JOURNAL OF CHILD AND ADOLESCENT PSYCHOPHARMACOLOGY
Volume 19, Number 3, 2009

Pp. 241–252
DOI: 10.1089/cap.2008.052

The Child and Adolescent Symptom Inventory-Progress Monitor: A Brief *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition-Referenced Parent-Report Scale for Children and Adolescents

John V. Lavigne, Ph.D.,[1] Taya Cromley,[2] Joyce Sprafkin, Ph.D.,[3] and Kenneth D. Gadow, Ph.D.[3]

Abstract

The Child and Adolescent Symptom Inventory-Progress Monitor–Parent Form (CASI-PM-P) is a 29-item rating scale designed to evaluate symptom change for commonly referred child and adolescent disorders. Its intended applications include monitoring longer-term changes in clinical status and assessing intervention responsiveness. To enhance practicality, there is one version of the CASI-PM-P for all age groups with a common set of norms for both genders. Scoring procedures allow clinicians to assess whether observed symptom changes exceeded chance fluctuations. Using a clinical sample of 2,693 children ages 3–17 years, the 29 symptom-related items were identified that had the best item-to-total minus item correlations on the three age-appropriate scales of the Symptom Inventories. Item-to-total minus item correlations of similar magnitude were also obtained for those items with the standardization sample. In clinical samples, the CASI-PM-P scores had both high levels of internal consistency and test–retest reliability and were sensitive to change in a treated sample. Collectively, the findings support the reliability and validity of the CASI-PM-P as a measure of behavioral change in clinical settings, while continued research will be necessary to improve clinical utility and provide better documentation of the scale's strengths and weaknesses.

Introduction

In the past two decades, the fields of psychology and medicine have placed increasing emphasis on the validation and use of empirically supported treatments and assessment instruments. A meta-analysis conducted in the mid-1980s demonstrated the efficacy of child psychotherapeutic interventions, with some treatment approaches receiving more empirical support than others. These studies and the numerous randomized clinical trials (RCT) conducted since then have made it clear that treatment can be efficacious, but that not all interventions are equally effective. Aided by the Chambless committee's review of empirically supported treatments in psychology (Chambless et al. 1996) and the emphasis on evidence-based practice and continuous quality improvement in health care (Sackett et al. 2000), there has been a growing awareness of the importance of monitoring changes in child behavior during the course of treatment and longer-term follow up in everyday clinical practice in mental health centers, outpatient treatment facilities, and private practice settings. Bickman (2008) argues that mental health services for children are unlikely to improve without such a system.

Hoagwood et al. (1996) proposed a comprehensive conceptual model for assessing child mental health outcomes. This SFCES model proposed that the assessment of outcomes should occur across several domains involving symptoms and diagnoses (*S*ymptoms), child functioning (*F*unctioning), consumer perspectives (e.g., patient satisfaction; *C*onsumer), environment (e.g., marital relationships in the child's home; *E*nvironment), and systems (e.g., change in level or type of service received; *S*ystems). The assessment of symptom change is a key element of this comprehensive model. Although symptom change is assessed routinely with evidence-based measures in RCTs, which comprise much of the child psychotherapy research literature, they are used much less commonly to assess symptom change in every-day child psychotherapy settings.

The recent emphasis on the use of evidence-based interventions has been accompanied by an increased awareness of

[1]Children's Memorial Hospital, Feinberg School of Medicine, Northwestern University, and Mary Ann and J. Milburn Smith Child Health Research Program, Children's Memorial Research Center, Chicago, Illinois.
[2]Feinberg School of Medicine, Northwestern University, Chicago, Illinois.
[3]Department of Psychiatry, State University of New York, Stony Brook, New York.

the discrepancy in expected outcomes for treatments administered in highly controlled research settings (efficacy) versus everyday clinical practice (effectiveness; Weisz et al. 1992). In general, effect sizes are much larger in the former versus the latter. It is noteworthy that there are very few published reports of psychotherapeutic outcomes in children in standard clinic settings. For example, Weisz et al. (1992) note only nine such studies, and our search yields only one additional report (Weiss et al.1999). Undoubtedly, there are a number of reasons why there are so few outcomes studies in standard clinical practice, but the lack of availability of a time-efficient test battery for use in real-world settings is a contributing factor.

Efficacy-based RCTs are typically conducted in a situation that is conducive to monitoring outcomes. Parents consent to treatment and to assessment procedures, thus agreeing to spend time completing outcome assessment batteries in advance, and often being compensated for the time it takes to complete such assessment batteries. It is not surprising that parents will complete outcomes measures under these circumstances. Similarly, in adequately funded effectiveness trials, extensive outcome batteries can also be obtained.

Conditions are quite different in outcome studies that Weiss (1998) refers to as routine monitoring of mental health treatments where the goal is "to determine, for a particular clinic or practice group, the parameters of the treatment's effectiveness." Such studies may be done to: (1) assess "quality of care," determining if a treatment is meeting a predetermined standard; (2) assess "absolute effectiveness," to see if treatment is producing change at all; (3) conduct an "effectiveness comparison," to examine which of two types of treatment in a clinic is more effective; (4) perform a "comparison to a published standard," in which the clinic's outcomes are compared to those conducted elsewhere and published previously; (5) conduct a "client subgroup" study to determine if a treatment is as effective with one group of patients as another. Such studies are crucial if quality is to improve at the standard clinic level, but are seldom implemented or published. The burden imposed on patients to complete complex, time-consuming outcome measures can pose an additional barrier.

There is a plethora of questionnaires designed to assess behavior problems in children—too many to review here. They can, however, be divided into two types: broadband and narrowband instruments. Broadband instruments are designed to measure a number of different diagnostic categories or types of behavior problems. Two examples of commonly used broadband instruments are the Child Behavior Checklist (Achenbach 1991) and Child Symptom Inventory-4 (Gadow and Sprafkin 2002). The most popular broadband instruments all have good psychometric properties and have normative data based on large samples of children. Their major disadvantage as measures of treatment outcome is the length of time they take to complete, often 15–20 minutes. In routine monitoring of clinical services, these instruments may be more suitable for assessing treatment outcomes between long time intervals (e.g., at the beginning and end of treatment). The time that these measures take to complete, however, places an undue burden on patients if monitoring is to occur at more regular intervals, such as at each treatment session or even monthly. Under those circumstances, a much briefer measure may be needed.

Narrowband instruments are designed to assess either a single diagnostic category or type of behavior problem, or else a small number of such problems. Examples of such instruments are the Conners' scale (Conners et al. 1998) and Du Paul scales (DuPaul et al. 1998) for assessing attention-deficit/hyperactivity disorder (ADHD), and the Screen for Child Anxiety Related Emotional Disorders (SCARED) (Birmaher et al. 1997) for assessing anxiety in children. These instruments can be completed relatively quickly. A major drawback of such instruments in clinical settings, however, is the inability of any single narrowband instrument to assess co-morbid conditions, which recent studies show are quite common in typical clinical settings (Weisz et al. 1997). Monitoring outcomes in patients with co-morbid disorders would require the use of multiple narrowband measures, thus increasing the time and response burden and making compliance less likely.

It is also becoming clear that there are advantages to monitoring symptom change on a regular basis rather than after long intervals, which is often the case in pre- and post-treatment procedures. Studies with children show that both sudden gains (i.e., large changes between two sessions) and early gains (i.e., large changes early in the treatment process) occur in child psychotherapy. For example, in one study conducted in a standard clinical treatment setting (Cromley and Lavigne 2008), early gains were associated with better long-term change, and the absence of early gains was associated with poor long-term outcome. This pattern of results suggests that monitoring of treatment progress on a session-by-session basis, at least early in treatment, could be important for deciding when treatment is likely to be effective and when changes in treatment approach are important. To conduct such session-by-session monitoring, however, particularly with patients showing one or more co-morbid conditions, it is necessary to use a brief, broadband measure.

Both clinicians and researchers working with children are aware that multiple informants can provide important information about the child's behavior, including teachers, parents, and the child himself or herself. Frequently, information from all sources can be important in monitoring the child's behavior. Parent reports are particularly important for several reasons: (1) children tend to under-report disruptive behavior symptoms compared to parents; (2) children are more likely to report internalizing symptoms than parents, but are sometimes guarded or defensive in reporting such symptoms, necessitating reliance upon parental report; and (3) younger children have more difficulty reporting symptoms, making parental report critical with younger children. Because self-report is not possible or reliable with young children, and teacher reports may be difficult to obtain on a regular basis in everyday clinical practice, a parent-completed measure may be the only instrument that can be used consistently across all children receiving treatment; and (4) parents are key stakeholders in the therapeutic process, often deciding whether the child will continue in therapy, and it is important to know if they believe that progress is occurring.

There are relatively few brief, broadband measures that may be suitable for monitoring therapeutic progress in mental health settings, The Strengths and Difficulties Questionnaire (Goodman 1997), a 25-item scale designed for use with children ages 4–16, assesses hyperactivity, conduct, and emotional symptoms. The questionnaire was designed for behavioral

screening, and its advantages include emphasis on assessing strengths as well as problems, extensive psychometric validation, and assessment of functional impairment. Its limitations are a less than optimal standardization sample (i.e., differences between the nonpsychiatric and psychiatric sample); subscales are not keyed to *Diagnostic and Statistical Manual of Mental Disorder* (DSM) diagnostic categories; and the scale was not developed for monitoring behavioral change. A second set of measures is the Ohio Scales (Ogles et al. 2001), which contain 44 items and were designed for progress monitoring with children ages 5–18 years. The Ohio Scales were designed specifically for assessing outcomes and are available in parallel parent, youth, and agency worker forms. Although individual items are keyed to *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV) (American Psychiatric Association 1994), only an overall "problem severity" scale can be derived, making it difficult to monitor children who have problems in a specific area. The Ohio Scales were shown to detect changes in symptoms over time, but there are no reports on sensitivity to change associated with treatment. A third measure, the Youth Outcome Questionnaire (YOQ) (Dunn et al. 2005), was designed for parent and adolescent self-report, but the psychometric analyses were conducted primarily with the latter. The YOQ is not DSM-IV–referenced and therefore requires the extrapolation of obtained scores to improvement in traditionally defined diagnostic constructs. Finally, the Peabody Treatment Progress Battery is available, but only for youths ages 11–18 and not for younger children (Bickman 2008). In summary, there clearly is a need for additional research into parent-completed instruments to monitor therapeutic progress across all age groups for clinical settings.

The parent (P) version of the Child and Adolescent Symptom Inventory-Progress Monitor (CASI-PM-P) was developed for use with patients referred to child and adolescent mental health facilities where time and staffing constraints greatly limit data collection, particularly instruments that need to be administered repeatedly in the course of clinical care. As a result, we sought to develop a rating scale that: (1) was suitable for a broad age range (3–18 years); (2) was relatively brief and easy to complete; (3) was available in a single form for all ages with a common set of clinical norms for both genders to facilitate scoring and administration; (4) had normative data for non-referred youths for comparison purposes; (5) allowed for monitoring specific types of disorders as well as overall symptom levels; (6) provided an overall assessment of functioning; and (7) included appropriate procedures to allow clinicians to assess whether observed changes exceeded chance fluctuations in symptom reports.

The present report describes the development of the CASI-PM-P and its reliability, validity, and sensitivity to change associated with treatment. One common approach to assessing a measure's sensitivity to change is to determine if it reflects within-subject variation in patients undergoing an effective intervention (Vermeersch et al. 2000). This approach, however, is difficult to implement in real-world clinical settings where, for ethical and practical reasons, withholding treatment (i.e., comparison condition) is problematic. It is also not possible to determine *a priori* whether a measure proven responsive to therapeutic improvement in a randomized trial is equally helpful in routine clinical care, which has led researchers to seek alternative comparison groups. For example, Veermersch et al. (2004) compared behavioral improvement resulting from therapy in regular clinic settings with changes over time in a group of college students not seeking treatment to demonstrate that their instrument was sensitive to change. With child patients, Weisz et al. (1987) reported that there were no differences in demographic characteristics and clinical measures between patients who dropped out of psychotherapy and those who completed treatment. These results agree with other reports (Levitt 1971; McAdoo and Roeske 1973) and suggest that dropouts may be an acceptable, naturally occurring control group for use in outcome research. Although there may be limitations in using dropouts to assess overall effectiveness, data from dropouts receiving only a small "dose" (i.e., number of sessions) of treatment may be useful in assessing whether an instrument is responsive to therapeutic change. In the present study, we hypothesized that children who continued in treatment would show greater progress than children who left treatment early. Given the negative findings for outpatient treatment, we did not expect these changes to be very large, but did expect that the trajectory of change would be greater for those receiving more sessions.

When assessing therapeutic improvement, it is also important not to over-interpret small changes in symptom manifestation that may be due to random fluctuations in symptoms, reporting error, or unreliability of the measure. In the present study we computed a reliable change index (RCI) (Jacobson and Truax 1991; Jacobson et al. 1999) and examined whether those continuing in treatment were more likely to achieve a reliable change.

Method

Participants

To generate a comprehensive picture of the psychometric properties of the CASI-PM-P, it was necessary to analyze data from several different samples.

Item selection sample. The clinical sample used to identify the optimal items for inclusion in the CASI-PM-P consisted of 2,693 parents of children referred to a large Midwestern child psychiatric clinic between 2001 and 2005. The sample included children ages 3–5 years ($n = 605$, 442 boys, 163 girls), 6–11 years ($n = 1537$, 1,069 boys, 468 girls), and 12–17 years ($n = 551$, 312 boys, 239 girls). The clinic serves a diverse clientele, and is approximately 45% self-identified white, non-Hispanic European American; 23%, African American; and 24%, Hispanic American.

Test-retest samples. After the measure was developed, a small sample of 24 families whose child was beginning a diagnostic evaluation were recruited from the same clinic, gave consent, and completed a CASI-PM-P at the time of the initial evaluation and again 1–2 weeks later.

Subsequently, test–retest reliability was reexamined in a second clinical sample of 178 children ages 3–17 who were routinely completing CASI-PM-Ps at clinic visits. This sample completed the Long Form of the Symptom Inventory (Gadow and Sprafkin 1997b; Gadow and Sprafkin 1998; Gadow and Sprafkin 2000) at the initial evaluation and the CASI-PM-P at a subsequent clinic visit over a 1- to 33-week period.

Behavior change sample. Sensitivity to change was assessed in a sample of 138 children and adolescents (3–18 years). One of their caregivers, usually the mother, completed the CASI-PM-P at consecutive clinic visits. Treatment was provided by 21 clinicians, of whom 62% ($n = 13$) were staff and 38% ($n = 8$) were trainees. Clinical care was provided by doctoral level psychologists (52%, $n = 11$), trainees in clinical psychology (33% $n = 7$), and other staff (15%, $n = 3$).

The type of treatment provided is best described as "treatment as usual" (TAU) (Garland et al. 2006; Kolko 2006). Therapists chose the type of treatment provided, which included interventions from one or more of the following orientations: supportive, family, psychodynamic, behavioral, and cognitive behavioral. Some therapists followed manuals derived from empirically supported treatment protocols whereas others did not. Most therapists described themselves as eclectic or integrative. When considered appropriate, treatment also involved medication.

Standardization samples. The standardization samples for the long form of the Symptom Inventories measures from which the CASI-PM-P were derived were also used in psychometric analysis. These included the standardization samples for the Early Childhood Inventory-4 (ECI-4) (boys $n = 268$, girls $n = 251$) (Gadow and Sprafkin 1997b), Child Symptom Inventory-4 (CSI-4) (boys $n = 272$, girls $n = 279$) (Gadow and Sprafkin 2002), and the Adolescent Symptom Inventory-4 (ASI-4) (boys $n = 389$, girls $n = 390$) (Gadow and Sprafkin 1998). These samples are described in considerable detail in prior publications (Gadow and Sprafkin 2006).

Measures

Symptom Inventory Long Form. The Symptom Inventories are DSM-IV–referenced, broadband instruments for assessing child and adolescent emotional and behavioral problems. The Long Form of these measures refers to three age-appropriate rating scales: the ECI-4 (Gadow and Sprafkin 1997b; Gadow and Sprafkin 2000) for 3–5 year olds in preschool programs, the CSI-4 (Gadow and Sprafkin 1994; Gadow and Sprafkin 2002) for 5–12 year olds in elementary school, and the ASI-4 (Gadow and Sprafkin 1997a; Gadow and Sprafkin 1998) for 12–18 year olds in middle or secondary schools. In the present study, items from the following symptom categories (scales) were used: attention-deficit/hyperactivity disorder, including the inattention type (ADHD-I), the hyperactive/impulsive type (ADHD-H), and the combined type (ADHD-C); oppositional defiant disorder (ODD); conduct disorder (CD); generalized anxiety disorder (GAD); separation anxiety disorder (SAD); social phobia; and major depressive episode (MDE).

The validity of the parent and teacher versions of the Symptom Inventories has been examined in scores of studies and includes comparisons with dimensional rating scales, laboratory measures, chart diagnoses, and structured psychiatric interviews; comparisons between symptomatic and asymptomatic samples and between samples with specific behavioral syndromes; and response to behavioral and pharmacological interventions (Gadow and Sprafkin 1994; Gadow and Sprafkin 2002; Gadow and Sprafkin 2006). In comparison to chart diagnoses (Gadow and Sprafkin 1994; Gadow and Sprafkin 2002) and structured psychiatric interviews (Grayson and Carlson 1991; Sprafkin et al. 2002), sensitivity and specificity of parent screening cutoff scores generally range from 0.70 to 0.90 for ADHD, ODD, and CD, which, according to criteria outlined by Cicchetti et al. (1995), is fair to excellent. The specificity of parent GAD and MDE screening cutoff scores is good to excellent, but sensitivity is low (0.56). The sensitivity of symptom severity cutoff scores, however, is fair to excellent for all five disorders. In developing the CASI-PM-P, items were selected from the Long Form of the Symptom Inventories that correlated highly with the individual symptom scales, and that constituted scales with good internal consistency and test-retest reliability.

Statistical analyses

Pearson correlations were conducted to examine item-to-total minus item correlations, and correlations between scales. To examine sensitivity to change, we compared changes for an early dropout versus continued treatment group, while entering the pretreatment total CASI-PM-P score as a covariate to control for initial severity. Inclusion of the dropout group was necessary to control for the effects of regression to the mean. By entering the pretreatment CASI-PM-P score as a covariate, statistically significant group differences that emerged would indicate that sensitivity to change on the CASI-PM-P scores was not restricted to the highest scores.

To aid in the interpretation of change, the use of a reliable change index (RCI) can be useful (Jacobson et al. 1999; Jacobson and Truax 1991) and was computed for each scale and summary score. The procedure described by Jacobson and Truax (1991) was used, in which the pretreatment scale score (x_1) is subtracted from the score on the measure obtained during or after treatment (x_2) and then divided by the standard error of difference (S_{diff}) for that particular scale. When the RCI is less than −1.96, posttest scores reflect a change that exceeds a level that might be expected based upon the reliability of the measure.

Results

Item selection

Item-to-total minus item correlations in a clinic sample. The items common to all three age-designated Long Forms formed the item pool for these analyses. In selecting items for the CASI-PM-P, we sought to identify the items that showed the strongest correlations with the Long Form of the Symptom Inventory. The initial analysis correlated each item with the corresponding symptom scale score minus the score for that item. This was done for each of the three different age groups for the Long Form of each of the three Symptom Inventories. Items with the best item-to-total minus item correlations across the three age groups were then chosen for inclusion in the CASI-PM-P. Table 1 shows the results of this analysis for the 29 symptom-related items included in the CASI-PM-P.

Overall, items showed moderate (0.5–0.6) to good (0.7 and above) item-to-total minus item correlations across 25 of the 29 items. Three items had moderate to high item-to-total minus item correlations among older children and adolescents but were lower for preschoolers ("is irritable most of the day," "shows little interest in [or enjoyment of] pleasurable activities," "has low energy level or is tired for no apparent reason"). All of these items are associated with MDE, which

TABLE 1. ITEM-TO-TOTAL MINUS ITEM CORRELATIONS IN A CLINICAL SAMPLE ($N = 2693$)

Item[a]	*ECI-4*	*CSI-4*	*ASI-4*
1. Fails to give attention to details	0.64	0.72	0.76
2. Difficulty paying attention	0.76	0.74	0.82
3. Difficulty following instructions	NA	0.80	0.84
4. Difficulty organizing tasks	0.72	0.74	0.78
5. Difficulty remaining seated	0.77	0.78	0.76
6. Difficulty doing things quietly	0.75	0.75	0.71
7. Is "on the go"	0.74	0.75	0.72
8. Difficulty awaiting turn	0.69	0.75	0.72
9. Defies what you tell them to do	0.72	0.76	0.75
10. Angry and resentful	0.78	0.79	0.78
11. Takes anger out on others	0.78	0.78	0.78
12. Deliberately annoys others	0.76	0.74	0.72
13. Argues with adults	0.74	0.75	0.77
14. Bullies, threatens, or intimidates	0.76	0.71	0.64
15. Starts physical fights	0.75	0.65	0.62
16. Destroys others' property	0.69	0.69	0.63
17. Acts restless or edgy	NA	0.70	0.66
18. Irritable	0.47	0.65	0.61
19. Tense or unable to relax	NA	0.70	0.67
20. Difficulty controlling worries	NA	0.60	0.60
21. Worries parents will be hurt	0.68	0.66	0.68
22. Worries disaster will separate parents	0.56	0.66	0.66
23. Upset when separated from parents	0.63	0.59	0.71
24. Excessively shy	0.84	0.62	0.67
25. Cries, freezes, or withdraws from interacting	0.84	0.50	NA
26. Depressed for most of day	0.70	0.71	0.76
27. Little interest in pleasurable activities	0.38	0.61	0.70
28. Low energy level	0.43	0.53	0.65

[a]Truncated version of item.
All correlations $p < .0001$.
Abbreviations: ECI-4, Early Childhood Inventory-4; CSI-4, Childhood Symptom Inventory-4; ASI-4, Adolescent Symptom Inventory-4; NA, not applicable.

occurs at low rates in that age group. One item ("when put in an uncomfortable social situation, child cries, freezes, or withdraws from interacting") is more common for the preschool-age group and shows a high item-to-total minus item correlation for that age group, but was lower in school-age children for whom such extreme responses are less likely to occur in situations arousing social anxiety.

Item-to-total minus item correlations in the standardization sample. After ascertaining which items seemed acceptable in the clinic sample based on their item-to-total correlations, item-to-total correlations were then calculated for the CASI-PM-P items in the standardization sample for the Long Form, with similar results (Table 2).

Final CASI-PM-P scale description

The CASI-PM-P scale that emerged after these analyses contains 29 items from the Long Form: ADHD-I (Items 1–4), ADHD-H (Items 5–8), ODD (Items 9–13), CD (Items 14–16), GAD (Items 17–20), SAD (Items 21–23), social phobia (SP) (Items 24–25), and MDE (Items 18, 26–28). One additional item (Item 29) assesses functional impairment.

A summary score is provided for ADHD-C, and there are two additional CASI-PM-P summary scores that are not included in the Long Form: Total Scale score (sum of items 1–29), and a Total Anxiety Scale score (sum of items 17–25).

Consistent with the Long Form, each item on the CASI-PM-P is scored as occurring never (0), sometimes (1), often (2), or very often (3). Because the CASI-PM-P is intended for assessing behavioral change, the parent is asked to rate the child's behavior since the last treatment visit.

Scoring is accomplished by summing the appropriate scores for the items of each individual or summary scale. As with the Long Form, related items are presented sequentially to facilitate visual inspection and scoring. Prior research suggests there is little difference in the psychometric properties whether items are grouped according to diagnostic construct or randomly inter-mixed (Sprafkin and Gadow 2007).

Correlations between scales

Intercorrelations between CASI-PM-P subscales show a predictable pattern, with moderate to high correlations between externalizing scales, including the ADHD scales, ODD, and CD (Table 3). Among the internalizing scales, the correlation between MDE and GAD scales were moderate to high, whereas SAD and social phobia correlated at low levels with other scales.

Reliability

Internal consistency. Internal consistency was calculated for the entire clinical sample and was high for each scale for all ages (Table 4). Internal consistency was high for school-age children on all scales, but only moderate for the two-item social phobia scale. Internal consistency was high for adolescents for each subscale, but could not be calculated for social phobia because only one item concerning social phobia in that age group was included on the CASI-PM-P. For preschoolers, internal consistency was high for ADHD, CD, SAD, ODD, and the summary scores for total problems and anxiety. It was moderate in this age group for social phobia, MDE, and low for GAD.

Test–retest reliability. Because the CASI-PM-P is designed primarily to assess behavior change, it was particularly important to determine its short-term stability in a clinical sample. For that reason, the stability coefficient for the CASI-PM-P was assessed in a clinical sample of children undergoing an evaluation or just beginning to receive care. The 1- to 2-week stability ($M = 10.33$ days, standard deviation [SD] = 5.68 days) in a subsample ($n = 24$) of clinic-referred children was high for all scales (Table 5).

The CASI-PM-P was then administered to 178 children at the time of their initial evaluation and then again at a subsequent clinic visit 1–33 weeks ($M = 28.7$ days, SD = 34.7 days) later (Table 6). In this larger sample over a longer time period, test–retest reliability was high (≥0.7) for ADHD, ODD, and CD scores, and moderate (0.5 to 0.6) for GAD, SAD, MDE, social phobia, and Total Anxiety scores.

Functional impairment

We hypothesized that the functional impairment item would be related to the severity of symptoms of each scale,

TABLE 2. ITEM-TO-TOTAL MINUS ITEM CORRELATIONS IN THE STANDARDIZATION SAMPLE

Item[a]	*ECI-4*	*CSI-4*	*ASI-4*
1. Fails to give attention to details	0.59	0.66	0.66
2. Difficulty paying attention	0.63	0.74	0.75
3. Difficulty following instructions	0.69	0.79	0.76
4. Difficulty organizing tasks	0.68	0.74	0.73
5. Difficulty remaining seated	0.68	0.72	0.65
6. Difficulty doing things quietly	0.63	0.69	0.67
7. Is "on the go"	0.67	0.66	0.56
8. Difficulty awaiting turn	0.64	0.68	0.60
9. Defies what you tell them to do	0.67	0.74	0.72
10. Angry and resentful	0.72	0.74	0.72
11. Takes anger out on others	0.65	0.70	0.62
12. Deliberately annoys others	0.70	0.68	0.64
13. Argues with adults	0.65	0.69	0.67
14. Bullies, threatens, or intimidates	0.66	0.58	0.50
15. Starts physical fights	0.58	0.64	0.54
16. Destroys other's property	0.57	0.72	0.34
17. Acts restless or edgy	NA	0.70	0.69
18. Irritable	0.22	0.62	0.57
19. Tense or unable to relax	NA	0.64	0.65
20. Controlling worries	NA	0.60	0.55
21. Worries parents will be hurt	0.65	0.70	0.60
22. Worries disaster will separate parents	0.50	0.61	0.57
23. Upset when separated from parents	0.56	0.59	0.43
24. Excessively shy	0.21	0.36	0.64
25. Cries, freezes, or withdraws from interacting	0.22	0.35	0.64
26. Depressed for most of the day	0.42/0.36	0.55/0.49	0.67/0.60
27. Little interest in pleasurable activities	0.36	0.45	0.65
28. Low energy level	0.41/0.33	0.41/0.42	0.58/0.60
29. How often do above interfere with school work or social activities?	NA	NA	NA

[a]Truncated version of item.
All correlations $p < .0001$.
Abbreviations: ECI-4, Early Childhood Inventory-4; CSI-4, Childhood Symptom Inventory-4; ASI-4, Adolescent Symptom Inventory-4; NA, not applicable.

with children described as having symptoms that never interfered with their functioning having the lowest, or least severe symptoms, and those with problems that interfered "very often" with their functioning having the highest severity scores. Analyses of variance showed this to be true for each scale ($p < 0.001$) except SAD, $F(3, 137) = 0.193$, $p = 0.9$

Validity

Correlations with the Long Form. The findings of numerous studies show that the Long Form of the Symptom Inventories is useful in screening children for a variety of DSM-IV disorders (Gadow and Sprafkin 1997a; Gadow and

TABLE 3. CASI-PM-P SCALE INTERCORRELATIONS

	ADHD-I	*ADHD-H*	*ADHD-C*	*ODD*	*CD*	*GAD*	*SAD*	*SP*	*Anx*	*MDD*	*Dysth*	*Dep*	*Total*
ADHD-I													
ADHD-H	0.54												
ADHD-C	0.86	0.89											
ODD	0.38	0.53	0.52										
CD	0.31	0.50	0.47	0.69									
GAD	0.25	0.19	0.25	0.42	0.28								
SAD	0.08	0.15	0.13	0.20	0.15	0.28							
SP	0.12	0.08	0.11	0.11	0.07	0.22	0.29						
Anx	0.23	0.21	0.25	0.38	0.26	0.81	0.71	0.59					
MDD	0.25	0.12	0.20	0.41	0.29	0.68	0.26	0.33	0.64				
Total	0.65	0.69	0.76	0.81	0.67	0.65	0.41	0.35	0.70	0.63			

Note: All correlations significant at the 0.01 level, $n = 2693$.
CASI-PM-P, Child and Adolescent Symptom Inventory-Progress Monitor–Parent Form; ADHD-I, Attention-deficit/hyperactivity disorder–inattentive type; ADHD-H, ADHD–hyperactive/impulsive type; ADHD-C, ADHD–combined type; ODD, oppositional Defiant disorder; CD, conduct disorder; GAD, generalized anxiety disorder; SAD, separation anxiety disorder; SP, social phobia; Anx, total anxiety (combining anxiety scales); MDD, major depressive disorder; Dysth, dysthymia; Dep, depression.

TABLE 4. INTERNAL CONSISTENCY (ALPHA) FOR EACH AGE GROUP FOR THE CASI-PM-P

Symptom category	*All ages*	*ECI*	*CSI*	*ASI*
ADHD-I	0.87	0.85	0.87	0.90
ADHD-H	0.88	0.86	0.87	0.85
ODD	0.90	0.89	0.90	0.89
CD	0.82	0.86	0.81	0.77
GAD	0.83	0.44	0.84	0.81
SAD	0.79	0.78	0.80	0.81
Social phobia	0.62	0.66	0.58	—[a]
MDD	0.76	0.67	0.76	0.81
Total	0.92	0.90	0.92	0.91
ADHD-C	0.89	0.89	0.89	0.89
Anxiety	0.80	0.74	0.80	0.78

Note: All correlations significant at the 0.01 level.

[a]Can't compute, only one item.

CASI-PM-P, Child and Adolescent Symptom Inventory-Progress Monitor–Parent Form; ECI, Early Childhood Inventory; CSI, Child Symptom Inventory; ASI,Adolescent Symptom Inventory; ADHD-I, Attention-deficit/hyperactivity disorder–inattentive type; ADHD-H, ADHD–hyperactive/impulsive type; ADHD-C, ADHD–combined type; ODD, oppositional defiant disorder; CD, conduct disorder; GAD, generalized anxiety disorder; SAD, separation anxiety disorder; Anxiety, Total anxiety (combining anxiety scales); MDD, major depressive disorder.

Sprafkin 2000; Gadow and Sprafkin 2002). In establishing the validity of the CASI-PM-P, it was therefore important to demonstrate that the CASI-PM-P correlated with the Long Form. Our analyses indicate that the CASI-PM-P correlates very highly with the Long Form (Table 7). The correlation for GAD was somewhat lower than for other scales because the internal consistency of GAD was lower for that age group (see Table 4).

Concurrent validity

Smith et al. (2000) note that one cannot assume that the validity of the longer or complete version of a rating scale applies to a shorter version. Ideally, assessments of the validity of a shorter form will be conducted in independent samples, but these data are not currently available for the CASI-PM-P. Nevertheless, an estimate can be made of the criterion validity of the shorter form, using the equation, $r(sc) = r(fc) \times r(ss)/r(ff)$. In this equation, $r(fc)$ is the correlation between the short and long form, $r(ss)$ is the reliability of the short form, $r(ff)$ is the reliability of the long form, and $r(sc)$ is the estimated correlation between the short form and the criterion. The concurrent validity of the CSI-4 Long Form with the Child Behavior Checklist (CBCL) (Achenbach 1991) has been reported previously (Sprafkin et al. 2002) for boys ages

TABLE 5. TEST–RETEST RELIABILITY (R) AFTER 1 MONTH ($N = 24$)

Symptom Category	*r*
Total	0.92
ADHD-I	0.70
ADHD-H	0.90
ADHD-C	0.84
ODD	0.71
CD	0.74
GAD	0.71
SAD	0.98
MDD	0.83
Social phobia	0.84
Anxiety	0.94

Note: All correlations significant at the 0.01 level.

ADHD-I, Attention-deficit/hyperactivity disorder–inattentive type; ADHD-H, ADHD–hyperactive/impulsive type; ADHD-C, ADHD–combined type; ODD, oppositional defiant disorder; CD, conduct disorder; GAD, generalized anxiety disorder; SAD, separation anxiety disorder; Anxiety, total anxiety (combining anxiety scales); MDD, major depressive disorder.

TABLE 6. TEST–RETEST RELIABILITY FOR A SAMPLE OF CHILDREN IN TREATMENT WITH A MEAN OF 27.6 DAYS BETWEEN ADMINISTRATIONS OF THE CASI-PM-P ($N = 178$)

Symptom category	*r*
Total	0.83
ADHD-I	0.73
ADHD-H	0.86
ADHD-C	0.83
ODD	0.72
CD	0.72
GAD	0.62
SAD	0.54
MDD	0.68
Social phobia	0.66
Anxiety	0.68

Note: All correlations significant at the 0.01 level.

CASI-PM-P, Child and Adolescent Symptom Inventory-Progress Monitor–Parent Form ADHD-I: Attention-deficit/hyperactivity disorder–inattentive type; ADHD-H, ADHD–hyperactive/impulsive type; ADHD-C, ADHD–combined type; ODD: oppositional defiant disorder; CD, conduct disorder; GAD, generalized anxiety disorder; SAD, separation anxiety disorder; Anxiety, total anxiety (combining anxiety scales); MDD, major depressive disorder.

TABLE 7. CORRELATION (R) OF THE CASI-PM-P WITH THE LONG FORM OF THE CSI

Symptom category	*ECI*	*CSI*	*ASI*
ADHD-I	0.95	0.95	0.96
ADHD-H	0.95	0.96	0.95
ADHD-C	0.97	0.97	0.97
ODD	0.98	0.98	0.98
CD	0.92	0.91	0.87
GAD	0.65	0.93	0.93
SAD	0.88	0.88	0.91
Social phobia	0.86	0.95	0.87
MDD	0.91	0.94	0.96
Dysthymia	0.92	0.96	0.96
Depression	0.94	0.96	0.97
Anxiety	0.84	0.94	0.94

Note: All correlations significant at the 0.01 level.

CASI-PM-P, Child and Adolescent Symptom Inventory-Progress Monitor–Parent Form; ECI, Early Childhood Inventory; CSI, Child Symptom Inventory; ASI, Adolescent Symptom Inventory; ADHD-I, Attention-deficit/hyperactivity disorder–inattentive type; ADHD-H, ADHD–hyperactive/impulsive type; ADHD-C, ADHD–combined type; ODD, oppositional defiant disorder; CD, conduct disorder; GAD, generalized anxiety disorder; SAD, separation anxiety disorder; Anxiety, total anxiety (combining anxiety scales); MDD, major depressive disorder.

6–11. The CBCL is a widely used, factor analytically derived, behavior checklist, but the 1991 version of the CBCL was not, however, keyed to DSM-IV. As such, only moderate correlations between either Long Form of the CSI and the CASI-PM-P with the CBCL would be expected. As Table 8 indicates, the concurrent validity of the CASI-PM-P scales with corresponding CBCL factor scores was moderate to high, ranging from 0.50 to 0.92. Further research will be needed to determine if item coverage for each scale is sufficient.

Responsiveness to change

We hypothesized that children who continued in treatment would show greater progress than children who left treatment early. To test this hypothesis, we compared changes for an early dropout group that completed no more than two treatment sessions versus a "continued treatment" group that continued in treatment. The mean number of CASI-PM-Ps completed by the participants in the two group was 4.4 (SD = 3.2; range, 2–17) For the early dropout group, a change score was calculated by subtracting the pretreatment CASI-PM-P from the CASI-PM-P completed at the second treatment session, the time at which the youth had completed only one visit. For the group continuing in treatment, the change score was calculated by subtracting the pretreatment CASI-PM-P from the CASI-PM-P completed at the youth's last visit during the study period. At that time, the youth and family had completed between 2 and 17 CASI-PM-Ps at clinic visits. The dropout ($M = 24.58$, SD = 12.1) and continued treatment groups ($M = 27.72$, SD = 14.53) did not differ on initial baseline scores for overall problem, $t(135) - 1.30$, $p = 0.194$.

Consistent with the hypothesis, those continuing in treatment showed greater improvement than did the early dropouts (M change for dropouts − 2.35, SD = 8.13, $n = 52$; M change for those continuing for 5 or more sessions − 7.83, SD = 12.38, $n = 85$), F(1, 134) = 6.37, $p = 0.013$. Total initial severity also predicted improvement, F(1, 134) = 62.36, $p = 0.000$. To achieve a reliable change, a 12-point improvement in Total score was needed. A reliable change was significantly more likely for those continuing in treatment (32.9%, $n = 28$) than dropping out (7.7%, $n = 4$) ($X^2 = 11.49$, df = 1, $p = 0.001$). Overall, the pattern of scores is consistent with the Total score for the CASI-PM-P being sensitive to therapeutic improvement. .

Calculating sensitivity to change for each subscale of the CASI-PM-P is more difficult with a sample of this size. Most children will not show elevated scores on a particular scale (e.g., SAD), creating a "floor effect" that makes observing change difficult. Examining changes on a subscale can best be done with children exhibiting a particular disorder. In the present sample, the sample size for the combined group of children with ADHD (combining children with clinician-assigned diagnoses of ADHD-I, ADHD-H, ADHD-C, and ADHD-not otherwise specificed [NOS]) was large enough to examine changes on the ADHD-C scale. There were no differences pretreatment between the dropouts ($M = 11.77$, SD = 5.34) and those continuing in treatment ($M = 14.22$, SD = 6.19) on the ADHD-C scale, $t(47) = -1.46$, $p = 0.15$. There was a greater improvement on the ADHD-C scale for those continuing in treatment (M change for dropouts = −2.14, SD = 3.36, $n = 22$; M change for those continuing for five or more sessions = −5.18, SD = 5.30, $n = 27$), F(1, 48) = 4.14, $p = 0.048$. Initial problem severity also predicted improvement, F(1, 48) = 13.20, $p = 0.001$. Unlike reliable change based on the Total score, however, the difference between dropouts (13.6%, 3/22) and those continuing in treatment (37.0%, 10/27) for achieving a reliable change only approached significance ($X^2 = 3.41$, df = 1, $p = 0.065$). For all other diagnostic groups the sample sizes for specific disorders or even groups of disorders (e.g., combining all types of anxiety or depression) were too small for adequate power (e.g., there were less than 10 dropouts for each disorder or groups of disorders, and less than 11 "continuers" for each disorder).

Discussion

The results of the present study indicate that the CASI-PM-P shows good internal consistency and test–retest reliability overall and for each of its subscales in both a clinic-referred and a nonreferred, community-based sample. Support for its validity comes from high correlations with the age-designated Long Forms of the measure, estimates of its concurrent validity with the CBCL, and indications of its sensitivity to change for the overall scale (Total score) and for ADHD-C subscale score.

Smith et al. (2000) described several methodological concerns that they refer to as "sins of short-form development."

TABLE 8. CONCURRENT VALIDITY WITH THE CBCL

CSI scale	*CBCL scale*	*Full-criterion*	*Full reliability*	*Short reliability*	*Short-criterion*	*Difference*
ADHD-I	Attention problems	0.63	0.75	0.70	0.59	0.04
ADHD-H	Attention problems	0.49	0.84	0.90	0.53	−0.04
ADHD-C	Attention problems	0.65	0.79	0.84	0.69	−0.04
ODD	Externalizing problems	0.81	0.80	0.71	0.72	0.09
CD	Externalizing problems	0.57	0.46	0.74	0.92	−0.35
GAD	Anxious/depressed	0.62	0.67	0.71	0.66	−0.04
SAD	Anxious/depressed	0.48	0.74	0.98	0.63	−0.15
MDD	Anxious/depressed	0.66	0.71	0.83	0.77	−0.11
Social phobia	Withdrawn	0.52	0.87	0.84	0.50	0.02

Note: All correlations significant at the 0.01 level.

CBCL, Child Behavior Checklist; CSI, Child Symptom Inventory; ADHD-I, Attention-deficit/hyperactivity disorder–Inattentive type; ADHD-H, ADHD–hyperactive/impulsive type; ADHD-C, ADHD–combined type; ODD, oppositional defiant disorder; CD, conduct disorder; GAD, generalized anxiety disorder; SAD, separation anxiety disorder; MDD, major depressive disorder.

Their survey suggests shorter versions of established measures rarely avoid all limitations, but assessing which sins are committed and which are avoided provides a useful framework for assessing a measure's strengths and limitations. Each of these methodological concerns and the way they were addressed in developing the CASI-PM-P are reviewed, below.

General sin 1: Assuming that the evidence for the reliability and validity of the original measure (Long Form) also applies to the shorter or abridged form. The development of the CASI-PM-P avoided this problem by reexamining the reliability, validity and sensitivity to change specifically for the short form.

General sin 2: Assuming the abbreviated measure requires less support for validity than the long form does because it is shorter. We have not assumed that the validity of the Long Forms of the Symptom Inventories applies to the CASI-PM-P. Because the clinical uses of the two forms differ, their respective types of validation studies also differ. The Long Forms of the Symptom Inventories were intended to be used in assessing whether problem behaviors were present for use in screening and diagnosis, whereas the CASI-PM-P was developed to assess sensitivity to behavioral change. The initial validation of the CASI-PM-P was based on assessing its correlation with the Long Forms, concurrent validity with another measure, and sensitivity to change. Although obtained findings were favorable, it is clear that further validity studies will be needed in both clinical settings and randomized trials to firmly establish its sensitivity to treatment effects.

Specific sin 1: Developing a short form of a measure from a longer measure that itself lacks sufficient validation. This problem was avoided because the Long Forms of the Symptom Inventories have undergone extensive validation.

Specific sin 2: Failing to show that the short form preserves the content coverage of each factor in the measure. By design, the CASI-PM-P does not have the breadth of the Long Forms. In shortening the Long Form, items from the most common types of disorders in referred children were retained, whereas items related to less common disorders were eliminated. This seems reasonable because the purpose of the Long and Short Forms is different. The CASI-PM-P was not intended to replace the Long Forms for use in diagnostic evaluations and was developed to help monitor therapeutic progress.

Breadth of coverage of some aspects of the most common disorders, however, may have been reduced. For example, Smith et al. (2000) noted that a popular approach to develop briefer versions of measures is to select items with the highest item-to-total correlations with each factor, and this is the approach taken in developing the CASI-PM-P. This procedure has the advantage of preserving high internal reliability, but does so at the cost of decreasing the breadth of the items selected. When breadth of coverage may have been reduced, Smith et al. (2000) suggest an examination of the content of the scales.

The DSM-IV includes nine criteria of ADHD-I type, of which four are measured directly on the CASI-PM-P. These include items assessing the child's ability to pay attention and to organize and complete tasks. Not included are items pertaining to losing things, distractibility, task avoidance, and forgetfulness. There is clear overlap between the omitted and the retained items that appears to allow for reasonable content coverage for the Short Form (e.g., children who lose things or are forgetful are likely to appear disorganized; those who avoid tasks will have difficulty completing them; children who are distractible will appear to be inattentive). For the nine hyperactive-impulsive DSM-IV criteria, four are directly measured on the CASI-PM-P and others not included are closely related to those that are (e.g., children who "run about or climb excessively" are likely to be seen as "on the go)." "Fidgety" children who remain seated, or children whose hyperactivity and impulsivity is evidenced verbally (talks excessively, blurts out answers to questions), however, may not be detected as well with the CASI-PM-P.

Coverage of the CASI-PM-P for CD is less complete than for other disorders. Problems with aggression directed toward others or property are likely to be detected, but serious rule violations (running away, curfew violations) and theft may not be because these items are not included on the CASI-PM-P. In settings where CD is a major clinical concern, the CASI-PM-P may be more useful for assessing co-morbid symptoms.

The DSM-IV includes eight criteria for ODD, of which five are included in the CASI-PM-P. Items not included are closely related to those that are (e.g., loses temper, easily annoyed) are omitted, but are likely to be detected by items inquiring whether the child is often angry or often argumentative. Children who blame others for mistakes, however, may be missed on the CASI-PM-P if they show few other ODD-related symptoms.

For GAD, DSM-IV notes seven specific symptoms, of which four are assessed directly in the CASI-PM-P and a fifth, concentration problems, is likely to be detected by ADHD-I items. Not assessed in the CASI-PM-P are two items pertaining to sleep-related difficulties (problems with sleeping, easily fatigued). For SAD, DSM-IV notes eight symptoms, of which only three are included on the CASI-PM-P. These two items are core symptoms (worries about harm to parents or parents not "coming back" to them; worries that they will be separated from the parent; upset when separation is anticipated), and it is likely that a child with SAD will exhibit one or more of these core symptoms. Other symptoms that may accompany those worries (e.g., nightmares, physical symptoms, avoiding school), however, are not assessed directly on the CASI-PM-P. For depression, core features of depressed mood, anhedonia, and fatigue are assessed, but not problems of weight loss or insomnia. Recurrent thoughts of death occur so infrequently in younger children that this item was eliminated in the item-to-total correlation analyses. When suicide is of concern, it is clearly an area that requires careful consideration in the clinical interview. Social phobia items appear to capture the core symptoms of the disorder in DSM-IV.

Specific sin 3: Failing to show that the short form measures each scale reliably. As Tables 4 and 5 indicate, both internal consistency and test-retest reliability were high for the CASI-PM-P.

Specific sin 4: Failing to show that the short form has adequate overlapping variance with the long form. Smith et al. (2000) note that it is important to demonstrate adequate overlapping variance between short and long forms. As indicated from the correlations reported in Table 7, the amount of variance in the long form accounted for by the short form is quite good, with r^2 values ranging from a low of 42.2% (ECI GAD scale) to 96% (ASI ODD scale).

Specific sin 5: Failing to show empirically that the short form reproduces the factor structure of a multifactorial instrument, and *Specific Sin 6: Failing to show that the short form preserves the content domains represented by the subfactors if the short form*

omits subfactors and preserves the overall factor. These two methodological problems are concerned with the factor structure of the short form and how it compares to that of the long form. This includes failing to show that the short form reproduces the factor structure of a multifactorial instrument, and that, when the factor structure is hierarchical, the short form retains both the broadband and narrowband factors. This problem exists when the original, long form exhibits a particular factor structure and the factor structure of the brief form is not assessed to determine if it is comparable. This problem may be particularly important in developing short forms of personality measures. It is less important for the development of the CASI-PM-P, however, because the DSM-IV itself was not empirically derived using factor-analytic procedures, and the Long Forms of the Symptom Inventories are keyed to the DSM-IV. As a result, replicating the factor structure of the CSI is not particularly relevant. More important than the factor structure of the CASI-PM-P and how it compares to that of the Long Forms are the correlations between the Long Forms and the CASI-PM-P, which were examined in the present study.

Specific sin 7: Failing to show that each factor in the short form has validity on an independent sample. Following procedures suggested by Smith et al. (2000), the concurrent validity of the CASI-PM-P was assessed, but further validity assessments should be conducted in independent samples if suitable instruments for comparison can be identified.

Specific sin 8: Failure to show that classification rates remain high with the short form. The intended purpose of the CASI-PM-P is for monitoring therapeutic progress. If it were to be used for screening or classification purposes, further research would be needed to show that it was equivalent to the Long Forms of the Symptom Inventories.

Specific sin 9: Failing to show the short form offers meaningful time savings. The CASI-PM-P requires approximately 5 minutes to complete, considerably less than the time needed to complete the Long Forms of the Symptom Inventories. Smith et al. (2002) indicate the loss of validity in using the short form can be calculated by comparing the validity coefficient of the long form and a criterion measure with that of the short form and that criterion measure. As Table 8 indicates, there are small losses in validity for social phobia (2%), ADHD-I (4%), ODD (9%), and no loss for other scales.

Other issues

An outcome instrument must, of course, be responsive to changes that occur during treatment (Vermeersch et al. 2000), and there are many different strategies for determining if an instrument is sensitive to treatment effects. One approach is to show that the instrument's change scores are correlated with behavioral improvement as measured with a more well-established instrument. Unfortunately, lacking a comparable broadband measure for monitoring overall change in children with multiple disorders, this approach was not feasible for assessing the CASI-PM-P at this time. However, a second, more common strategy is to demonstrate that the measure reflects within-subject change among individuals undergoing an effective intervention (Vermeersch et al. 2000), particularly when the effective treatment can be compared to a control condition. Although this can be done when the measure is being used in a RCT and a control group is compared to an active treatment, it is more difficult to do when the instrument is to be used in regular clinic settings and, for ethical and practical reasons, withholding treatment is problematic. As noted earlier, this situation has led researchers to seek alternative comparison groups. Following the suggestion of Weisz et al. (1987) and others, we examined differences in responsiveness to treatment of children whose families left treatment early versus those who continued in treatment. There may be limitations in using dropouts to assess overall effectiveness of a treatment, but with dropouts receiving only a small "dose" (i.e., number of sessions) of treatment and those remaining in treatment receiving a larger "dose," differences in responsiveness seemed useful in determining whether a new measure could detect change. As expected, the present study showed that those continuing in treatment showed greater changes on the CASI-PM-P than did early treatment terminators for the Total score and ADHD-Combined. In addition, as hypothesized, those continuing in treatment were more likely to exhibit a clinically significant change on the total score of the CASI-PM-P.

Limitations

Presently, the psychometric properties of the measure appear to be sound and comparable to, or better, than that of other broadband, brief measures for assessing treatment outcome. Nonetheless, establishing the psychometric properties of a behavior rating scale is a long and tedious process, and no single study can fully address all the questions that need to be resolved. Therefore, the present study is best considered to be an initial effort and much additional work remains to be done. For example, although responsiveness to change of the Total scale score and the ADHD-C scale for children with ADHD was demonstrated, sample sizes were not sufficient to assess the responsiveness to change of each of the subscales of the CASI-PM-P. Further confirmation of their responsiveness to change is needed. In addition, independent assessment of the functional impairment scale by comparison with other scales completed by parents, teachers, or both, would be useful. Moreover, the present study examined behavioral change in one clinic setting. Although the children who were evaluated and treated were administered a variety of different interventions from multiple therapists, how representative our findings are can only be determined after additional studies with different therapies, locales, and therapists are conducted. Also, poor agreement between different types of raters of child behavior problems (e.g., parents, teachers, clinicians) is well known (Achenbach et al. 1987), and additional research will be needed to develop measures suitable for use by such raters and to study cross-informant agreement between them. The problems of aggregating data from multiple sources are complex, so careful consideration will be needed to determine when data from multiple sources should be considered independently or aggregated (Holmbeck et al. 2002).

Future directions

In summary, the CASI-PM-P is a brief, broadband scale for monitoring change in the most salient symptoms of the more common childhood psychiatric disorders and assessing response to intervention in youths referred for mental health evaluations. At present, there is great interest in documenting

therapeutic progress in real-world clinical settings with reliable and valid assessment instruments. This, plus the increasing need for accountability in a financially stressed health-care system and growing awareness of high levels of co-morbidity that seemingly dictate the necessity for lengthy testing batteries, the development of a brief, parent-completed rating scale such as the CASI-PM-P seems timely. It is easy to score and requires only a few minutes for parents to complete, and a single-form for all age groups makes repeated administration easier to accomplish in busy clinic settings. The CASI-PM-P demonstrates good internal consistency, short-term test–retest reliability, and sensitivity to behavioral change during follow-up clinical care.

Although these preliminary findings for psychometric characteristics are encouraging, much additional research needs to be done. Of particular interest is its performance in clinical trials (both efficacy and effectiveness studies) and suitability for monitoring symptom change for a variety of different disorders. Unfortunately, these are not easily achieved objectives, owing in part to the extraordinary expense in conducting controlled trials and a general reluctance to add secondary measures to an assessment battery (i.e., concerns about response burden). Nevertheless, efforts are currently underway to examine these questions.

Disclosures

Drs. Sprafkin and Gadow are shareholders in Checkmate Plus, the publisher of the CASI-TM-P. Dr. Lavigne and Ms. Cromley are consultants to Checkmate Plus, the publisher of the CASI-TM-P.

Acknowledgments

We wish to thank Sarah Szwed Murray for her assistance with data analysis, and the anonymous reviewers for their helpful suggestions.

References

Achenbach TM: Manual for the Child Behavior Checklist/4–18 and 1991 Profile. Burlington (Vermont), University of Vermont Department of Psychiatry, 1991.

Achenbach TM, McConaughy SH, Howell CT: Child/adolescent behavioral and emotional problems: Implications of cross-informant correlation for situational specificity. Psychol Bull 101:213–232, 1987.

American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV). Washington (DC): American Psychiatric Association, 1994.

Bickman L: A measurement feedbck system (MFS) is necessary to improve mental health outcomes. J Am Acad Child Adolesc Psychiatry 47:1114–1119, 2008.

Birmaher B, Khetarpal SDB, Cully M, Balach L, Kaufman J Neer SM: The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale construction and psychometric characteristics. J Am Acad Child Adolesc Psychiatry 35:545–553, 1997.

Chambless DL, Sanderson WC, Shokam V, Bennett-Johnson S, Pope KS, Crits-Cristoph P, Baker M, Johnson B, Woody SR, Sue S, Beutler L, Williams DA, Mccurry S: Update on empirically validated therapies. Clin Psychol 49:5–18, 1996.

Cicchetti D, Volkmar F, Klin A, Showalter D: Diagnosing autism using ICD-10 criteria: A comparison of neural networks and standard multivariate procedures. Child Neuropsychol 1:26–37, 1995.

Conners CK, Siarenios G, Parker JD, Epstein JN: The Revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity. J Abnorm Child Psychol 26: 257–268, 1998.

Cromley T, Lavigne JV: Predictors and consequences of early gains in child psychotherapy. Psychotherapy: Theory, Research, Practice, Training 45:42–60, 2008.

Dunn TW, Burlingame GM, Walbridge M, Smith J, Crum MJ: Outcome assessment for children and adolescents. Clin Psychol Psychotherapy 12:388–401, 2005.

DuPaul GH, Anastopoulos AD, Power TJ, Reid R, Ikeda MJ, McGoey KE: Parent ratings of attention-deficit/hyperactivity disorder symptoms: Factor structure and normative data. J Psychopathol Behav Assess 20:83–102, 1998.

Gadow KD, Sprafkin J: Child Symptom Inventories Manual. Stony Brook (New York), Checkmate Plus, 1994.

Gadow KD, Sprafkin J: Adolescent Symptom Inventory-4 Screening Manual. Stony Brook (New York), Checkmate Plus, 1997a.

Gadow KD, Sprafkin J: Early Childhood Inventory-4 Norms Manual. Stony Brook, (New York), Checkmate Plus, 1997b.

Gadow KD, Sprafkin J: Adolescent Symptom Inventory-4 Norms Manual. Stony Brook, (New York), Checkmate Plus, 1998.

Gadow KD, Sprafkin J: Early Childhood Inventory-4 Screening Manual. Stony Brook (New York), Checkmate Plus, 2000.

Gadow KD, Sprafkin J: Child Symptom Inventory-4 Screening and Norms Manual. Stony Brook, (New York), Checkmate Plus, 2002.

Gadow KD, Sprafkin J: The Symptom Inventories: An annotated bibliography. From www.checkmateplus.com, 2006.

Garland A, Hurlburt M, Hawley K: Examining psychotherapy processes in a services research context. Clin Psychol: Science and Practice 13:30–46, 2006.

Goodman R: The Strengths and Difficulties Questionnaire: A research note. J Child Psychol Psychiatry 38:581–586, 1997.

Grayson P, Carlson GA: The utility of a DSM-III-R-based checklist in screening child psychiatric patients. J Am Acad Child Adolesc Psychiatry 30:669–673, 1991.

Hoagwood K, Jensen PS, Petti T, Burns BJ: Outcomes of mental health care for children and adolescents: I. A comprehensive conceptual model. J Am Acad Child Adolesc Psychiatry 35:1055–1063, 1996.

Holmbeck GN, Li ST, Schurmann JV, Friedman D, Coakley RM: Collecting and managing multisource and multimethod data in studies of pediatric populations. J Pediatr Psychol 27:5–18, 2002.

Jacobson N, Truax P: Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 59:12–19, 1991.

Jacobson N, Roberts L, Berns S, McGlinchey J: Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. J Consult Clin Psychol 67: 300–307, 1999.

Kolko D: Studying usual care in child and adolescent therapy: It's anything but routine. Clin Psychol: Science and Practice 13:47–52, 2006.

Levitt EE: Research on psychotherapy with children. In Handbook of Psychotherapy and Behavior Change. Edited by Bergin AE, Garfield SL. New York, Wiley, 1971, pp. 474–494.

McAdoo WG, Roeske NA: A comparison of defectors and continuers in a child guidance clinic. J Consult Clin Psychol 40:328–334, 1973.

Ogles BM, Melendez G, Davis DC, Lunnen KM: The Ohio Scales: Practical outcome assessment. J Child Fam Stud 10:199–212, 2001.

Sackett DL, Richardson WS, Rosenberg W, Haynes RB: Evidence-Based Medicine: How to Practice and Teach EBM. New York, Churchill Livingstone, 2000.

Smith GT, McCarthy DM., Anderson, KG: On the sins of short-form development. Psychol Assess 12:102–111, 2000.

Sprafkin J, Gadow KD: Is item randomization in psychiatric rating scales necessary? A comparison of the psychometric properties of two attention-deficity/hyperactivity disorder rating scale formats. J Child Adolesc Psychopharm 17:75–84, 2007.

Sprafkin J, Gadow KD, Salisbury H, Schneider J, Loney J: Further evidence of reliability and validity of the Child Symptom Inventory-4: Parent Checklist in clinically referred boys. J Clin Child Adolesc Psychol 31:513–524, 2002.

Vermeersch DA, Lambert MJ, Burlingame GM: Outcome Questionnaire: Item sensitivity to change. J Pers Assess 74:242–261, 2000.

Vermeersch DA, Whipple JL, Lambert MJ, Hawkins EJ, Burchfield CM, Iliishi JC: Outcome Questionnaire: Is it sensitive to changes in counseling center clients? J Couns Psychol 51:38–49, 2004.

Weiss B: Annotation: Routine monitoring of the effectiveness of child psychotherapy. J Child Psychol Psychiatry 39:943–950, 1998.

Weiss B, Catron T, Harris V, Phung T: The effectiveness of traditional child psychotherapy. J Consult Clin Psychol 67:82–94, 1999.

Weisz JR, Weiss B, Donenberg G: The lab versus the clinic: Effects of child and adolescent psychotherapy. Am Psychol 47:1578–1585, 1992.

Weisz JR, Weisz JR, Langmeyer DB: Giving up on child psychotherapy: Who drops out? J Consult Clin Psychol 55:916–918, 1987.

Weisz JR, Thurber C, Sweeney L, Proffitt VD, LeGagnoux GL: Brief treatment of mild-to-moderate child depression using primary and secondary control enhancement training. J Consult Clin Psychol 65:703–707, 1997.

Address reprint requests to:
John V. Lavigne, Ph.D.
Department of Child and Adolescent Psychiatry (#10)
Children's Memorial Hospital
2300 Children's Plaza
Chicago, IL 60614

E-mail: jlavigne@childrensmemorial.org