychopathological Disorders of Childhood*, 3rd ed, Quay HC, Werry JS, eds. New York: Wiley, pp 73-110

Reynolds CR (1981), Long-term stability of scores on the Revised-Children's Manifest Anxiety Scale. *Percept Mot Skills* 53:702

Reynolds CR, Richmond BO (1985), *Revised Children's Manifest Anxiety Scale Manual*. Los Angeles: Western Psychological Services

Reynolds WM (1986), A model for screening and identification of depressed children and adolescents in school settings. *Prof Sch Psychol* 1:117-129

Reynolds WM (1989), *Rtynoh Child Depression Scale*. Odessa, FL: Psychological Assessment Resources

Reynolds WM (1990), Introduction to the nature and study of internalizing disorders in children and adolescents. *Sch Psychol Rev* 2:137-141

Reynolds WM (1992), The study of internalizing disorders in children and adolescents. In: *Internalizing Disorders in Children and Adolescents*, Reynolds WM, ed. New York: Wiley, pp 1-18

Reynolds WM, Graves A (1989), Reliability of children's reports of depressive symptomology. *J Abnorm Child Psychol* 17:647-655

Saylor CF, Finch AJ, Baskin CH, Furey W, Kelly MM (1984), Construct validity for measures of childhood depression: application of multitrait-multimethod methodology. *J Consult Clin Psychol* 52:955-967

Spielberger CD (1973), *State-Trait Anxiety Inventory for Children*. Palo Alto, CA: Consulting Psychologists Press

Walker LS, Greene JW (1989), Children with recurrent abdominal pain and

their parents: more somatic complaints, anxiety, and depression than other patient families? *J Pediatr Psychol* 14:231-243

Weiss B, Wiesz JR, Politano M, Carey M, Nelson WM, Finch AJ (1991), Developmental differences in the factor structure of the Children's Depression Inventory. *PsycholAssess* 3:38-45

Wisniewski JJ, Mulick JA, Genshaft JL, Coury DL (1987), Test-retest reliability of the Revised Children's Manifest Anxiety Scale. *Percept Mot Skills* 65:67-70