

Measures of Impairment for Children and Adolescents

Kay Hodges, Ph.D.
Jean Gust, B.S.

Abstract

Measuring the degree of impairment in children and adolescents due to emotional or behavioral disorders is a recent development that will become increasingly important to states and their mental health administrators. A survey of the 50 states was conducted to identify impairment measures that were used on a statewide basis. The results of the survey are presented; five measures were found to be used consistently. The three global measures and two multidimensional measures identified in the survey are reviewed here. Each measure is described and its pertinent psychometric data are presented. The implications for mental health administrators in terms of implementing statewide procedures to assess impairment are considered.

State administrators applying for mental health block grants, which are administered by the Center for Mental Health Services (CMHS), are now required to identify and estimate the incidence and prevalence of child and adult populations with serious emotional disturbance (SED). CMHS published the following definition for children with SED:

persons from birth up to age 18 who currently or at any time during the past year have had a diagnosable mental, behavioral, or emotional disorder of sufficient duration to meet diagnostic criteria specified within *DSM-III-R* (or the most recent edition of *DSM*) that resulted in functional impairment which substantially interferes with or limits the child's role or functioning in family, school, or community activities. (p. 29425)¹

Further, the states are required to operationally define functional impairment and to develop standardized methods for estimating SED. Recognizing the burden this would place on the states, CMHS commissioned several papers, one of which was to review the literature on operational measures of functional impairment and to survey the states to identify measures currently being used.² The purpose of this article is to (1) present the results of the state survey, (2) describe and critique each measure being used by the states, and (3) discuss implications for evaluating programs and systems of service delivery to children and adolescents.

These activities relevant to application for federal monies parallel changes that have been taking place in other sectors of mental health care. Increasingly, third-party payers are expecting mental health providers to document severity of impairment for the patient to qualify for more intensive or costly treatments. In addition, third-party payers are requiring continual monitoring of the client's level of functioning with an expectation that intensity of treatment will be reduced as impairment lessens. Justifying resource allocation by documenting degree of impairment appears to have become an expectation in all spheres of delivery of mental health services.

Address correspondence to Kay Hodges, Ph.D., Professor and Director of Clinical Training Eastern, Michigan University, Department of Psychology, 537 Mark Jefferson, Ypsilanti, MI 48197.

Jean Gust, B.S., is a graduate student in the Department of Psychology at Eastern Michigan University.

Traditionally, client records have focused on type of symptoms, usually summarized as a diagnosis, rather than on degree of impairment associated with the symptoms. Impairment reflects the consequences or effects of symptoms on functioning. In fact, epidemiological research over the past decade has demonstrated clearly that presence of a diagnosis is not comparable with impairment or need for treatment.³ These findings have in part been responsible for the inclusion of the concept of impairment in the most recent edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*,⁴ which was published in 1994. Presence of impairment is a required criterion for many *DSM-IV* diagnoses. However, impairment is not defined other than to specify that it should be in social, academic, or occupational areas of functioning or that it should be present in two or more settings.

A survey of state mental health agencies' use of a level of functioning measure for evaluating adults was reported previously by Pokorny.^{5,6} State mental health agencies were surveyed as part of the Mental Health Statistics Improvement Program, which was initiated in 1975 by the National Institute of Mental Health (NIMH). Pokorny⁶ found that six states were using the Global Assessment Scale (GAS),⁷ five were using Axis V of *DSM-III*,⁸ four were using a variant of a level of functioning scale developed by Carter and Newman (Carter-Newman LOF Scale),⁹ and four were using (or considering using) the Colorado Client Assessment Record (CCAR).¹⁰ No comparable survey has been reported for children and adolescents, thus justifying the need for a review at this time.

Method

The survey respondents were the state representatives to a national organization, the State Mental Health Representatives for Children and Youth (SMHRCY). Members of SMHRCY are appointed by the state mental health program directors. Although it varied by state, the SMHRCY contact person typically was either the director or coordinator of child and adolescent mental health for the state or the program director for the Child and Adolescent Service System Program. If the SMHRCY contact person was no longer in his or her role or was not the best contact person for the information sought, contact was made with the person or persons to whom the authors were referred.

In September and October 1993, letters were sent to the SMHRCY representative for each state. Information was requested about any instrumentation or measures of impairment used by various state agencies dealing with children and adolescents. A structured questionnaire was not used; rather, the states were asked to respond via written reply and to send any materials they thought might be relevant. The rationale given was that the CMHS was interested in developing resources for the states given the new requirement to provide information on prevalence of SED. Most states provided written replies and, when relevant, sent related administrative materials for the authors to examine, such as state plans for developing children's services, criteria for determining eligibility for services, and measures being used. Follow-up phone calls were made to those states from which no responses had been received.

Results of the State Survey

Because the purpose of the survey was to identify impairment instruments that could be used on a statewide basis, the authors restrict this review to measures that provide operational definitions of impaired functioning, yield at least one quantitative score, have published psychometric data, and are readily available to the public (i.e., not commercially owned). Measures that were designed to assess number of symptoms are excluded.

Information was received from 44 states; of these, 22 states identified themselves as not consistently using impairment measures. However, it is noteworthy that some of the states that were not consistently using measures at the time the survey was conducted had previously used measures on a statewide basis but were currently in periods of transition. The most frequent reason for

abandoning use of a measure was the perception that demand characteristics of the situation negatively influenced the accuracy of the ratings (e.g., ratings were highly influenced by perceptions about justifying one's own job or program). Two states were in the process of developing their own measures.

Sixteen states reported using impairment measures, as mentioned previously, and one of those states, Nevada, reported using two different impairment measures. The breakdown was as follows: two states (Montana and Nevada) were using the Carter-Newman LOF Scale, five states (Ohio, Oregon, Rhode Island, Utah, and West Virginia) were using the GAS or the Children's Global Assessment Scale (CGAS),¹¹ five states (Kentucky, Idaho, Nevada, North Dakota, and Texas) were using Axis V of the revised third edition of the *DSM (DSM-III-R)*,¹² one state (Colorado) was using the CCAR, and four states (Arizona, New Hampshire, North Carolina, and Wisconsin) were using the Child and Adolescent Functional Assessment Scale (CAFAS)^{13,14} (see Table 1). These measures are reviewed in detail after the criteria used in evaluating them are delineated.

Criteria Used in Evaluating the Measures

In the review of each of these five impairment measures, the authors briefly describe its characteristics and critique the measure. The criteria used to evaluate these measures are taken from criteria developed by a team of experts who wrote a widely publicized NIMH-sponsored monograph on outcome measures for use in mental health settings.¹⁵ A review of the literature did not yield any better set of criteria for the practical approach needed in applied settings. Specifically, the following criteria were used.

- The measure involves a simple, teachable methodology. Relevant issues include the following: the procedures can be implemented uniformly by staff members, the measure can be completed by front-line staff members, training materials exist that are accessible and well defined, and materials are readily available for use in establishing interrater reliability so that scores generated on the same client by different raters are comparable.
- The measure has objective referents. Ciarlo and colleagues¹⁵ state that the referents to functioning should be clear and objective. The present authors would add that the referents should be as behaviorally defined as possible.
- The degree to which the measure is free from respondent bias and nonreactivity to extraneous situational factors such as accountability pressure and staff or client expectations is considered. The authors would add that this factor is probably best conceptualized as on a continuum of likelihood because this may be a characteristic of all measures to some extent. The goal is to lessen the degree to which bias contributes to the given score.
- Evidence of reliability exists that can consist of agreement across raters (i.e., interrater reliability) or stability over time (i.e., test-retest reliability).
- Evidence of validity exists that could include the following: correspondence to other impairment measures, correspondence to measures assessing related variables such as number of symptoms, comparisons demonstrating that clients typically perceived as more impaired (e.g., inpatients, persons seeking treatment) score higher than do persons thought of as less impaired (e.g., nonreferred, outpatients), and correspondence between degree of impairment and service use (e.g., restrictiveness of treatment, cost or number of services).

Critique of Measures

Each measure is critiqued on the basis of these criteria, with the global measures presented first (i.e., Carter-Newman LOF Scale, GAS/CGAS, and Axis V), followed by the multidimensional scales (i.e., CCAR and CAFAS). The global measures are unidimensional, with the client being assigned one score that is meant to capture the client's functioning. By contrast, with multidimensional measures, the client is rated on multiple scales that define specific spheres of functioning. This order of presentation also coincides with the chronological order in which the measures were developed.

Table 1
Impairment Measures Used by States

	Global Measures		Multidimensional Measures	
Carter-Newman LOF	GAS/CGAS	Axis V	CCAR	CAFAS
Montana	Ohio	Kentucky	Colorado	Arizona
Nevada	Oregon	Idaho		New Hampshire
	Rhode Island	Nevada		North Carolina
	Utah	North Dakota		Wisconsin
	West Virginia	Texas		

Note: At the time of the survey, 6 states did not respond and 28 states reported that they were not consistently using impairment measures. LOF = level of functioning; GAS = Global Assessment Scale; CGAS = Children's Global Assessment Scale; CCAR = Colorado Client Assessment Record; CAFAS = Child and Adolescent Functional Assessment Scale.

Carter-Newman Level of Functioning Scale

For the purposes of this review, the level of functioning scale developed by Carter and Newman is referred to as the Carter-Newman LOF Scale.^{9,16} It was modeled after the Luborsky Health-Sickness Rating Scale.¹⁷ The Carter-Newman LOF Scale was intended to be a generic model of a level of functioning scale that could be used for a variety of programmatic and administrative purposes. This scale contains nine brief definitions of levels of functioning that range from dysfunctional in all areas to functioning well in all areas with no mental health services needed. The clinician rates the client's overall functioning level during the previous 24 hours for inpatients or the previous week for clients receiving less intensive services. In generating a score, the rater is asked to balance the relative contributions of four criteria: personal self-care, social functioning, vocational and/or educational functioning, and evidence of emotional stability and stress tolerance. Although these four areas are to be considered, the anchor points for the scale emphasize vocational functioning and ability to live in an ordinary social unit. This scale was intended to be used with clients that range from normal to incapacitated. The scale is designed for adults; if it is used with children, the rater is expected to make adjustments concerning personal care and functioning that appear warranted given the children's ages.¹⁵ However, because of the themes emphasized in the scale, it is not easily applicable to children.

The Carter-Newman LOF Scale offers a simple, teachable methodology in that the nine levels of functioning refer to observable behaviors. There are no known training materials. This scale could be viewed as vulnerable to respondent bias given that it yields a global score. However, if specific information is required to document the client's current vocational activities and living situation (e.g., residing in a supervised group home), then the potential for bias would be reduced.

No published reliability or validity data specific to the Carter-Newman LOF Scale were found by the authors. However, this would be expected given that the scale was intended to serve as a model. The reader may want to refer to Newman and his colleagues^{18,19} for a review of data relevant to using global measures for assessing adults. Other articles by Newman and colleagues^{20,21} present analyses of treatment episode costs across levels of functioning as examples of how level of functioning measures can be used to affect program development for mental health services.

Children's Global Assessment Scale

The CGAS, developed by Shaffer and colleagues,¹¹ is an adaptation of the GAS⁷ for adults, which was in turn an adaptation of Luborsky's Health-Sickness Rating Scale. The CGAS was designed to

reflect the child's lowest level of functioning during a specified period of time (e.g., previous month). The score can range from 1 to 100, with 10 anchor descriptions provided. The 10 anchor descriptions contain references to symptoms and examples of common clinical manifestations of various disorders. Lower scores reflect more impairment; scores of 71 or higher reflect functioning within the normal range.¹¹ Based on empirical data, Bird et al.³ proposed that a CGAS score below 61 and the presence of a diagnosis be used as the criteria to identify "cases" (i.e., children who are likely in need of services). The CGAS was intended to be used by highly trained clinicians such as psychiatrists, psychologists, and social workers.

The format of the CGAS is simple and straightforward, and it is assumed that it can be readily used by professionals who have had formal education and clinical experience in childhood psychiatric disorders. Consequently, specific training materials for the CGAS were not seen as necessary. The 10 anchor descriptions consist mainly of examples of behaviors or symptoms. Some of the referents in the scale are easily observable behaviors, whereas others are subjective symptoms or contain clinical terminology commonly used to summarize a set of symptoms or behaviors (e.g., frequent anxiety attacks). No definitions or descriptions are provided for the psychiatric references or terms. Like other measures yielding global scores, the CGAS is likely vulnerable to respondent bias.

Numerous psychometric studies have been conducted with the CGAS. Satisfactory interrater reliability has been demonstrated in studies in which the ratings were based on case histories^{11,22} or on interviews that were conducted in person or were recorded.²³⁻²⁶ In all of these studies, psychiatrists or other traditional clinicians (i.e., psychologists, social workers, graduate students in training) were used as raters and the source of information was a structured format (e.g., structured interviews or written reports). However, in an applied field study conducted by Herman,²⁷ mostly unsatisfactory reliability coefficients were reported. Satisfactory test-retest reliability was reported by Bird et al.²³

In terms of validity, several studies have demonstrated significant correlations with symptom scores from various measures including studies by Bird et al.²³ with the Child Behavior Checklist (CBCL)^{28,29} total problem score, by Shaffer et al.¹¹ with the Connors Abbreviated Checklist,³⁰ by Steinhausen²² with the Child Behavior Questionnaire,³¹ and by Vandvik³² with the total score from a structured diagnostic interview, the Child Assessment Schedule.³³ Although most of the correlations are moderate in magnitude, these results are supportive of the validity of the CGAS given that the other measures reflect mostly number of symptoms rather than impairment resulting from the child's symptomatology.

Evidence of contrast group validity was reported by Shaffer and colleagues¹¹ with a comparison of inpatients and outpatients and by Bird et al.²³ with a comparison of a clinical and a nonreferred sample.

The CGAS has satisfactory reliability and validity when used by professionals and when used in a situation in which there is minimal information variance (i.e., information on which the score is based is consistent across all raters). More research in applied settings is needed given that the only study conducted in an applied, nonresearch setting yielded mostly unsatisfactory reliability. A limitation in using the CGAS is the lack of formal training materials. A very recent development with the CGAS has been the generation of a version for nonclinicians.³⁴ This version of the CGAS has the same 10 anchor points but with somewhat simplified descriptors.

Axis V

With the publication of the *DSM-III*,⁸ a multiaxial system was incorporated in which Axis V was to be a rating of adaptive functioning. In the subsequent revisions of the *DSM* (i.e., *DSM-III-R*¹² and *DSM-IV*⁴), Axis V is also referred to as the Global Assessment of Functioning (GAF). The GAF is cited¹² as a revision of the GAS,⁷ which was derived from the Health-Sickness Rating Scale.¹⁷ In the *DSM-III-R* the scores range from 1 to 90, whereas in the *DSM IV* the range extends to 100. Higher scores reflect higher functioning. The time frame considered in generating a score is to be specified by the rater (e.g., current, highest level in past year, at discharge).

The instructions for Axis V state, "Consider psychological, social and occupational functioning on a hypothetical continuum of mental health-illness" (p. 32).⁴ The 10 anchor descriptions consist primarily of examples of symptoms with some examples of impaired functioning (e.g., no friends). Even though Axis V contains a few examples that are pertinent to children, it is oriented primarily toward assessing adult functioning.

There are no published reliability or validity data specific to using Axis V from the *DSM-III* or *DSM-IV* with children. However, this may be a moot issue because Axis V was derived from the GAS and CGAS. In regard to the other criteria, the points made in the critique already presented for the CGAS also apply to Axis V.

The remaining two measures, the CCAR and the CAFAS, are multidimensional measures and were developed as a result of the efforts of state mental health administrators.

Colorado Client Assessment Record

The CCAR was developed by the Colorado Division of Mental Health¹⁰ to assess the effectiveness of the state's mental health programs. It consists of two major components: a "personal problem profile" and a "level of functioning" measure.³⁵ The personal problem profile is essentially a checklist of 77 items (e.g., anxious) in which the clinician indicates whether the problem is true for the client with scoring options of *yes*, *no*, *don't know*, and *not applicable*. These problem items are organized into nine content domains: Feelings/Mood/Affect (anxiety, depression, mood), Thinking (problems with mental ability, thought disorders), Medical/Physical, Substance Use, Family (family problems, interpersonal), Role Performance (work/school, home/living arrangement), Sociolegal (antisocial, legal), and Self-Care/Basic Needs. Although the same nine content domains are referred to in both the personal problem profile and the measure, the information contained in the profile is not directly connected to the ratings on functioning.

Each of the nine level of functioning scales is scored using the following scale: above average functioning (1-9), average functioning (10-19), slight dysfunction (20-29), moderate dysfunction (30-39), severe dysfunction (40-49), and extreme dysfunction (50). The rater assigns a score that ranges from 1 to 50 for each scale. There is no total or summary score. On the CCAR itself, there are no definitions or anchor descriptions given. A separate document, the manual for the CCAR,³⁶ provides elaborate descriptions and examples for each scale. Examples are given for both adults and children. Ellis, Wackwitz, and Foster³⁵ state that the CCAR is well suited for use with individuals above the age of 14 years. The time period used in rating the CCAR is the prior 3 weeks.

The CCAR has a straightforward, teachable methodology. Good training materials are available for the CCAR in the form of the manual, which is quite elaborate with definitions given for commonly used psychiatric terms (e.g., anxious, hallucinations). However, no materials for establishing interrater reliability are provided. Many of the referents are behaviorally oriented, whereas others refer to subjective states or to summary descriptions of functioning (e.g., family unit functions cohesively with strong mutual support for its members). For some of the scales, the six anchor descriptions provided in the manual are general in form (e.g., occasional or mild difficulties in functioning, frequent difficulties, disabled or incapacitated). Even so, a specific example is typically given for at least three or four levels for each scale.

The fact that ratings are given for nine separate scales rather than one global score makes the CCAR less vulnerable to respondent bias. However, although the manual contains numerous examples, there are no referents specific to any of the scales appearing on the CCAR form itself. The generic description of levels already given (above average functioning, etc.) are applied to each of the scales. As a result, the rater does not have a frame of reference to which to refer in generating the score from 1 to 50. This characteristic likely increases respondent bias.

Despite currently being used extensively in Colorado and formerly in Louisiana,^{10,37} there is a paucity of studies addressing psychometric issues of the CCAR. No studies on reliability have been published. Two-factor analysis studies^{35,38} found three underlying factors that were very similar:

self-care, acting out/antisocial behavior, and emotional/psychological functioning. Other validity data are limited in that no studies have compared scores on the level of functioning scales to other measures or to service indicators. The major study published on the CCAR was conducted on derived typologies based on cluster analysis.³⁵ There have been no studies conducted on the CCAR with children, although the authors of the CCAR currently are working on a version for children.

Child and Adolescent Functional Assessment Scale

The CAFAS was developed by Hodges¹³ to assess degree of impairment in functioning for children and adolescents. It was modeled after the North Carolina Functional Assessment Scale,³⁹ which was developed by the state for the purposes of planning and prioritizing programs and fiscal resources. The CAFAS was originally developed to assist in evaluating the Fort Bragg Child and Adolescent Mental Health Demonstration Project.⁴⁰

The CAFAS is a multidimensional measure in which the youth is rated on five subscales: Role Performance (how effectively the youth fulfills societal roles), Thinking (ability of the youth to use rational thought processes), Behavior Toward Others/Self (appropriateness of the youth's daily behavior), Moods/Emotions (modulation of the youth's emotional life), and Substance Use (the youth's substance use and the extent to which it is inappropriate and disruptive). There are four rating categories (30 = *severe*, 20 = *moderate*, 10 = *mild*, and 0 = *average*). Each severity category contains numerous descriptions of behaviors, with a total of 97 anchor descriptions over the five scales. For each scale, the rater determines the severity level that best describes the youth. The five subscale scores are summed to derive a total score, which ranges from 0 to 150. The time period assessed can vary, typically ranging from 1 to 3 months. The CAFAS also has two subscales on which to rate the child's caregiver: Basic Needs and Family/Social Support. A telephone interview designed to obtain information for rating the CAFAS has been developed.

The CAFAS has a simple, teachable methodology. Training materials are extensive and include instructions, definitions of terms, 10 example vignettes with scoring and rationale, and 20 detailed reliability vignettes. Using the reliability vignettes, evaluators can determine whether staff members have achieved sufficient reliability to serve as raters. Most of the referents are behaviorally oriented (e.g., expelled from school). Some of the items refer to symptoms (e.g., depression or sadness); in these cases, the items typically focus on the behavioral effects associated with the symptoms (e.g., poor attendance at school).

The CAFAS is less vulnerable to respondent bias because the rater does not generate a global score but rather determines the appropriate severity level for each scale. The total score is then generated by summing over the scales. More importantly, the rationale for selecting a severity level must be provided by the rater. This is done by checking the item description in the severity level that is characteristic of the youth (e.g., currently confined for legal violations). In the event that none of the descriptors for any given scale is true for the youth, the rater is instructed to select the most appropriate severity level and write in the justification (e.g., self-mutilation behavior).

Satisfactory interrater reliability has been demonstrated with several different samples of raters: intake workers at a community mental health center, staff members from various agencies serving children, lay raters used in an evaluation study, and graduate and undergraduate students.^{14,41} In all of the reliability studies, written vignettes that presented information from interviews with the parents and the children were rated. Satisfactory interrater and test-retest reliability has also been demonstrated in a study in which trained research assistants rated referred and nonreferred children after administering the CAFAS telephone interviews to the parents.

Several types of validity have been demonstrated using the data from the evaluation of the Fort Bragg Demonstration Project.⁴² The CAFAS total score was highly correlated with CGAS scores and was moderately and significantly correlated with the CBCL total problem score and with the total score on a diagnostic interview, the CAS.³³ Contrast group validity revealed that inpatients were rated significantly more impaired than were outpatients on the CAFAS total score and on each

of the five CAFAS scales assessing the child. Regression analyses were also conducted to determine which measures administered at intake best predicted service use (i.e., cost and number of services received) over the subsequent 6 months. The CAFAS total score was a strong predictor compared to the other measures in the battery. The CAFAS has been revised recently to address problems identified by users. Currently, studies are under way to address the appropriateness of the CAFAS with African-American and Hispanic clients. Like all of the other measures, its reliability needs to be tested in the field with actual clients.

Summary of Comparison of Measures

All of the measures have a relatively simple format that is easily taught. However, the CGAS and its derivative, Axis V, assume that the rater is a highly trained professional. The scale has many referents to clinical terms and concepts. In terms of psychometric data, only the CGAS and the CAFAS have any evidence of reliability or validity of the functioning scores. For both the CGAS and the CAFAS, there is sufficient evidence of validity to justify their use. There is also considerable evidence of reliability for both measures; however, there is a need to conduct reliability studies in applied settings (i.e., in which personnel rate actual clients based on the typical clinical work-up at the facility).

A major liability of the global measures is their vulnerability to rater bias. Assigning a number from 1 to 100 (or 1 to 10) does not require the rater to consider the client's functioning across different areas or spheres in a disciplined fashion. Without such a process, positive or negative halo effects are much more likely to contribute to the score. This situation would likely exaggerate other biases such as accountability pressure that may result when treating clinicians serve as raters. In fact, five survey respondents spontaneously mentioned these factors as reasons for discontinuing the use of global measures. Some respondents reported frankly that staff members gave scores below or above a certain cut-off point depending on whether the client was being admitted or discharged. The raters apparently assumed that there was a connection between client scores and evaluation or funding of programs.

Implications for Mental Health Administrators

In the past, staff members who made the assumption that ratings of client impairment had some implications for resource allocation may have been accurately described as overly suspicious. However, at the present time and in the future, such assumptions may reflect reality. States have designed procedures for the purpose of prioritizing delivery of services so that resource allocation corresponds to intensity of need for services, such as in the Pioneer Project in the state of North Carolina.³⁹

Data collected on client impairment can be used to guide administrative and programmatic decision making as well as clinical treatment protocols for individuals. In fact, there is now an expectation that decisions such as these, which affect resource allocation, should be based on documented information. To assume that the data are accurate, it is important that the instruments used are psychometrically sound and that the specific conditions under which the instruments are used are likely to lead to collecting valid data.

As a result, measures for assessing impairment need to be more psychometrically sophisticated for evaluation information to be considered credible.⁴³ Global measures are desirable from the perspective of ease; however, the liability is a greater likelihood of rater biases. By contrast, multidimensional measures have an advantage in that the rater does not designate a summary score, although a total score could be derived by combining multiple scale scores.

However, it may be possible to use a global measure as long as the scoring is justified. For example, in the evaluation of the Fort Bragg Demonstration Project, the CGAS and the CAFAS were

to be used. However, because nonexpert interviewers were used, two problems with the CGAS were readily apparent: training materials were needed, and the lack of clinical professional background was a serious liability in generating an overall impression of functioning. The end result was that the CGAS was altered by adding four scales that describe levels of impairment for four separate areas of functioning (i.e., home, school, peers, and community) and by changing the wording of the CGAS scale so that scoring was based on the ratings given in the four separate areas (e.g., mild impairment in two areas or moderate impairment in one area).¹⁴ Experimentation with global measures for the purpose of lessening rater bias may generate other useful adaptations.

Choosing an appropriate instrument is easier than the remaining tasks, which include training, assessing reliability, and monitoring rater drift. Rater drift refers to the tendency of raters to unintentionally change over time the manner in which they apply the criteria definitions contained in a scale or measure. The state survey respondents reported common problems; typically, staff members were insufficiently trained or not trained at all, and interrater reliability was not assessed. Establishing interrater reliability is essential to assume that the data are meaningful. All of the raters need to use the same internal criteria or guidelines in making judgments. Even highly educated professionals need to demonstrate interrater reliability.

From a practical perspective, the training and the assessment of interrater reliability need to be efficient and effective. Typically, to assess reliability, trainees are asked to rate written vignettes related to the type of client with whom they will be working. Training could potentially be accomplished through the use of self-training manuals, with the trainee given sufficient time to complete the materials, or via structured seminars in which the trainee completes the reliability training during the seminar. Additional training may be necessary for trainees not achieving satisfactory reliability. A mechanism for routinely training new staff members needs to be in place. Periodic "refresher" training is necessary to guard against rater drift.⁴⁴ The term *judgmental relativism* has been used to describe the situation in which a person consistently is exposed to a subgroup with either lower or higher scores, resulting in that person's internal framework for scoring changing over time and making it less reliable.⁴⁵

Given the tasks involved in operationalizing training on a statewide basis, it may be more efficient to pay trained raters other than the treating staff. For example, the CAFAS telephone interview can be administered by raters who are independent of the treating staff to rate impairment of clients at intake and at subsequent follow-up times.

The state mental health systems need to be an integral part of efforts to develop adequate measures of impairment for children. In fact, some of the most robust measures have been derived from the efforts of state personnel (e.g., Carter-Newman LOF Scale, CCAR, CAFAS). Additionally, states have an obligation to ensure that any measures used to assess impairment do not place minority or cultural subgroups at a disadvantage or unfairly identify them as impaired. Thus far, there are no data on the cultural appropriateness of the various measures. State agencies need to be involved in developing feasible training strategies and in assessing whether the instruments can be used reliably by front-line staff members rating actual clients.

Acknowledgments

The authors express special thanks to Barbara J. Burns and Krista Kutash for their critical review of this article and to Kyle Bonesteel and Sarah Ring-Kurtz who helped with the task in their role as graduate assistants. We also express appreciation to the various state contact persons who assisted us in the survey and to the authors of the measures who generously shared information. Portions of this article were prepared under a contract from the Center for Mental Health Services and were presented at a conference on the Development of Standardized Methodology for Estimates of the Incidence and Prevalence of Serious Emotional Disturbance Among Children and Adolescents, Bethesda, Maryland, May 1994.

References

1. *Federal Register*, May 20, 1993, pp. 29422-29425.
2. Hodges K: Measures for Assessing Impairment in Children and Adolescents. Paper presented at a meeting sponsored by the Center for Mental Health Services on the Development of Standardized Methodology for Estimates of the Incidence and Prevalence of Serious Emotional Disturbance Among Children and Adolescents, Bethesda, MD, May 1994.
3. Bird HR, Yager TJ, Staghezza B, et al.: Impairment in the epidemiological measurement of childhood psychopathology in the community. *Journal of the American Academy of Child and Adolescent Psychiatry* 1990; 29:796-803.
4. *Diagnostic and Statistical Manual of Mental Disorders*. Fourth ed. Washington, DC: American Psychiatric Association, 1994.
5. Pokorny LJ: *Feasibility Assessment: The Inclusion of a Level of Functioning Measure in the Minimum Data Set of the Mental Health Statistics Improvement Program*. Final (unpublished) report prepared under NIMH Contract 85M042256401D. 1986.
6. Pokorny LJ: A summary measure of client level of functioning: Progress and challenges for use within mental health agencies. *Journal of Mental Health Administration* 1991; 18:80-87.
7. Endicott J, Spitzer RL, Fleiss JL, et al.: The Global Assessment Scale. *Archives of General Psychiatry* 1976; 33:766-771.
8. *Diagnostic and Statistical Manual of Mental Disorders*. Third ed. Washington, DC: American Psychiatric Association, 1980.
9. Carter DE, Newman FL: *A Client-Oriented System of Mental Health Service Delivery and Program Management: A Workbook and Guide*. Rockville, MD: National Institute of Mental Health, 1976.
10. Ellis RH, Wilson NZ, Foster FM: Statewide treatment outcome assessment in Colorado: The Colorado Client Assessment Record (CCAR). *Community Mental Health Journal* 1984; 20:72-89.
11. Shaffer D, Gould MS, Brasic J, et al.: A Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry* 1983; 40:1228-1231.
12. *Diagnostic and Statistical Manual of Mental Disorders*. Third ed., rev. Washington, DC: American Psychiatric Association, 1987.
13. Hodges K: Manual for the Child and Adolescent Functional Assessment Scale. Unpublished manuscript, Department of Psychology, Eastern Michigan University, 1990.
14. Hodges K, Bickman L, Kurtz S, et al.: Multidimensional measure of level of functioning for children and adolescents. In: Algarian A, Freidman R (Eds.): *A System of Care for Children's Mental Health: Expanding the Research Base*. Tampa, FL: Research and Training Center for Children's Mental Health, 1991, pp. 149-154.
15. Ciarlo JA, Brown TR, Edwards DW, et al.: *Assessing Mental Health Treatment Outcome Measurement Techniques*. National Institute of Mental Health, Series FN, No. 9. DHHS Pub. No. (ADM) 86-1301. Washington, DC: U.S. Govt. Printing Office, 1986.
16. Newman FL, Sorensen JE: *Integrated Clinical and Fiscal Management in Mental Health: A Guidebook*. Norwood, NJ: Ablex Publishing Corporation, 1985.
17. Luborsky L: Clinicians' judgements of mental health: A proposed scale. *Archives of General Psychiatry* 1962; 7:35-45.
18. Newman FL: Global scales: Strengths, uses and problems of global scales as an evaluation instrument. *Evaluation and Program Planning* 1980; 3:257-268.
19. Newman FL, Hunter RH, Irving, D: Simple measures of progress and outcome in the evaluation of mental health services. *Evaluation and Program Planning* 1987; 10:209-218.
20. Newman FL, Burwell BA, Underhill WR: Program analysis using the client oriented cost outcome system. *Journal of Evaluation and Program Planning* 1978; 1:19-30.
21. Newman FL, Rinkus A: Level of functioning, clinical judgment, and mental health service evaluation. *Evaluation and the Health Professions* 1978; 1:175-194.
22. Steinhilber H: Global assessment of child psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry* 1987; 26:203-206.
23. Bird H, Canino G, Rubio-Stipec M, et al.: Further measures of the psychometric properties of the Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry* 1987; 44:821-824.
24. Garber J, Kriss MR, Koch M, et al.: Recurrent depression in adolescents: A follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry* 1988; 27:49-54.
25. Apter A, Orvaschel H, Laseg M, et al.: Psychometric properties of the K-SADS-P in an Israeli adolescent inpatient population. *Journal of the American Academy of Child and Adolescent Psychiatry* 1989; 28:61-65.
26. Fox SJ, Barnett RJ, Davies M, et al.: Psychopathology and developmental delay in homeless children: A pilot study. *Journal of the American Academy of Child and Adolescent Psychiatry* 1990; 29:732-735.
27. Herman SE: *Reliability Study: Children's Global Assessment Scale*. Lansing, MI: Department of Mental Health, 1983.
28. Achenbach T, Edelbrock C: *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington, VT: Queen City Printers, 1983.
29. Achenbach T: *Manual Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry, 1991.

30. Goyette CH, Connors CK, Ulrich RF: Normative data on revised Connors' parent and teacher rating scales. *Journal of Abnormal Child Psychology* 1978; 6:221-236.
31. Rutter M, Tizard J, Whitmore K: *Education, Health and Behaviour*. London: Longman, 1970.
32. Vandvik IH: Mental health and psychosocial functioning in children with recent onset of rheumatic disease. *Journal of Child Psychology and Psychiatry* 1990; 31:961-971.
33. Hodges K, Kline J, Stern L, et al.: The development of a child assessment interview for research and clinical use. *Journal of Abnormal Child Psychology* 1982; 10:173-189.
34. Bird HR, Schwab-Stone M, Andrews H, et al.: Global Measures of Impairment for Epidemiologic and Clinical Use With Children and Adolescents. Unpublished manuscript, 1994.
35. Ellis RH, Wackwitz JH, Foster M: Uses of an empirically derived client typology based on level of functioning: Twelve years of the CCAR. *Journal of Mental Health Administration* 1991; 18:88-100.
36. Colorado Division of Mental Health: *User's Manual for the Colorado Client Data Forms: Client Admission Form, Part A (PES-7A), Colorado Client Assessment Record—CCAR (PES-7B and PES-10)*. Denver: Colorado Division of Mental Health, 1991.
37. Shern D, Dilts SL: The concentration of severely disturbed CMI in a core urban area. *Administration in Mental Health* 1987; 15:18-28.
38. Stahler GJ: The factor structure of the Colorado level of functioning measure. *Journal of Clinical Psychology* 1987; 43:696-698.
39. North Carolina Department of Human Resources: *Pioneer Funding System: Operating Manual*. Vol. 3: *Level of Eligibility*. Raleigh, NC: Department of Human Resources, Division of Mental Health, Mental Retardation and Substance Abuse Services, 1989.
40. Bickman L, Heflinger CA, Pion G, et al.: Evaluation planning for an innovative children's mental health system. *Clinical Psychology Review* 1992; 12:853-865.
41. Hodges VK, Watson L, Peregord J: Interrater Reliability of a Multidimensional Measure to Assess Impairment: The Child and Adolescent Functional Assessment Scale. Unpublished manuscript, 1994.
42. Hodges K, Lambert W, Summerfelt WT: Validity of a Measure to Assess Impairment: The Child and Adolescent Functional Assessment Scale (CAFAS). Paper presented at the seventh annual research conference, A System of Care for Children's Mental Health: Expanding the Research Base, Tampa, FL, March 1994.
43. Moskowitz JM: Why reports of outcome evaluations are often biased or uninterpretable. *Evaluation and Program Planning* 1993; 16:1-9.
44. Kazdin AE: Artifact, bias, and complexity of assessment: The ABCs of reliability. *Journal of Applied Behavior Analysis* 1977; 10:141-150.
45. Rich BE, Paul GL, Mariotto MJ: Judgmental relativism as a validity threat to standardized psychiatric rating scales. *Journal of Psychopathology and Behavioral Assessment* 1988; 10:241-257.