

THE OHIO YOUTH PROBLEM, FUNCTIONING,
AND SATISFACTION SCALES

TECHNICAL MANUAL



Benjamin M. Ogles, Ph. D. Gregorio Melendez, M. S.
Diane C. Davis, M. S., and Kirk M. Lunnen, Ph. D.^{1,2}

Ohio University

March 2000

¹ Portions of this project were funded by the Office of Program Evaluation and Research, The Ohio Department of Mental Health, Grant # 96-1105

² This project was also supported by the Southern Consortium for Children.

TABLE OF CONTENTS

TABLES.....	III
FIGURES.....	IV
EXECUTIVE SUMMARY.....	V
INTRODUCTION.....	1
INITIAL CONCEPTUALIZATION	3
USE OF A THEORETICAL AND CONCEPTUAL SCHEME	3
THE INPUT OF STAKEHOLDERS.....	4
RESEARCH INPUT.....	5
SERVICE PROVISION FOR MARGINALIZED POPULATIONS.....	6
SUMMARY OF CONCEPTUALIZATION	6
INSTRUMENT DEVELOPMENT	9
CONTENT AREAS	9
ITEM DEVELOPMENT.....	9
SHORT FORM OF THE OHIO SCALES	10
ITEM DESCRIPTIONS.....	10
ADMINISTRATION AND SCORING	12
PROBLEM SEVERITY	12
FUNCTIONING.....	12
HOPEFULNESS	13
SATISFACTION.....	13
RESTRICTIVENESS OF LIVING ENVIRONMENTS SCALE (ROLES).....	14
PSYCHOMETRIC PROPERTIES - ORIGINAL OHIO SCALES	16
PROCEDURES	17
RELIABILITY	19
VALIDITY	23
SENSITIVITY TO CHANGE.....	31
PSYCHOMETRIC PROPERTIES OF THE OHIO SCALES - SHORT FORM.....	40
PROCEDURES	41
RELIABILITY	42
VALIDITY	43
SUMMARY.....	45
CONCLUSION	46

TABLES

Table 1. ROLES' Weights	14
Table 2. Means and Standard Deviations on the Original Ohio Scales for the different samples.....	19
Table 3. Internal Consistency Estimates (Cronbach's Alpha) for each Scale on the Three Instruments for Community and Clinical Samples.	20
Table 4. Test-Retest Reliability Estimates for the Parent and Youth Rated Instruments.....	21
Table 5. Inter-rater Reliability for Four Measures of Functioning for Three Rater Groups across Methods of Presentation.....	22
Table 6. Inter-rater Reliability for Four Measures of Functioning using Vignettes and Clinical Folders.....	22
Table 7. Correlations Among Agency Worker Rated Measures in Sample #4.....	24
Table 8. Correlations Among Four Measures of Functioning Rated by Graduates, Undergraduates, and Case Managers in Sample #6	24
Table 9. Means and Standard Deviations on the Ohio Scales for clinical and community samples rated by the case manager.....	25
Table 10. Concurrent Validity Estimates for the Parent Rated Ohio Scales	25
Table 11. Means and Standard Deviations on Parent Ratings of Problem Severity and Functioning.....	26
Table 12. Factor Loadings on the Parent Rated Hopefulness Scale.....	27
Table 13. Factor Loadings for the Parent Rated Problem Severity Scale	28
Table 14. Factor Loadings on the Parent Rated Functioning Scale.....	28
Table 15. Concurrent Validity Estimates for the Youth Rated Ohio Scales	29
Table 16. Means and Standard Deviations on Youth Ratings of Problem Severity and Functioning.....	30
Table 17. Sensitivity to change estimates for the Agency Worker Rated Ohio Scales.....	31
Table 18. Number of Individuals Completing the Follow-up Ratings.....	32
Table 19. Means, Standard Deviations, and Significance Tests for Three Sources of Information in Three Content Areas from Intake to 3 month Assessment.....	32
Table 20. Hierarchical Modeling of Change in Youth Self-Report Functioning	37
Table 21. Means and Standard Deviations on the Short Form of the Ohio Scales for the different samples.	42
Table 22. Internal Consistency Estimates (Cronbach's Alpha) for each Scale on the Short Form for Community and Clinical Samples.....	42
Table 23. Correlations Between the Agency Worker Rated Short Form and Original Ohio Scales.....	43
Table 24. Correlation Coefficients for the Original and Short Forms of the Parent Rated Ohio Scales.....	43
Table 25. Comparison of Case Manager Ratings of Minority and Majority Youth.....	44
Table 26. Comparison of Parent Ratings of Minority and Majority Youth	44
Table 27. Comparison of Minority and Majority Youth Self-Report Ratings	44

FIGURES

Figure 1. Categories of Outcome Measurement for Four Dimensions.	4
Figure 2. Change in Problem Severity by Duration in Treatment Rated by the Community Support Worker	33
Figure 3. Change in Problem Severity by Duration in Treatment Rated by the Parent.....	34
Figure 4. Change in Functioning by Duration in Treatment Rated by the Community Support Worker	35
Figure 5. Change in Functioning by Duration in Treatment Rated by the Parent	36
Figure 6. Modeled Change in Youth Self-report Functioning	38

EXECUTIVE SUMMARY

As the service system for children and adolescents with emotional and behavioral problems has evolved, additional emphasis has been placed on developing ongoing evaluation procedures to determine the effectiveness of community-based interventions. Similarly, behavioral health care providers (in both the public and private sectors) are more often required to collect information regarding the effectiveness of services as a part of health care reform and an increased focus on accountability. With this emphasis on outcome assessment, many providers and administrators are searching for outcome measures. Typically, administrators hope to find measures that are both practical and scientifically sound. With this goal in mind – practical yet empirical – we developed the Ohio Youth Problem, Functioning and Satisfaction Scales (Ohio Scales).

This manual provides a detailed description of the the background, conceptualization, and psychometric properties of the Ohio Scales. This manual is a Technical Manual designed to provide an in-depth description of the theoretical foundation for the Ohio Scales along with the nuances regarding reliability, validity, and sensitivity to change. A more user-friendly, practical manual (User's Manual) is available that briefly describes the conceptualization of the Ohio Scales along with instructions for administration, scoring, and interpretation. For additional information regarding the Ohio Scales, readers may contact the first author at (740) 593-1077 or ogles@ohio.edu. Questions can also be addressed by the Office of Program Evaluation and Research, Ohio Department of Mental Health at (614) 466-8651.

INTRODUCTION

Everywhere in the service sector one hears the cry of outcomes! Across a broad range of industries and services, increasing emphasis is being placed on responsibility and accountability for the end product or outcome of services. Education, health care, and behavioral health care are especially influenced by the increasing focus on outcome. There are outcome task forces within states, credentialing bodies, associations, and organizations. Numerous articles and books are written that make recommendations regarding when, where, who, and how to assess the outcome of psychosocial and medical interventions (e.g., Ogles, Lambert, & Masters, 1996; Sederer & Dickey, 1996). Payers desire quality outcomes. Consumers deserve good outcomes. Providers want to show that they produce quality outcomes. Outcome is the topic of the season.

Especially with the advent of managed care and the privatization of public services, the collection of outcome data is becoming an increasingly important method of accounting for the expenditure of funds. Both public and private funders of behavioral health services want evidence that the behavioral health interventions they fund are effective. Outcome data are one of the primary avenues for demonstrating effective interventions. Unfortunately, the term "outcome" is often used as a "buzzword" rather than as a specific descriptor of certain scientific methodologies. Just deciding what the word "outcome" means is a difficult beginning. Additionally, once the goal of assessing outcome has been established, there are difficulties identifying, selecting, measuring, and reporting useful data that indicate whether the outcomes have been achieved. Who should report the outcome? What content area should be assessed? How often should we collect this outcome data? These and numerous other questions must be answered.

Unfortunately the answers to questions about assessing outcome vary widely depending on the service, the location, the clientele, and other situation specific circumstances. Decisions about outcome assessment for the outpatient treatment of adults may not apply to services for individuals with chronic mental disorders such as schizophrenia. Similarly, children who receive mental health services will need a unique set of methodologies and measures for evaluating the outcomes of service (Burns & Friedman, 1988).

The assessment of outcome within children's behavioral health services can be especially challenging. Because the development of outcome assessment tools for children's behavioral health services lags behind the efforts for adults (Weber, 1998), there is a paucity of quality measures. Children's outcome assessment also requires data from multiple sources (e.g., parents, youth, agency worker, and teacher). Especially when examining the effectiveness of services for youth with severe emotional disturbances, the involvement of multiple child-serving systems can complicate the assessment of outcome (Burchard & Shaefer, 1992).

Assessing outcome is a challenging task for researchers and perhaps overwhelming for administrators and individuals who provide services. For example,

many community mental health providers find that the format of research-based measurement tools is impractical. These research-based tools may be lengthy, difficult to score and interpret, or costly. As a result, some organizations throw together a few items that assess satisfaction and make their own "outcome" measure. These agencies acknowledge the importance of assessing outcome yet desire methods of evaluating services using cost-efficient, practical measures.

A number of conflicting tensions also influence the assessment of outcome and the demands placed on instruments used for assessment are often unrealistic. For example, some would like an instrument that can screen for serious issues at intake (e.g., self-harm, drug or alcohol use), provide information regarding a broad range of potential problems (diagnostic symptoms), and provide pretreatment data for later comparison with post treatment data (outcome). At the same time, users require that the instruments are short, easy to understand, easily scored, cheap, and psychometrically rigorous. The natural tensions that evolve from the multiple competing uses and characteristics of outcome instruments influence what, who, how and when to measure. As a result, the development of meaningful measures of outcome is complicated by the many competing demands of the end user.

Within this climate of demand for outcome measures and considering the need for pragmatic, child friendly measures, we set out to develop measures of clinical outcome for youth that receive behavioral health services (The Ohio Scales). The goal was to develop outcome measures that could be practical (e.g., easily administered, scored, and interpreted) while still meeting stringent psychometric and research criteria. The target population for the instruments is children who have severe emotional and behavioral problems ages 5 to 18. These youth are more likely to be involved with multiple child-serving systems and tend to receive a longer duration of intervention. As a result, there is a need for instruments that can be administered at predetermined intervals to evaluate ongoing progress.

The remaining portions of this manual describe the conceptualization and initial development of the Ohio Scales, the scoring and administration procedures, and the current psychometric data regarding reliability, validity, and sensitivity to change. This manual presents the "nuts and bolts" details of the scale construction. A more practical manual (User's Manual) is available for the front-line user of the Ohio Scales that limits the presentation of information to practical administration, scoring, and interpretation issues.

Data presented in this manual suggest that the instruments are reliable, valid, and sensitive to change. As with any scale, however, the validation of the instrument is never complete. Nevertheless, data collected to date support the application of the Ohio Scales as outcome instruments in services for children and adolescents.

INITIAL CONCEPTUALIZATION

As part of the conceptualization process, four areas of concern were considered relevant to the assessment of clinical outcomes for children with severe emotional and behavioral disorders:

- 1) a theoretical and conceptual scheme of outcome;
- 2) the perspective of various stakeholders (both directly or indirectly affiliated with children's mental health services);
- 3) research concerning the effectiveness of mental health treatment for children with specific emphasis on current methods of outcome measurement; and
- 4) the problems associated with service provision and assessment in resource deprived areas and underserved populations.

Use of a Theoretical and Conceptual Scheme

Because of the numerous processes that occur during mental health intervention, divergent methods of measurement have been used as a way of capturing the complexity of human functioning and change. However, when multiple assessment methods are used, how should one go about choosing the most appropriate outcome measures? To a large degree researchers are bound by practical constraints. The theoretically ideal battery of instruments for a given study is usually limited by pragmatic considerations such as time, money, and client comfort. Yet, an ideal scheme may give purpose and direction to the selection of a final assessment package (Ogles, Lambert, & Masters, 1996). Such a conceptual scheme is presented in Figure 1 (Lambert & Hill, 1994; Lambert, Ogles, & Masters, 1992).

The conceptual scheme includes four theoretical dimensions upon which outcome instruments vary: 1) the content area addressed by the instrument, 2) the source of outcome ratings sampled by the instrument, 3) the outcome instrument's method or technology of data collection, 4) and the time orientation or stability of the instrument. Each dimension along with sub-dimensions is depicted in Figure 1. The numbers represent potential instruments or subcategories of the specified dimension that are not enumerated. For the Ohio Scales, we used the scheme as the basic underlying model for conceptualizing the important sources, contents, and methods for collecting outcome data.

Figure 1. Categories of Outcome Measurement for Four Dimensions.

Content	Source	Technology	Time
Intrapersonal	Self-report	Evaluation	Trait
affect	1	1	1
1	2	2	2
2	•	•	•
•	Therapist Rating	Description	State
behavior	1	1	1
1	2	2	2
2	•	•	•
•	Trained Observer	Observation	Pattern
Cognition	1	1	1
1	2	2	2
2	•	•	•
•	Relevant Other	Status	•
Interpersonal	1	1	
1	2	2	
2	•	•	
•	Institutional	•	
Social Role	1		
1	2		
2	•		
•	•		

Note: The numbers represent potential instruments or subcategories of the specified dimension that are not enumerated.

The Input of Stakeholders

Strupp and Hadley (1977) proposed a tripartite model of mental health outcomes in which they suggested that three interested parties are concerned with the outcome of mental health interventions: society, the consumer, and the mental health professional. Based on the viewpoint of the interested party different criteria are selected to measure successful treatment. Certainly, one’s perspective plays a role in determining what one values as successful intervention. As a result, we attempted to gain input from a variety of “stakeholders” (Gold, 1983) in order to assess success from several perspectives.

More specifically, a Social Validation Survey of the various stakeholders was conducted (Gillespie, 1993) in order to get their input into what they found important and to what degree they were satisfied with certain aspects of services and potential outcomes.

This approach evolves from a body of behavioral and social validation research that first made the case for subjective measurement of behavioral interventions (Kazdin, 1977; Schriener & Fawcett, 1988; Wolf, 1978). The Social Validation Survey instrument used in this project was developed in Pennsylvania by VanDenBerg (1992) and was originally based on the work of Wolf (1978). The instrument was obtained and the survey was conducted with slight changes based on an item analysis of the original data (Gillespie, 1993). The revised survey was then administered in rural, southeastern Ohio. Stakeholders were asked a series of questions regarding the importance and satisfaction levels associated with various service issues.

One hundred and ninety-two stakeholders of child and family services were selected for participation in our survey. In all, 95 responses were received from a variety of stakeholders (e.g., children, parents, judges, mental health professionals, social service professionals, influential community members, etc.). While the details of this master's thesis (Gillespie, 1993) are too lengthy to include in the manual, the overall goal was to identify issues that stakeholders deemed most important but with which they were least satisfied and then to include these issues within the instruments. For example, the item they considered most important but were least satisfied with involved the youth "learning to not be aggressive and to not harm others;" consequently, items tapping these tendencies were included in the instruments.

Research Input

In addition to obtaining input from the various individuals both directly and indirectly involved with children's mental health services, we identified and examined several recent studies investigating the effectiveness of mental health services for children and youth (e.g., Bickman et al., 1995; Duchnowski, Johnson, Hall, Kutash, & Friedman, 1990; Evans, Dollard, Huz, & Rahn, 1990; Kutash, Duchnowski, Johnson, & Rugs, 1993; Stroul & Friedman, 1986). This review focused on the instruments used to evaluate outcome and identified areas of outcome thought important to assess. For example, Duchnowski, Johnson, Hall, Kutash, and Friedman (1990) describe their multi-source, multi-method data collection strategy which included assessment instruments from five domains: 1) demographic data, 2) a history of services received, 3) family characteristics and functioning, 4) emotional and behavioral problems and competence, and 5) academic achievement (including IQ). A variety of well-established instruments were selected to assess various aspects of these domains in order to "obtain an ecological overview of the youth and their families" (p. 18; Duchnowski, Johnson, Hall, Kutash, & Friedman, 1990). While the focus of this project did not include all areas of assessment, reviewing several well-designed studies helped to ascertain the most important domains of assessment to include in an initial outcome instrument.

Service Provision for Marginalized Populations

The initial development of the Ohio Scales occurred in rural southeastern Ohio. The rural nature of services presents some unique problems for both the provision of services and the development of an evaluation program. Nearly 25% of the individuals within a ten county area have incomes that fall below the federal poverty guidelines. The rural nature of the counties also limits financial resources and results in large distances between agencies. Similarly, there is limited availability of many medical and mental health services. For example, in some counties, only one or two case managers provide services, and because of geographic and practical limitations, training and communication with other agencies is infrequent. In addition, needed services are often not available in smaller communities resulting in placements that may isolate the family from the child. These difficulties influence both the provision of services and the assessment of outcome.

The problems encountered in southeastern Ohio are not unique to rural areas. In fact, when serving at-risk populations many of the issues are identical irrespective of geographic location (e.g., poverty, transportation, availability of services). As a result, when developing the Ohio Scales issues that might preclude adequate application in areas with limited resources were carefully considered.

Summary of Conceptualization

Based on our consideration of assessment in resource deprived settings, input from stakeholders, review of current studies, and using a conceptual scheme of outcome assessment, a list of desirable characteristics for the initial assessment of clinical outcomes was developed.

1. Measurement instruments need to be pragmatic in terms of time, expense, and clinical utility. The practical constraints of service provision must be considered when developing useful instruments. Because many instruments are difficult to score or require large amounts of the client's time to complete, they are not used on an ongoing basis to provide feedback regarding program effectiveness. While they may be used on occasion for specific projects, their cumbersome nature makes them impractical for ongoing use (Rosenberg, 1979).
2. Current mental health care practices require increased involvement of paraprofessionals in assessment. This necessitates measures that require minimal professional training for interpretation. Similarly, instruments are needed that provide immediate and understandable results for parents and children receiving services.

3. Effective assessment devices should include input from multiple sources (VanDenBerg, Beck, & Pierce, 1992; Lambert, Christensen, & DeJulio, 1983; Ogles, Lambert, & Masters, 1996). Information from specific youth as well as their parents and case manager can provide a more comprehensive clinical picture. In addition, the different sources of input provide an index of the authenticity of the youth's self-report information. Multiple sources are also important given the growing emphasis on consumer satisfaction with treatment and involvement of parents and children in the treatment planning process (Barth, 1986; Friesen, Koren, & Koroloff, 1992).

4. Multiple content areas of outcome should be considered. Potential content areas included: overall well-being or hopefulness, severity of problems, life functioning, satisfaction with services, family functioning, restrictiveness of living setting, school performance, etc. Including multiple content areas allows for the development of individual profiles necessary for individualized treatment planning. In addition, the assessment of multiple content areas helps to identify areas of change for youth who have multiple and severe problems. The assessment of client and family strengths is an area that may be especially useful (Burchard & Clark, 1990; Cochran, 1987; Dunst, Trivette, & Deal, 1988; Friesen & Koroloff, 1990; Poertner & Ronnau, 1992). Unfortunately, many existing measures focus on the child's psychopathology while excluding their strengths. With many new programs that focus on developing individualized plans of intervention or "wrap-around" services, the child's strengths within his or her social context should be considered (Friesen & Koroloff, 1990; Burchard & Clark, 1990; Cochran, 1987; Dunst, Trivette, & Deal, 1988).

5. Any measurement instruments should be psychometrically sound. While an emphasis on pragmatics is necessary, this emphasis should be counterbalanced by the need to develop instruments with demonstrated psychometric properties. Many attempts to demonstrate program success originating with service providers rely upon homemade surveys with questionable reliability and validity. At the same time, the development of brief, practical, and usable instruments does not rule out the possibility of using psychometrically rigorous methods of test development. More specifically, evidence of test-retest reliability, inter-rater reliability, or internal consistency (used respectively as appropriate) is needed to establish the instrument's reliability. Similarly, adequate evidence should be provided to demonstrate the validity of the measures. Finally, with the current emphasis on outcome, it is of particular importance that the instruments demonstrate sensitivity to change (Kutash, Duchnowski, Johnson, & Rugs, 1993).

Based on this list of desirable characteristics for outcome assessment instruments, we began the process of developing practical measures of clinical outcome that could

cover multiple content areas and provide input from multiple sources while attempting to maintain a level of psychometric integrity. Our final goal was a practical set of instruments that would be useful for agencies and practitioners without the hassles of many research based instruments (e.g., lengthy, difficult scoring, difficult to interpret, costly, time consuming).

INSTRUMENT DEVELOPMENT

With this background, the Ohio Youth Problem, Functioning, and Satisfaction Scales (Ohio Scales) were developed (Ogles, Lunnen, Gillespie, & Trout, 1996). Three parallel forms of the Ohio Scales were developed for completion by the youth's parent or primary caregiver (P - form), the youth if 12 or older (Y - form), and the youth's agency worker/case manager (W - form).

Content Areas

After considering a large number of potential content areas, four primary areas or domains of assessment were selected:

- 1) Problem severity,
- 2) Functioning,
- 3) Hopefulness, and
- 4) Satisfaction with behavioral health services.

The parent, youth, and agency worker rate the problem severity and functioning scales. The youth and parent rate the satisfaction scales. Youth rate their own hopefulness about life or overall well being. Parents (or primary caregivers) rate their hopefulness about caring for the identified child. In addition, the Restrictiveness of Living Environments Scales (ROLES; Hawkins, Almeida, Fabry, & Reitz, 1992) is included on the agency worker form along with data regarding several key indicators that are not used when scoring the form.

Item Development

Item writing and selection for the Ohio Scales necessitated isolating the most common problem areas and typical areas of functioning. Five sources of information were considered when writing items for the instruments:

- 1) problem behaviors listed as criteria for diagnosis of child and adolescent disorders in the DSM-IV,
- 2) a list of the most common "presenting problems" of youth with SED compiled by a regional mental health board (Cuyahoga County),
- 3) the results of the social validation survey,
- 4) several commonly used instruments were collected and examined to ascertain the typical areas of assessment when evaluating children and youth along with typical items, and

- 5) consultation with child service providers in three separate agency meetings involving 3 child program directors, 4 case manager supervisors, 23 case managers, and 5 parent/ parent advocates.

Short Form of the Ohio Scales

During the initial validation studies of the Ohio Scales, case managers and parents were given the opportunity to provide qualitative feedback regarding the instrument. Two common criticisms were voiced during the studies: 1) even though the Ohio Scales were only 72 items long, several individuals thought the scales could be shorter, and 2) some case managers suggested the reading level of the parent and case manager versions of the scales should be changed to match the youth version.

As a result, we modified the original scales. The descriptions that follow will include both the Short Form and the Original Ohio Scales. Psychometric studies will be presented for both scales. We anticipate that many will select the Short Form because of the increased usefulness in terms of readability and time needed for administration and scoring.

Item Descriptions

The "Problem Severity Scale" is comprised of 20 items (short form) or 44 items (original form) covering common problems reported by youth who receive behavioral health services. Each item is rated for severity/frequency (0 "Not at all" to 5 "All the time") on a six-point scale. A total score is calculated by summing the ratings for all items.

The "Functioning Scale" is comprised of 20 items (short form and original form) designed to rate the youth's level of functioning in a variety of areas of daily activity (e.g., interpersonal relationships, recreation, self-direction and motivation). Each item is rated on a five-point scale (0 "Extreme troubles" to 4 "Doing very well"). Although the problem severity scale is similar to many other existing symptom rating scales that focus on the severity of behavioral problems, the functioning scale provides a broader range of ratings including "OK" and "Doing very well". This provides an opportunity for raters to identify areas of functional strength. A total functioning score is calculated by summing the ratings for all 20 items. Higher scores are indicative of better functioning.

In addition to the problems and functioning scales, two brief (four item) scales (short form and original form) on the parent and youth forms assess satisfaction and hopefulness. Four items assess satisfaction with and inclusion in behavioral health services on a six-point scale (1 "extremely satisfied" to 6 "extremely dissatisfied"). The total satisfaction score is calculated by summing the 4 items.

Four additional items on the parent and youth forms tap levels of hopefulness and well-being either about parenting or self/future respectively. Each of these is also rated on a six-point scale. The total hopefulness score is calculated by summing the 4 items.

Finally, the agency worker version of the Ohio Scales includes a copy of the Restrictiveness of Living Environments Scale (ROLES). Information regarding the initial development of the ROLES can be obtained by reviewing the original article written by Hawkins et al. (1992). The ROLES assesses the level of restrictiveness for the youth's placements during the past 90 days. A higher score means on average the youth is placed in a more restrictive setting. Administration and scoring procedures for all three instruments are described below.

ADMINISTRATION AND SCORING

The Ohio Scales were developed for quick administration, scoring and interpretation. With relatively minimal training, parents, youth, or agency workers can administer, score, and interpret the meaning of scores for each of the scales. Each of the scales will be briefly discussed in this section.

There are three parallel forms of the Ohio Scales completed by the youth's parent or primary caregiver (P-form), the youth (Y-form), and the youth's agency worker (W-form). This allows assessment of the client's strengths and weaknesses from multiple perspectives. The youth form is designed for youth ages 12-18. The parent and agency worker versions are designed for youth ages 5-18.

The instrument is two pages long, placed on the front and back of a single sheet. The questions for problem severity and functioning are identical on the three parallel forms. The satisfaction and hopefulness scales are slightly different depending on the perspective (parent or youth). On the front side of all three forms is the problem severity scale which has 20 items on the Short-Form and 44 items on the original forms. The remaining scales are on the back.

Problem Severity

All three forms (parent, youth, and agency worker) include the problem severity scale. Each of these items is rated on a 6-point scale for frequency during the past 30 days: not at all, once or twice, several times, often, most of the time, or all of the time. The columns for each frequency are coded respectively from 0 (Not at all) to 5 (All of the Time). Each column's score can then easily be added at the bottom of the page. The sum of the six columns then becomes the individual's score on the problem severity scale. No items are reverse-scored. The only differences between the original and short forms for this scale are the number of items (44 - original; 20 - short form) and the easier wording for the Short-Form.

Functioning

All three forms include the 20 item functioning scale in the bottom half of the back page. Each of these 20 items is rated using a 5-point scale: extreme troubles, quite a few troubles, some troubles, OK, or doing very well. Since raters might have somewhat different conceptions regarding what constitutes the various levels of functioning, we use comparable ratings on the Children's Global Assessment Scale (CGAS) as a reference:

<u>Ohio Scales</u>	<u>CGAS</u>
Doing very well (4)	Superior functioning in all areas; (CGAS 90's)
OK (3)	Good functioning in all areas; (CGAS 80's)
Some Troubles (2)	Some difficulty in a single area, but generally functioning pretty well (CGAS approximately 70's)
Quite a few Troubles (1)	Moderate problems in most areas or severe impairment in one area (CGAS approximately 50's)

Extreme Troubles (0) Major impairment in several areas and unable to function in one or more areas (CGAS 30's or below)

A common question about the functioning scale involves the rating of items 3 and 13. For young children, raters often wonder how to rate items concerning vocational preparation (Item 13) or developing relationships with boyfriends or girlfriends (Item 3). On these items the rater should rate "OK (3)" if they are unsure or rate the youth based on what might be expected for their developmental level. For example, developmentally appropriate vocational preparation for a 7 year old typically involves school work, chores at home, and other work-like assignments. Note: If insufficient information is available to answer a specific item on the functioning scale, that item should be rated "OK (3)".

The functioning scale total is calculated in the same manner used on the problem severity scale. Each of the 20 items is rated on its 5-point scale. The rating for each item is circled. The columns for each frequency are coded respectively from 0 (extreme troubles) to 4 (doing very well). Each column's score can then easily be added at the bottom of the page. The sum of the five columns then becomes the individual's score on the functioning scale. No items are reverse scored.

As can be seen from the scoring method, a high score on the problem severity scale is considered to be more problematic (more frequent problems), while a low score on the functioning scale is considered to be more impairment. The method of scoring is thus congruent with what one would intuitively expect given the content of each scale. The short form and original Ohio Scales differ on this scale only in the wording of the items. The number of items remained unchanged. The parent (P-form) and agency worker (W-form) on the original were reworded to match identically the youth (Y-form) on the short form.

Hopefulness

On the back side of the parent and youth versions, eight questions are printed at the top of the page. The first four questions ask for ratings of hopefulness (parent) or overall well being (youth). The specific questions vary somewhat on the two versions to fit the respondents. Each question is answered according to a 6-point scale with the specific scale items varying to fit the questions. In each question, response "1" is the most hopeful/well and response "6" is the least. The four items can then be totaled for a hopefulness scale score. On this scale, a lower total means more hope or wellness. There are no differences in this scale between the original and short forms.

Satisfaction

The second four questions on the top half of the back page (P-form and Y-form) ask for ratings of overall satisfaction with behavioral health services received and ratings of their inclusion in treatment planning. The specific questions vary somewhat on the two versions to fit the respondents. Each question is answered according to a 6-point scale with the specific scale items varying to fit the questions. In each question, response

"1" is the most satisfied/included and response "6" is the least. The four items can then be totaled for a satisfaction scale score. On this scale, a lower total means more satisfaction. There are no differences in this scale between the original and short forms.

Restrictiveness of Living Environments Scale (ROLES)

On the agency worker version of the Ohio Scales (W-form), the space in the top half of the back side of the page is utilized quite differently since satisfaction and hopefulness ratings are only appropriate from the perspectives of the parent/caregiver and youth. The W-form includes a copy of the ROLES (Hawkins et al., 1986). The ROLES consists of a list of 23 categories of residential settings. Next to each specific setting is a blank line on which the agency worker writes the number of days (during the past 90 days) the youth was residing in that setting (The total of all the days will therefore add to 90). Although the authors of the Ohio Scales did not develop this scale, it was felt that tracking this information could be helpful to the agency worker. The worker should identify the categories that most closely resemble the settings in which the youth stayed.

Scoring for this scale is not included on the form, but it is possible to compute a score if the worker thinks it would be a meaningful measure of the child's treatment progress. Each setting is given a statistical 'weight' as listed in the table below. To get the ROLES total score, each weight is multiplied by the number of days in the blank next to the setting. The sum of these products is then calculated to get a total. The total is then divided by 90 to get the average restrictiveness for the previous 90 days. This is the ROLES score (see Hawkins et al., 1986).

Table 1. ROLES' Weights

<u>Setting</u>	<u>Weight</u>	<u>Setting</u>	<u>Weight</u>
Jail	10.0	Foster care	4.0
Juvenile detention/youth corrections	9.0	Supervised independent living	3.5
Inpatient psychiatric hospital	8.5	Home of a family friend	2.5
Drug/alcohol rehab. center	8.0	Adoptive home	2.5
Medical hospital	7.5	Home of a relative	2.5
Residential treatment	6.5	School dormitory	2.0
Group emergency shelter	6.0	Biological father	2.0
Vocational center	5.5	Biological mother	2.0
Group home	5.5	Two biological parents	2.0
Therapeutic foster care	5.0	Independent living with friend	1.5
Individual home emergency shelter	5.0	Independent living by self	.5
Specialized foster care	4.5		

For example, if during the last 90 days a child was placed in a juvenile detention facility for 2 days, a group home for 12 days, and with the biological father for 76 days, the ROLES score would be calculated in this way:

	<u>Days</u>		<u>Weight³</u>		<u>Product</u>
Detention Center	2	X	9.0	=	18.0
Group Home	12	X	5.5	=	66.0
With Father	76	X	2.0	=	<u>152.0</u>
Total	90				236.0

$236 / 90 = \underline{2.62}$ - The ROLES score for the past 90 days is 2.62.

The agency worker version also includes a several questions in the middle of the back side of the page. These items are 'Marker' questions and, similar to the ROLES, are meant to be helpful to the agency worker in tracking key information. There are blank spaces to write in information on "school placement" and "current psychoactive medications". In addition, several lines are available for recording the frequency during the past 3 months of arrests, suspensions from school, days in detention, days of school missed, and self-harm attempts.

³ From the Table on the previous page.

PSYCHOMETRIC PROPERTIES - ORIGINAL OHIO SCALES

To begin evaluating the psychometric properties of the original instrument, seven samples of data were collected. After these studies were conducted, qualitative feedback regarding the scales served as a catalyst to make two changes - make the problem severity scale shorter and changing the wording of the problem severity and functioning scales on the parent and agency worker forms to match the youth report (Y - form). A description of psychometric studies unique to the Short Form is included in the next section.

- 1) A total of 301 Jr. High and High School students (average age 14.36, SD 1.54; 118 boys, 159 girls, 24 missing sex data) completed the youth version of the instruments. Youth from all grades were represented (7th – 58, 8th – 54, 9th – 65, 10th – 59, 11th – 45, 12th – 17, Missing data – 3). Average grade point average for the youth participants was 3.11 (SD = .789) on a five point scale (range .5 – 5.0). All but ten of the youth (291) also asked one parent or primary caregiver (average age 39.43, SD 7.36; 218 women, 58 men, 25 missing data) to rate them using the parent version of the Ohio Scales (88% of the adults responding were one of the biological parents of the child).
- 2) In addition to the middle and high school data, a sample of 225 parent ratings of K through 6th grade students were also collected. The children were 104 boys and 115 girls (6 missing data) with an average age of 8.86 years (SD = 2.23). All grades were represented in the sample (K – 27, 1st – 31, 2nd – 35, 3rd – 25, 4th – 33, 5th – 33, 6th – 39, missing data - 2). The parents (32 men and 190 women, 3 missing data) averaged 35.01 years old (SD = 5.93).
- 3) An initial clinical sample was collected consisting of 59 case manager ratings of youth receiving behavioral health services. In addition, 28 of the 59 parents rated their youth and 16 adolescents completed the youth self-report version of the Ohio Scales. The 59 youth (40 boys, 17 girls, 2 missing data) were an average 12.54 years old (SD = 3.85).
- 4) A second clinical sample was collected from two agencies across four sites from parents (n = 66) of youth who entered child community support services. The youth who were 12 or older (n = 26) also participated by completing self-report forms. Case managers also rated the 66 youth receiving services. The youth (42 boys and 24 girls) were an average 10.75 years old (SD = 3.73).
- 5) Forty parents or primary caregivers and 17 adolescents who were receiving mental health services completed the Ohio Scales twice with a one week interval to investigate the test-retest reliability,
- 6) Eight students and four case managers rated vignettes and clinical intake paperwork to investigate the inter-rater reliability of the case worker rated functioning scale, and

- 7) A large group of adolescents who received outpatient counseling through a multi-state Behavioral Health Care Management Organization completed the functioning scale of the youth self-report version of the Ohio Scales.

Procedures

Instruments and procedures for the seven samples were slightly different and are described separately here. The means and standard deviations on the Ohio Scales for each sample are displayed in Table 2.

Sample 1. Research assistants distributed packets to Jr. High and High School students (grades 7 – 12) near the end of a school day. The packet included a brief letter explaining the study (including implied consent by returning the forms), the Ohio Scales Problem Severity and Functioning Scales for the parent, the Ohio Scales Problem Severity and Functioning Scales for the youth, and several demographic questions.⁴ Students (and parents) were instructed to complete the forms in the evening and return them in an envelope to the research assistants prior to school the next morning. All students who returned the forms received one dollar for their participation. Research assistants collected forms two consecutive mornings after the packets were distributed. Students could also return the forms to the school secretary thereafter. A total of 301 students and 291 parents returned completed forms (some individual items were inadvertently left blank for some participants).

Sample 2. The second sample was completed in the same fashion as sample #1 except students were enrolled in grades K – 6. As a result, the packet did not include forms for the student to complete. Children who returned the parent-completed forms received a dollar the next morning. Again, research assistants collected forms two consecutive mornings and students could return the forms to the school secretary thereafter. Of the 491 children registered to attend the school, parents of 225 (46%) completed ratings of their children.

Sample 3. This initial clinical sample was collected at a community behavioral health center in rural, southeastern Ohio. Case managers rated the cases using the Ohio Scales and the Progress Evaluation Scales (Ihlevich & Gleser, 1982). A total of 59 youth currently receiving services were rated. Each case was rated twice by the primary case manager with a four-month interval between ratings. In addition, the cases were also rated by a second case manager who was familiar with the case using the Ohio Scales. Finally, 28 parents participated by rating their child using the Ohio Scales and Child Behavior Checklist (Achenbach & Edelbrock, 1983). Youth, 16, also completed the Ohio Scales and Youth Self Report.

⁴ The satisfaction scale was not included since most of the children were not participating in mental health services.

Sample 4. The second clinical sample included youth and their parent or primary caretaker who were referred for community support services at one of four sites across two agencies. Families participated in an intake interview or other initial services (e.g., outpatient counseling) and were then referred for community support services. A total of 66 families agreed to participate in the research. Families who participated were asked to complete several forms. Parents completed the Ohio Scales and the Vanderbilt Functioning Index. Youth who were 12 or older completed the Ohio Scales. The community support worker completed the Ohio Scales, Child and Adolescent Functional Assessment Scales (Hodges & Wong, 1996), Restrictiveness of Living Environments Scale (Hawkins et al., 1992), and the Children’s Global Assessment Scale (Shaffer et al., 1983). The parent, youth, and community support worker were each asked to complete the Ohio Scales every 3 months as long as the family continued to receive services.

Sample 5. To obtain estimates of test-retest reliability and validity for the satisfaction scale, 40 parents/primary caretakers and 17 youth over 12 who attended appointments with the contractual psychiatrist at one agency were asked to complete the Ohio Scales and the Client Satisfaction Questionnaire (CSQ-8; Attkisson & Zwick, 1982) after their appointment. They were also asked to take home a second copy of the Ohio Scales, which they completed one-week later then returned via mail. Thirty-seven parents (93%) and 14 youth (82%) returned the second set of forms. Each participant received \$5 for completing the forms.

Sample 6. Four case managers, four undergraduate students, and four graduate students completed ratings of 20 cases using four measures of functioning – Ohio Scales Functioning Scale, Child and Adolescent Functional Assessment Scale (Hodges & Wong, 1996), Children’s Global Assessment Scale (Shaffer et al., 1983), and the Vanderbilt Functioning Index (Bickman, 1997). Ten of the cases were vignettes developed by Kay Hodges for training raters to use the Child and Adolescent Functional Assessment Scales. These vignettes were organized based on a structured interview used to collect relevant data for rating the CAFAS. The other ten cases were copies of the actual intake paperwork generated at one clinical facility (with names removed).

In addition, the four case managers rated 10 children each using the Ohio Scales Problem Severity and Functioning scales. The case managers were instructed to think of children and adolescents that they knew personally and who were not currently participating in any form of behavioral health treatment. These ratings were obtained to make a first estimate concerning “normal” means and standard deviations on the case manager rated scale. Many rater-based scales do not include norms. For example, the Hamilton Rating Scale for Depression has been used in hundreds of studies in various forms, but no normative sample is available (Grundy et al., 1994; Grundy, Lambert, & Grundy, 1996). As a result, we collected this initial data to begin the process of developing a rater based comparison sample that could be contrasted with clinical samples.

Sample 7. A sample of nearly 1900 adolescents who were entering outpatient treatment through a large managed behavioral healthcare provider completed the Ohio Scales Functioning scale upon intake and periodically throughout treatment.

Table 2. Means and Standard Deviations on the Original Ohio Scales for the different samples.

<u>Population: Sample Number</u>	<u>N</u>	<u>Problems*</u> <u>M (SD)</u>	<u>Functioning</u> <u>M (SD)</u>	<u>Hope</u> <u>M (SD)</u>
Community: Sample # 1				
• Youth	297	33.93 (29.15)	60.44 (13.32)	9.70 (3.77)
• Parents	285	24.28 (31.76)	62.73 (14.17)	8.31 (3.52)
Community: Sample # 2				
• Parents	225	19.48(18.06)	63.38 (14.63)	7.83 (2.86)
Clinical: Sample # 3				
• Youth	16	48.44 (29.48)	52.00 (10.75)	8.94 (3.86)
• Parent	28	56.11 (35.19)	45.11 (12.67)	12.48 (5.11)
• Case Manager	59	42.98 (23.41)	37.83 (14.33)	NA
Clinical: Sample # 4				
• Youth	17	67.29 (30.92)	47.00 (15.78)	13.35 (4.99)
• Parent	52	65.10 (36.56)	43.75 (15.02)	12.44 (4.58)
• Case Manager	53	49.30 (24.54)	42.82 (13.00)	NA
Clinical: Sample # 5				
• Youth	17	45.47 (28.52)	57.88 (11.08)	9.29 (4.54)
• Parent	40	66.50 (32.12)	41.05 (18.21)	12.90 (5.63)
Community: Sample # 6				
• Case Manager	40	17.58 (9.62)	67.03 (9.01)	NA
Clinical: Sample # 7				
• Youth	1897	NA	51.12 (13.95)	NA

Note: All clinical means and standard deviations reflect intake levels of problems and functioning.
 * All problem severity score means in this table were based on administration of the original 44-item problem severity scale. Means and standard deviations for the short form are included in the next section.

Reliability

Internal Consistency. Internal consistency data for each scale for the three perspectives are presented in Table 3. Data are presented for both clinical and comparison samples. As can be seen, the internal consistencies for each scale are adequate or better. Examination of the individual item-total correlations suggested few items were poor. On the problem severity scale several items had infrequent endorsement and as a result had lower item-total correlations. These items were retained for their informative value despite low base rates of endorsement.

Table 3. Internal Consistency Estimates (Cronbach's Alpha) for each Scale on the Three Instruments for Community and Clinical Samples.

Community Samples			
Scale	Sample #1		Sample #2
	Parent (n = 242)	Youth (n = 245)	Parent (n = 217)
Problem Severity	.97	.95	.93
Functioning	.95	.92	.95
Hopefulness	.71	.75	.65
Satisfaction	NA	NA	NA

Clinical Sample #3			
Scale	Parent (n = 23)	Rater	
		Youth (n = 15)	Agency worker (n = 59)
Problem Severity	.96	.90	.93
Functioning	.89	.75	.94
Hopefulness	.86	.84	NA
Satisfaction	.79	.72	NA

Clinical Sample #4			
Scale	Parent (n = 59)	Rater	
		Youth (n = 21)	Agency worker (n = 64)
Problem Severity	.95	.93	.92
Functioning	.93	.91	.94
Hopefulness	.87	.75	NA
Satisfaction	.72	.82	NA

Test-retest Reliability. Test-retest reliability (one week) was evaluated for the parent and youth versions of the Ohio Scales. Test-retest reliability estimates for both parent and youth samples on four scales are presented in Table 4. As can be seen, test-retest reliability was adequate or better for all four scales on both the parent and youth rated versions with the exception of the youth rated functioning scale (This may have been influenced by the small number of youth and one outlier). Test-retest reliability was poorest for the satisfaction scale. This may have been influenced by the testing over two times in different locations. Participants completed the satisfaction scale at time one while waiting for or just after their appointment with the psychiatrist. At time two, however, they completed the scales in their own home and then mailed the forms back to the researcher. This may have influenced their willingness to be critical of the agency. The data from sample 7 (adolescents in outpatient treatment) also provides some

information regarding test-retest reliability. These adolescents were administered the youth self-report functioning scale at irregular intervals while participating in outpatient treatment. This provided the opportunity to examine the correlation between ratings at session 1 with session 2 and session 3. As can be seen in Table 4, the test-retest reliability in this circumstance was adequate.

Table 4. Test-Retest Reliability Estimates for the Parent and Youth Rated Instruments.

<u>Scale</u>	<u>Rater</u>			
	Parent ^a Sample 5 (<u>n</u> = 37)	Youth ^a Sample 5 (<u>n</u> = 14)	Youth ^b Sample 7 (<u>n</u> = 15)	Youth ^c Sample 7 (<u>n</u> = 611)
Problem Severity	.88	.72	-	-
Functioning	.77	.43	.79	.68
Hopefulness	.79	.74	-	-
Satisfaction	.67	.67	-	-

^a 1 week test-retest; ^b Session 1 to session 2; ^c Session 1 to Session 3

Inter-rater Reliability. The inter-rater reliability was investigated for the agency worker version of the Functioning Scale using two different methods. In sample 3, two case managers (one primary case manager and another case manager who was acquainted with the youth) rated the same youth. The statistical relationship between the ratings of two caseworkers that were familiar with the case was a modest .44 correlation. Since it was not clear if the primary case manager and the second agency worker had similar information when rating the youth, we decided to investigate the inter-rater reliability of the case manager ratings using a more stringent methodology.

In sample 6, four undergraduate students, four graduate students, and four case managers rated 20 cases as described on paper (10 sets of clinical intake paperwork and 10 vignettes that were presented in a standard format based on structured telephone interviews for collecting clinical information developed by Kay Hodges; Hodges & Wong, 1996). The raters used four measures of functioning including the Ohio Scales Functioning Scale, The Child and Adolescent Functional Assessment Scale, The Vanderbilt Functioning Index, and the Children’s Global Assessment Scale). More details regarding the study are available in published format (Ogles, Davis, & Lunnen, 1998).

To examine the inter-rater reliability, inter-rater correlations were calculated for each pair of undergraduates, graduates, and case managers respectively. Correlations were then averaged (across measures and methods) to examine the influence of rater level of training on inter-rater reliability. Table 5 presents average correlations for each measure within each rater group. As can be seen, undergraduates were able to make

equally reliable ratings as the graduate students (significance tests were not performed). Case managers were slightly lower, but no significance tests were performed.

Overall, the level of training did not seem to influence inter-rater reliability. This suggests that no sophisticated clinical training is necessary when raters have sufficient training on the instruments. Students and paraprofessionals may be used to conduct ratings in typical studies. This may represent a substantial savings in research dollars for larger studies. The reader may note that the interrater correlations are on average quite low. Please note that averaging across methods attenuates the averages within groups.

Table 5. Inter-rater Reliability for Four Measures of Functioning for Three Rater Groups across Methods of Presentation.

<u>Measure</u>	<u>Undergraduates</u>	<u>Graduates</u>	<u>Case Managers</u>
CGAS	.69	.62	.38
CAFAS	.77	.81	.74
Ohio Scales	.58	.57	.50
Vanderbilt	.76	.68	.58
Average	.70	.68	.58

Inter-rater correlations were also calculated for each pair of raters within the two methods of presenting the case materials. Correlations were then averaged to examine the influence of method of case presentation on the inter-rater reliability. Table 6 presents average correlations for each measure within each method of presentation. Clearly, the standardized format improved reliability. When raters examined and rated intake forms, reliability was significantly attenuated. The intake forms varied widely in their degree of completeness, accuracy, and adequacy.

Table 6. Inter-rater Reliability for Four Measures of Functioning using Vignettes and Clinical Folders.

<u>Measure</u>	<u>Vignettes</u>	<u>Clinical Folders</u>
CGAS	.77	.33
CAFAS	.90	.66
Ohio Scales	.88	.22
Vanderbilt	.86	.59
Avg. inter-rater reliability	.77	.33

Overall, the measures seemed to produce rather similar levels of reliability across methods of presentation and rater groups. The CAFAS was the most immune to decreases in reliability when using the clinical cases that had variable amounts of data presented in an unstandardized format. When using standardized vignettes (similar information organized in the same format), inter-rater reliability was excellent (.77 to .90). When using clinical intake forms that varied widely in completeness and organization, inter-rater reliability was attenuated (.22 to .66). This suggests that a standardized, comprehensive method of data collection and presentation may be needed in applied settings. For example, Hodges (Hodges & Wong, 1996) has developed a standardized telephone interview for collecting and organizing information to be used when making CAFAS ratings. This or another similar structured interview may improve inter-rater agreement through minimizing differences in available information. This may also help explain the poor correlation between case manager ratings on the Ohio Scales in a clinical setting (Sample #3). Using a standardized format for the collection of data will produce reliable agency worker ratings of youth functioning.

Validity

Data were collected for several samples to provide evidence of validity. Validity data are presented for each source of data collected: agency worker, parent, and youth.

Agency Worker. The agency worker Ohio Scale ratings in sample #3 were correlated with the Progress Evaluation Scales (Ihilevich & Gleser, 1979; both completed by the case manager). Problem severity and functioning were both significantly correlated with scores on the Progress Evaluation Scales ($r = .58$ & $.44$, $p < .05$, respectively). This suggests a modest overlap of constructs.

In sample #4, case managers completed the Ohio Scales, the Child and Adolescent Functional Assessment Scales (Hodges & Wong, 1996), Children's Global Assessment Scale (Shaffer et al., 1983), and Restrictiveness of Living Environments Scale (Hawkins et al., 1992). Correlations among the measures of functioning are presented in Table 7. As can be seen, the agency worker version of the Ohio Scales was modestly correlated with the two measures of functioning (.59 and -.52 with the CAFAS and .31 & .32 with the CGAS). The Ohio Scales were not related to the restrictiveness in living environments. It should be noted that there was a restricted range of living environments – most youth were living at home. In addition, issues other than level of functioning often determine placement. For example, the CAFAS appears to be correlated with the current placement. Under closer examination, however, it was apparent that the CAFAS item that refers to current alcohol and drug use was the best predictor of current placement. In essence youth with serious drug and alcohol problems were the most likely to be placed in a more restrictive setting or removed from their homes.

Table 7. Correlations Among Agency Worker Rated Measures in Sample #4

	CGAS	CAFAS	Ohio Scales Functioning	Ohio Scales Problem Severity
CAFAS	-.26	-	-	-
OS - Functioning	.31*	-.52**	-	-
OS - Problem Severity	-.32*	.59**	-.43**	-
ROLES	-.25	.31*	.00	-.13

* p < .05 (2-tailed)

** p < .001 (2-tailed)

Correlations were also calculated among the measures used in Sample #6 (across all raters, methods, and cases). As can be seen in Table 8, the four measures of functioning are significantly related to one another. The correlations range from .54 to .66 and suggest a moderate degree of overlap (30% to 44% shared variances). The four measures of functioning appear to be tapping into the same basic core construct as evidenced by the moderate degree of shared variance among the measures. This would suggest that choices among the measures might be governed by other factors such as: inter-rater reliability, cost, required training, ease of use, etc. At the same time, correlations among the measures were modest and may suggest that different types of functioning are assessed. Further research is needed to investigate the similarity of measures.

Table 8. Correlations Among Four Measures of Functioning Rated by Graduates, Undergraduates, and Case Managers in Sample #6

	CGAS	CAFAS	Ohio Scales
CAFAS	-.66*	-	-
Ohio Scales	.62*	-.59*	-
Vanderbilt	-.54*	.64*	-.60*

* p < .01 (2-tailed)

Additional evidence for validity is obtained through comparing the community and clinical samples. For sample 6, four case managers rated 10 children each using the

Ohio Scales Problem Severity and Functioning scales. The case managers were instructed to think of children and adolescents that they knew personally and who were not currently participating in any form of behavioral health treatment. The case managers were also asked to think of children within each of the 10 age ranges. These ratings were obtained to make a first estimate concerning “normal” means and standard deviations on the agency worker rated scale.

As can be seen, in Table 9, the 40 youth included in this comparison sample had significantly lower scores on the problem severity scale, $t(97) = 6.49, p < .001$ & $t(91) = 7.73, p < .001$, and significantly higher scores on the functioning scale, $t(97) = 2.99, p < .05$ & $t(91) = 2.99, p < .05$, than both clinical samples respectively.

Table 9. Means and Standard Deviations on the Ohio Scales for clinical and community samples rated by the case manager.

<u>Sample</u>	<u>N</u>	<u>Problems</u> <u>M (SD)</u>	<u>Functioning</u> <u>M (SD)</u>
Initial clinical sample	59	42.98 (23.41)	37.83 (14.33)
Grant funded clinical sample	53	49.30 (24.54)	42.82 (24.54)
Grant funded community sample	40	17.58 (9.62)	67.03 (9.01)

Parent Ratings. In sample #3, parent ratings of the youth's problem severity and functioning were correlated with the CBCL (Achenbach & Edelbrock, 1983). As can be seen in Table 10, the correlations are significant for both problem severity and functioning scales correlated with total CBCL. Hypothesized correlations are underlined. The CBCL is primarily a “symptom” oriented instrument and was consequently included primarily to establish the concurrent validity of the problem severity scale. The VFI was included to investigate the concurrent validity of the functioning scale.

Table 10. Concurrent Validity Estimates for the Parent Rated Ohio Scales

<u>Instrument</u>	<u>Instrument</u>	
	<u>Ohio Scales- Parent</u> <u>Problem Severity</u>	<u>Ohio Scales -Parent</u> <u>Functioning</u>
Child Behavior Checklist (CBCL)	<u>.89</u> **	<u>.77</u> **
Vanderbilt Functioning Index	<u>.39</u> **	<u>.54</u> **

** $p < .001$

The significant differences between the community and clinical samples also provides evidence for the discriminant validity of the parent rated Ohio Scales. The 2 community samples differ from all 3 clinical samples in terms of parent rated problem severity, functioning, and hopefulness (all p values $< .01$).

Within group differences in the community sample (sample #1) provide more evidence for the discriminant validity. Five t-tests were conducted using parent ratings of problem severity and functioning to examine differences between students who had repeated a grade, been arrested, received behavioral health services, assigned to classes for students with behavioral problems (SBH), or assigned to classes for students with learning problems (LD) and those who had not experienced these events (Table 11).

Students with learning difficulties, or who had received behavioral health services or had been arrested had significantly poorer functioning and more severe problems than students who had not experienced these events. Students who had previously been assigned to classes for youth with behavior problems had poorer functioning (but not more severe problems) than students who had not been assigned to these classes. There was no significant difference in functioning or severity of problems for students who had repeated a grade from youth who had not repeated a grade.

Table 11. Means and Standard Deviations on Parent Ratings of Problem Severity and Functioning.

Sample	Problem Severity		Functioning	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Assigned to LD class ($\underline{n} = 52$) ^a	33.5	30.9	55.9	13.6
Never assigned to LD class ($\underline{n} = 229$)	20.5	27.3	64.5	13.1
Assigned to SBH class ($\underline{n} = 9$) ^b	34.9	27.5	43.9	13.7
Never assigned to SBH class ($\underline{n} = 271$)	22.6	28.2	63.6	13.3
Arrested ($\underline{n} = 19$) ^a	55.8	58.3	50.7	18.1
Never arrested ($\underline{n} = 262$)	20.9	24.0	63.8	13.0
Received behavioral health services ($\underline{n} = 59$) ^a	32.9	27.7	56.7	16.4
Never received services ($\underline{n} = 221$)	20.8	28.7	64.5	12.5
Repeated a Grade ($\underline{n} = 50$) ^c	30.0	35.6	60.1	15.5
Never Repeated a Grade ($\underline{n} = 228$)	21.7	26.8	63.5	13.2

^a different on both problem severity and functioning, $p < .05$, ^b different on functioning only, ^c not significantly different on problem severity or functioning.

To further examine the construct validity of the parent rated scales, we factor analyzed the problem severity, functioning, and hopefulness scales using a principal components extraction with a varimax rotation ($n = 609$; combined samples). The factors were selected by an examination of the scree plots along with considering the interpretability of the factors. We expected to find a factor structure similar to other problem behavior scales when analyzing the problem severity scale. For the functioning and hopefulness scales we hoped to find evidence for a single underlying factor.

The factor analysis of the hopefulness scales resulted in a one factor solution that accounted for 57% of the variance. All four items had loadings above .39 on the single factor. Factor loadings for the hopefulness scale are displayed in Table 12.

Table 12. Factor Loadings on the Parent Rated Hopefulness Scale

Item	Loading	Item	Loading
1	.39	3	.47
2	.76	4	.67

The factor analysis of the problem severity scale resulted in a three factor solution which accounted for 54% of the variance. The factors were labeled: conduct disturbance, externalizing, and internalizing. Factor loadings above .40 are displayed in Table 13. Seven of 44 items had loadings above .40 on more than one factor.

The factor analysis of the functioning scale resulted in a two factor solution that accounted for 57% of the total variance. The factors were labeled: overall functioning and transitional areas of functioning. Only three items loaded on factor two. All three items referred to areas of functioning that are more applicable to teenaged youth who are preparing for the transition into adulthood: romantic relationships, vocational preparation, and financial management. Factor loadings are displayed in Table 14.

Table 13. Factor Loadings for the Parent Rated Problem Severity Scale

<u>Item</u>	<u>Factor 1</u>	<u>Factor 2</u>	<u>Factor 3</u>	<u>Item</u>	<u>Factor 1</u>	<u>Factor 2</u>	<u>Factor 3</u>
1		.73		23	.63		
2		.75		24	.49		
3		.78		25	.49		
4		.68		26		.41	.51
5	.67			27	.40		
6		.74		28	.53		
7		.71		29			.53
8	.75			30	.45		.47
9	.66			31		.54	.55
10	.69			32		.44	.53
11	.68			33			.62
12	.66			34	.42		.60
13	.62			35			.59
14	.61			36	.46		.54
15	.45	.59		37			.63
16		.66		38			.51
17		.66		39	.65		
18	.60			40	.66		
19	.44			41	.46		
20		.63		42		.51	
21		.69		43			.40
22		.69		44			.50

Table 14. Factor Loadings on the Parent Rated Functioning Scale

<u>Item</u>	<u>Factor 1</u>	<u>Factor 2</u>	<u>Item</u>	<u>Factor 1</u>	<u>Factor 2</u>
1	.68		11	.67	
2	.71		12	.77	
3		.81	13		.79
4	.68		14	.77	
5	.63		15	.76	
6	.61		16	.82	
7	.74		17	.48	.58
8	.83		18	.74	
9	.76		19	.76	
10	.69		20	.69	

These factor analyses support the construct validity of the three scales. The hopefulness scale was in fact represented by one primary factor. This is not surprising, however, given the small number of items. The functioning scale was also represented by one main factor. The second factor represented three items that are more applicable to adolescents. In interviews with agency workers and parents, they often expressed concern about these items when rating younger children. It was clear that the distribution of scores on these items differed from other items because the parents were not sure how to rate young children. The problem severity scale factor analysis resulted in three main factors that are similar to the internalizing/externalizing superordinant factors that have been identified elsewhere in the literature (Achenbach & Edelbrock, 1983).

Youth Ratings. In sample #3, youth ratings of problem severity and functioning were correlated with the Youth Self Report (Table 15). As can be seen, the correlations are significant. We were especially interested in the relationship between the problem severity and Youth Self Report since both tap behavioral problems. One might argue that youth rated functioning should not be correlated with the Youth Self Report. No measure of functioning has been used to substantiate the validity of the youth rated functioning on the Ohio Scales to date.

Table 15. Concurrent Validity Estimates for the Youth Rated Ohio Scales

	<u>Ohio Scales - Youth Problem Severity</u>	<u>Ohio Scales - Youth Functioning</u>
Youth Self-Report (YSR)	<u>.82</u> **	<u>.46</u> *

** p < .001; * p < .05

The community sample (sample #1) also provides some evidence for the discriminant validity of the youth rated Ohio Scales. As with the parent ratings, five t-tests were conducted to examine differences between students who had repeated a grade, been arrested, received behavioral health services, assigned to classes for students with behavioral problems (SBH), or assigned to classes for students with learning problems (LD) and those who had not experienced these events (Table 16). Students who had been assigned to classes for youth with behavioral difficulties had significantly lower scores on the functioning scale. Students who had received previous behavioral health services had higher scores on the problem severity scale. No other significant differences were noted.

Table 16. Means and Standard Deviations on Youth Ratings of Problem Severity and Functioning.

<u>Sample</u>	<u>Dependent Variable</u>			
	<u>Problem Severity</u>		<u>Functioning</u>	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Assigned to LD class (<u>n</u> = 50) ^c	31.66	25.89	60.18	14.03
Never assigned to LD class (<u>n</u> = 228)	34.26	30.59	60.85	13.29
Assigned to SBH class (<u>n</u> = 9) ^b	55.55	35.60	51.44	12.88
Never assigned to SBH class (<u>n</u> = 271)	33.02	29.26	61.07	13.29
Arrested (<u>n</u> = 19) ^c	33.63	26.96	59.16	15.97
Never arrested (<u>n</u> = 262)	33.69	29.87	60.83	13.17
Received behavioral health services (<u>n</u> = 59) ^d	42.22	38.97	58.37	14.74
Never received services (<u>n</u> = 221)	31.47	26.32	61.30	12.94
Repeated a Grade (<u>n</u> = 50) ^c	32.62	21.76	60.66	13.48
Never Repeated a Grade (<u>n</u> = 228)	34.04	31.20	60.59	13.48

^b significantly different on functioning only.
^c not significantly different on problem severity or functioning.
^d significantly different on problem severity only

Finally, youth rating differences between the community sample and clinical samples provide evidence of the discriminant validity of the youth rated Ohio Scales. Returning to Table 2, all clinical samples differed from the community sample in terms of problem severity (sample #5 differed from sample #1 at the $p < .10$ level). Similarly, all four clinical samples differed from the community sample in self-report functioning, $p < .001$. Only one clinical group, however, differed from the community sample on the well-being scale (sample 4 > sample 1).

Parent and Youth Rated Satisfaction. To assess the validity of the parent and youth rated satisfaction scales, parents (40) and youth (17) who participated in the test-retest reliability study were also administered the Client Satisfaction Scale – 8 (Attkisson and Zwick, 1982). The correlation between the Ohio Scales 4-item satisfaction scale and the CSQ-8 for parents was -.68. The correlation between the Ohio Scales 4-item satisfaction scale the CSQ-8 rated by the youth was -.52. In both cases the correlations were statistically significant yet modest. This indicates that the two measures overlap to some degree.

Sensitivity to Change

In order to investigate the sensitivity of the Ohio Scales to change, three samples of data were collected and analyzed: sample #3, sample #4, and sample #7.

Correlation with the PES. In sample #3, case managers rated youth problems and functioning twice with a four-month interval between ratings. Ratings were collected for the Ohio Scales and Progress Evaluation Scales. All youth were participating in behavioral health services. Changes in scores on the problem severity and functioning scales were then correlated with changes in scores on the Progress Evaluation Scales. As can be seen in Table 17, change scores on both the problem severity and functioning scales were significantly correlated with change scores on the Progress Evaluation Scales. This suggests that changes on an instrument that has been used to assess outcome co-occur with changes on the Ohio Scales.

Table 17. Sensitivity to change estimates for the Agency Worker Rated Ohio Scales.

<u>Instrument</u>	<u>Instrument</u>	
	<u>Δ Ohio Scales Problems</u>	<u>Δ Ohio Scales Functioning</u>
Δ Progress Evaluation Scales (PES) (<u>n</u> = 48)	-.54*†	.56*

* p <.001

† The PES contains 7 items, higher values indicate lower numbers of problems and higher levels of present functioning.

Longitudinal Change. Additional evidence of sensitivity to change was collected in sample #4. A total of 53 children who were enrolled in community support services at four offices within two agencies were enrolled in a longitudinal study. Families that agreed to participate were asked to complete the Ohio Scales at intake and every three months thereafter while they were receiving services up to a one-year follow-up. Parents completed all four content areas of the Ohio Scales, Agency workers rated the youth using the problem severity and functioning scales, and youth who were 12 or older completed the four content areas of the youth self-report version of the Ohio Scales.

Table 18 displays the number of individuals who completed the forms at each time point. As can be seen, a large number of families dropped out of services over time and were not included in the follow-up. As a result, conclusions regarding the analysis of these data must remain guarded. Families that did not continue with services may have dropped out when their situation improved, deteriorated, or remained unchanged. Unfortunately, we do not know why they dropped out of services.

Table 18. Number of Individuals Completing the Follow-up Ratings.

<u>Rater</u>	<u>Intake</u>	<u>3 months</u>	<u>6 months</u>	<u>9 months</u>	<u>12 months</u>
Parent	52	25	12	16	5
Agency Worker	53	26	13	14	4
Youth	13	7	5	6	3

While the number of dropouts was high (ca. 50%), we conducted analyses to examine change in problem severity, hopefulness, and functioning. Paired t-tests examining changes from intake to 3 months were first examined. Means, standard deviations, and significance tests for the measures are presented in Table 19.

As can be seen (next page), the parents, case managers, and youth all reported significant changes in problem severity. No changes were noted, however, in functioning, or hopefulness/well being.⁵ Because of the small N's, no additional analyses were conducted to examine the significance of 6, 9, or 12 month change.

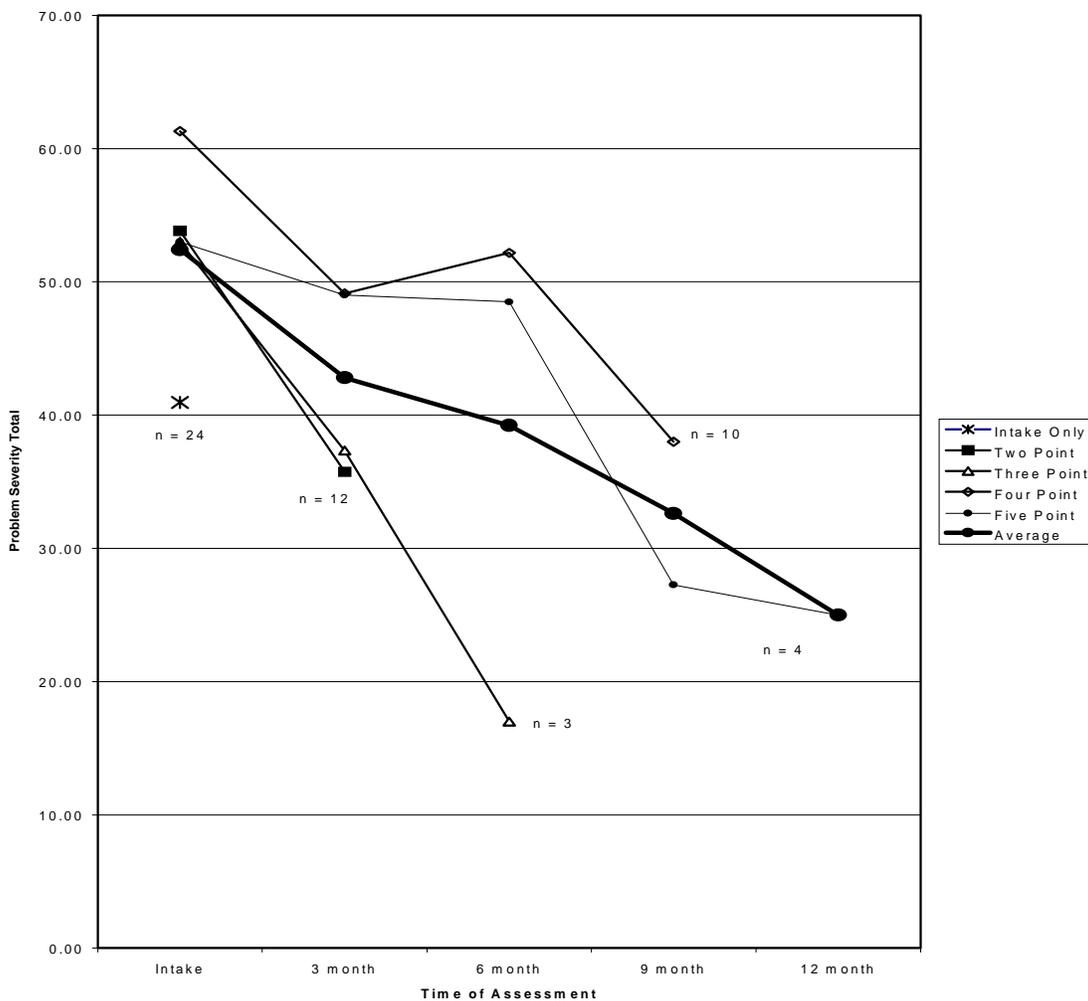
Table 19. Means, Standard Deviations, and Significance Tests for Three Sources of Information in Three Content Areas from Intake to 3 month Assessment.

<u>Rater</u> <u>Scale</u>	<u>Intake</u> <u>X (SD)</u>	<u>3 months</u> <u>X (SD)</u>	<u>T</u>	<u>Sig.</u>
Parent (n = 25)				
Problem Severity	69.4 (32.8)	50.0 (32.0)	3.64	.001
Functioning	41.6 (15.8)	45.0 (14.2)	-1.24	.225
Hopefulness	12.8 (4.84)	11.9 (4.17)	.854	.401
Agency Worker (n = 26)				
Problem Severity	57.5 (24.1)	41.6 (18.0)	3.06	.005
Functioning	39.3 (12.8)	40.3 (11.9)	-.634	.532
Youth (n = 7)				
Problem Severity	60.3 (30.8)	36.7 (23.2)	2.35	.057
Functioning	50.6 (14.7)	47.0 (13.7)	.624	.556
Well Being	11.4 (3.30)	10.0 (2.58)	1.59	.162

⁵ Lack of power is an issue for statistics calculated using the youth report scales.

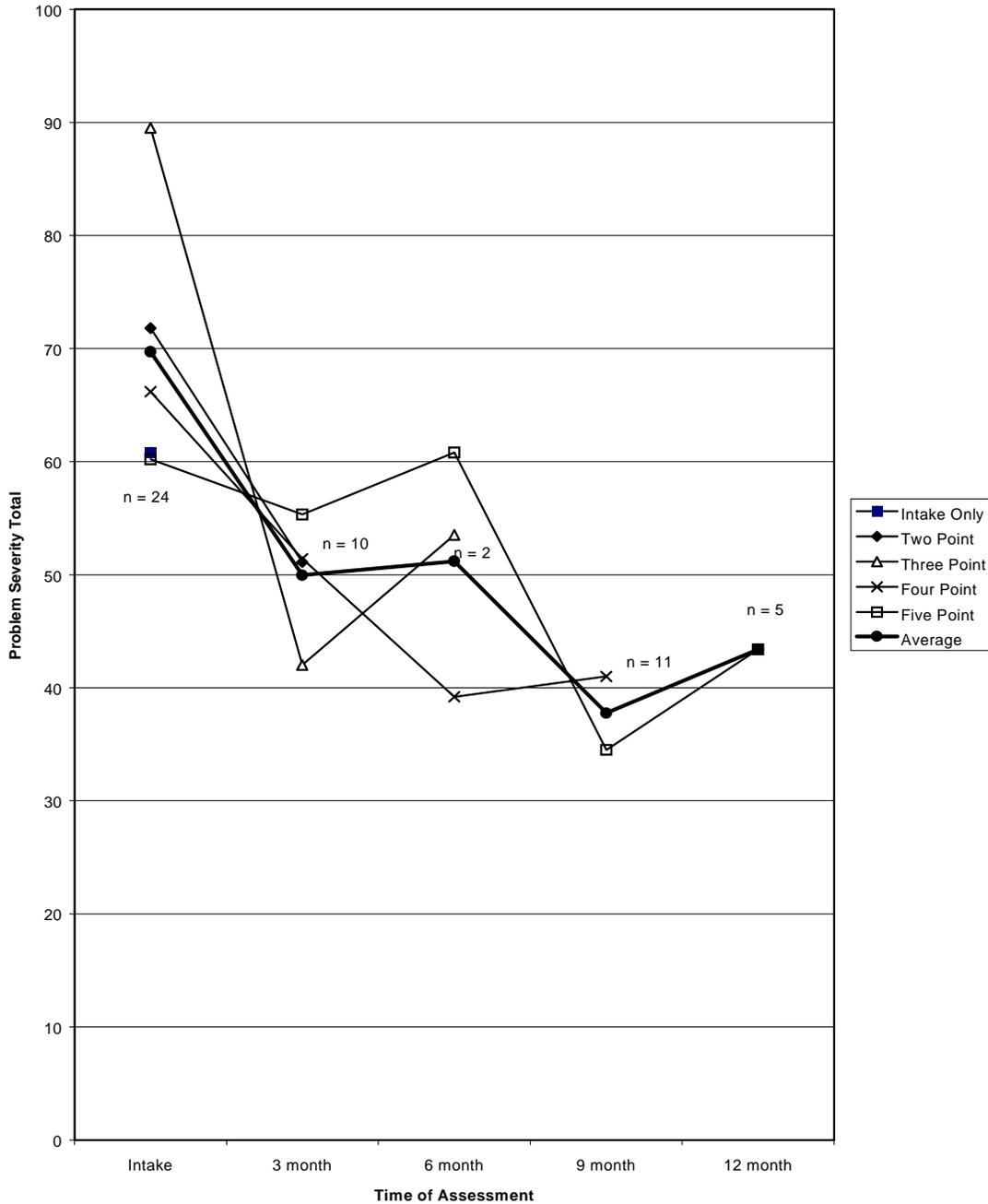
Figure 2 displays the slopes of change for each of five groups who participated in the longitudinal study as rated by the community support worker. Group 1, labeled Intake Only, includes those individuals who completed the Ohio Scales upon entry into the child community support program, but they did not continue in treatment or complete the scales thereafter. The second group, labeled two-point, completed the Ohio Scales at intake and three months later, but then dropped out of treatment or the study. The third group, labeled three point, completed the Ohio Scales at intake, three months, and six months later then dropped out. The fourth and fifth groups followed the same pattern.

Figure 2. Change in Problem Severity by Duration in Treatment Rated by the Community Support Worker



Examination of the figures suggests that the average slope of change as rated by the community support workers within each group was steeper with shorter duration. A similar pattern was exhibited by parent ratings of problem severity (Figure 3.)

Figure 3. Change in Problem Severity by Duration in Treatment Rated by the Parent



Youth rated problem severity was not graphed due to the small numbers. While changes were readily apparent on the problem severity scale, parent, case manager, and youth rated functioning remained more stable. Figures 4 and 5 depict the average change in functioning as rated by the community support worker and parents respectively.

Figure 4. Change in Functioning by Duration in Treatment Rated by the Community Support Worker

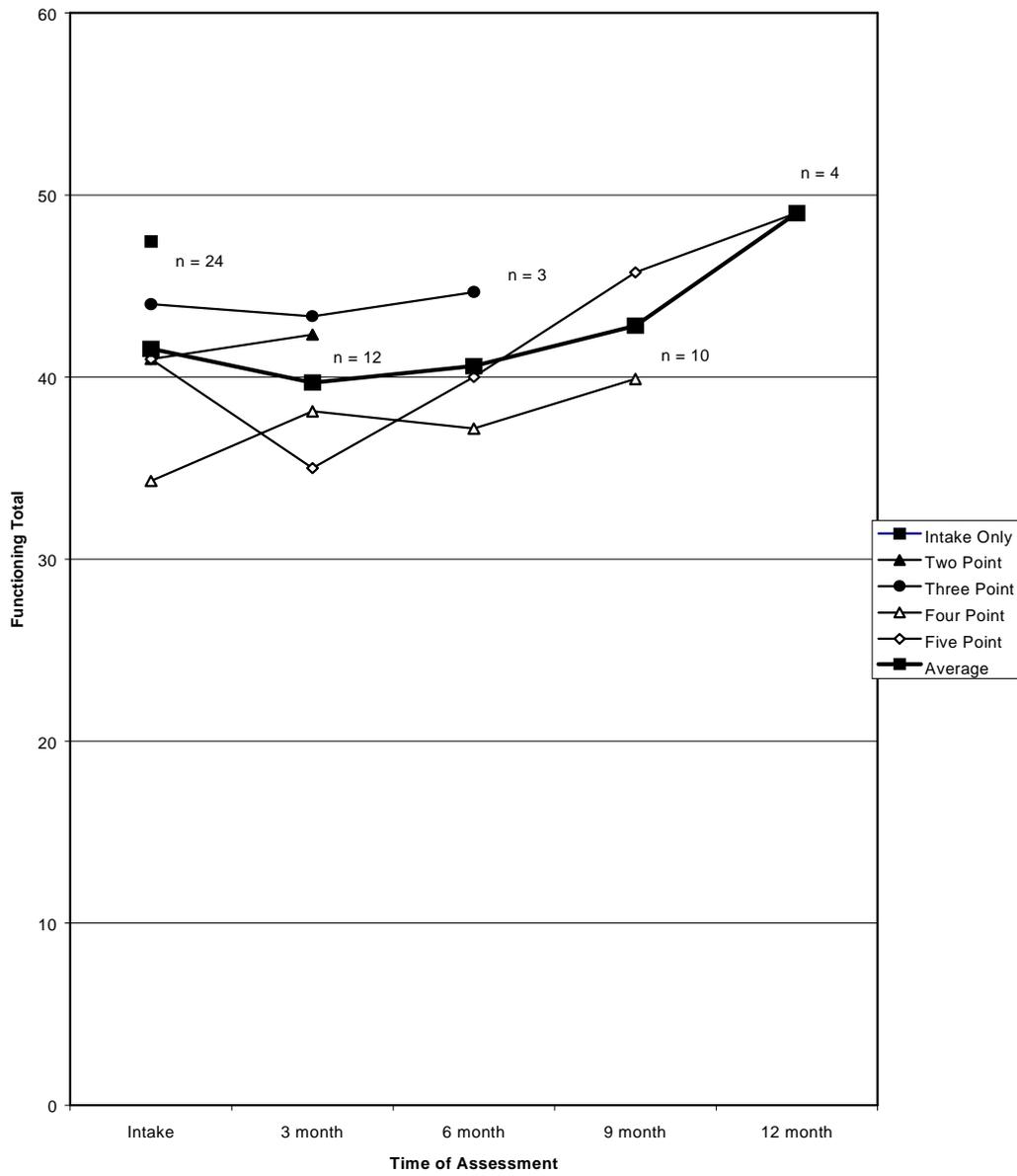
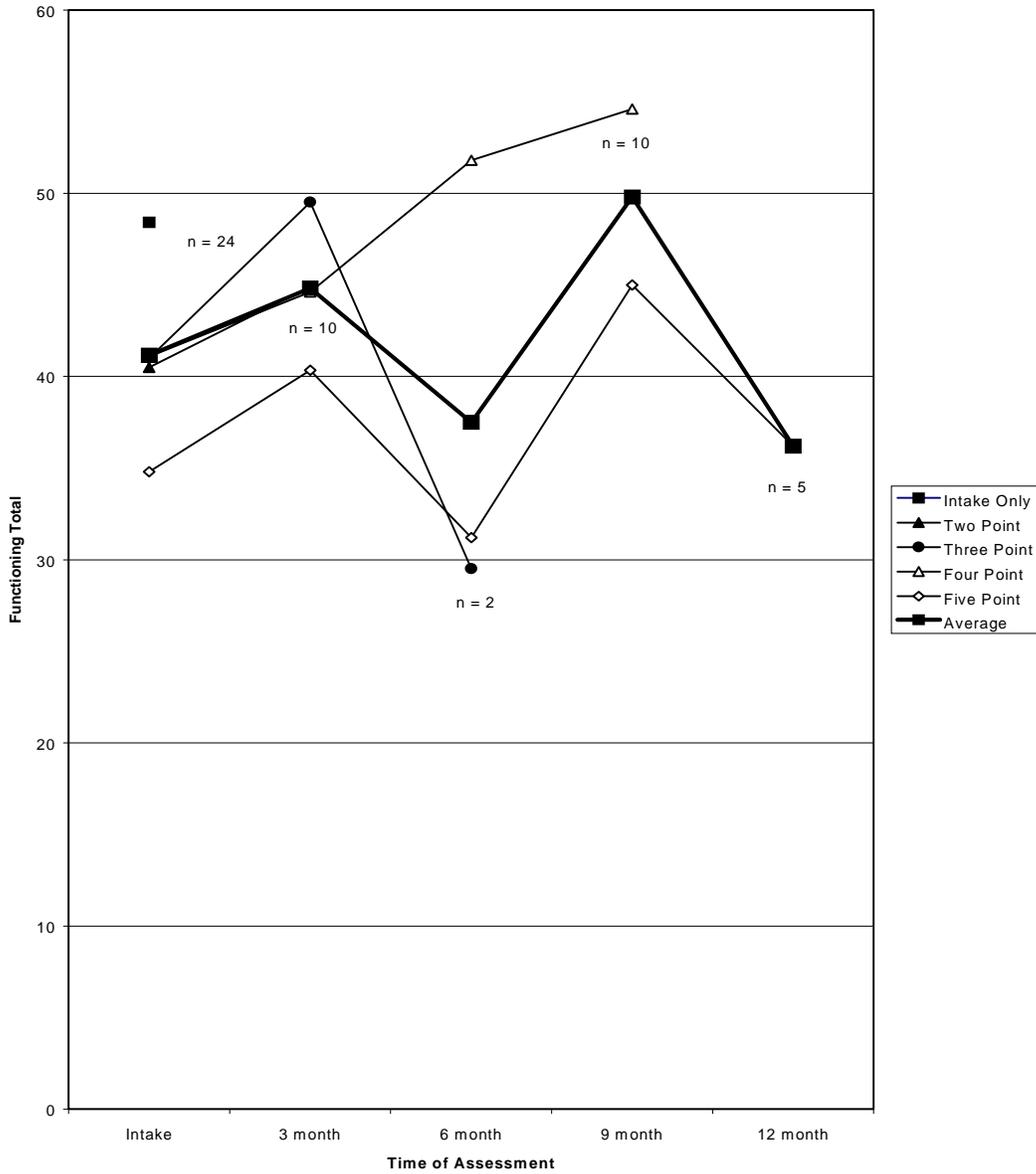


Figure 5. Change in Functioning by Duration in Treatment Rated by the Parent



As can be seen, patterns of change in functioning are more difficult to identify. From the case managers' perspective, the youth maintained the same level of functioning throughout the study. Although there appears to be a trend toward improved functioning as the duration of intervention increases, the numbers are too small to make such conclusions. Similarly, the parents' view of functioning was inconsistent. No pattern of improvement in functioning is readily apparent. If the average intake functioning were compared to the 9 month measurement point, it appears that improvement in functioning could be evident. In fact, a paired t-test on parent rated functioning using intake and 9

month scores suggests that the youth did evidence improved functioning, $t(13) = -2.098$, $p = .056$. The 13 youth who continued to receive services from intake until the 9-month assessment improved from a mean of 41.6 at intake to a mean of 51.9 at 9 months as rated by their parents. However, the pattern is not maintained at the 12 month measurement point (of course another 8 families dropped out of treatment). In addition, the small N's make attempts at sophisticated analysis and interpretation difficult, if not impossible. Our initial hypothesis was that changes in problem severity would result in subsequent changes in functioning. Clearly, more longitudinal data are necessary before such a conclusion can be reached. For now, we must suggest that changes in problem severity were readily apparent. As for functioning we suggest that one of three scenarios is in operation within this analysis:

1. The youth did not change in their level of functioning,
2. The methodology and small n were insufficient to detect changes in functioning, or
3. The Ohio Scales Functioning Scale is not sensitive to changes in functioning.

Youth Self-report Change in Outpatient Treatment. The final study investigating the sensitivity to change was obtained in sample #7. In this sample, adolescents who were receiving outpatient counseling through a large western managed behavioral health care company completed the self-report functioning scale at intake and periodically throughout treatment thereafter. A large number of youth completed the instrument at intake (nearly 1900) and their mean rating at intake is listed in Table 2. A much smaller number completed the scale at a later session ($n = 757$). Using hierarchical linear modeling (HLM), variation in intake levels of self-report functioning and average slope of change were modeled. Table 20 displays the parameter estimates and significance tests.

Table 20. Hierarchical Modeling of Change in Youth Self-Report Functioning

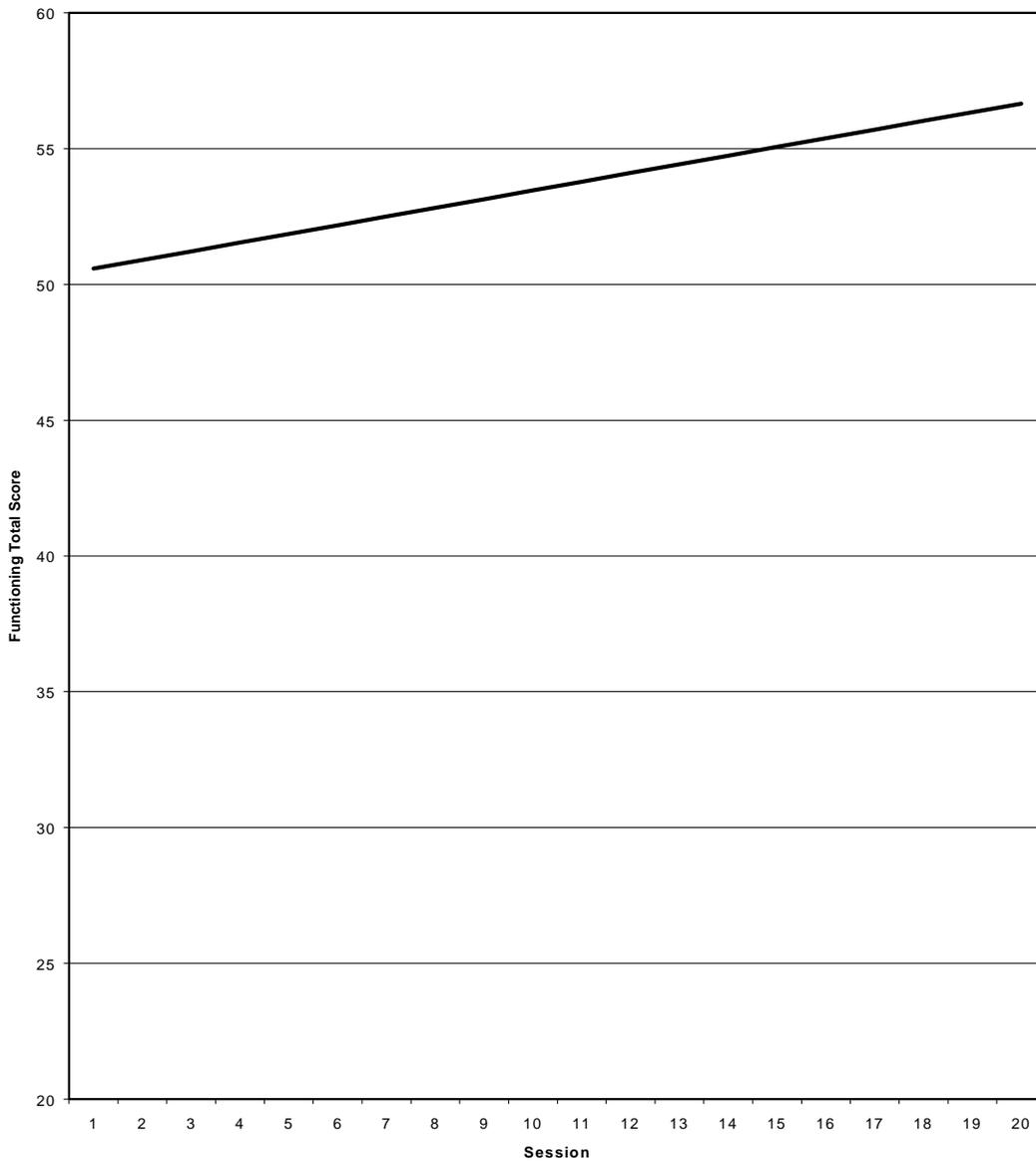
<u>Fixed Effect</u>				
<u>Parameter</u>	<u>Coefficient</u>	<u>SE</u>	<u>T-ratio</u>	<u>P-value</u>
Intercept	50.58	.333	212.18	.000
Session slope	00.45	.072	6.32	.000

<u>Random Effect</u>				
<u>Parameter</u>	<u>Variance</u>	<u>df</u>	<u>Chi-square</u>	<u>P-value</u>
Intercept	95.29	756	1198.22	.000
Session slope	.23	756	761.210	.440

As can be seen in Table 20, the fixed effects indicate that the average youth entering outpatient treatment has a self-report functioning scale score of 50.58 and an average of .45 points of change in functioning per session. The significance tests indicate

that both scores are necessary for describing the growth trajectory (Bryk & Raudenbush, 1992). The random effects indicate that the youth vary significantly on their intake scores, but rates or slopes of change do not vary significantly across individuals. Modeled change in functioning is displayed in Figure 6.

Figure 6. Modeled Change in Youth Self-report Functioning



This analysis suggests that the youth self-report functioning scale is sensitive to changes evidenced in outpatient treatment. In contrast to the previous findings in which problem severity changed but functioning did not, changes in functioning were noted during treatment based on the youth self-report. At the same time, the rate of change is

not dramatic (.45 points per session). Given the current findings, a youth would need to attend 31 sessions in order to improve one standard deviation on the functioning scale. It would appear based on the earlier data that problem severity and functioning change at different rates. Additional data is needed to ascertain the rates of change for problem severity and functioning from all three perspectives.

Summary. Overall, the data from three samples suggest that the Ohio Scales Problem Severity Scale is sensitive to changes occurring during treatment. In contrast, the data are mixed when examining the functioning scale. In the first data set, the functioning scale changes were correlated with changes on the progress evaluation scales. No changes during treatment were readily apparent on the Functioning Scale in the second data set. Finally, changes in youth self-report functioning were noted, but gradual in the third data set.

PSYCHOMETRIC PROPERTIES OF THE OHIO SCALES - SHORT FORM

Based on a factor analysis of the parent rated problem severity scale along with comparing the scores of a clinical and non-clinical sample on the parent rated problem severity items, we selected 18 items from the problem severity scale to represent the core elements of the scale. We added to this 2 items that were considered necessary for initial assessment: an item about drug and alcohol use, and an item about breaking rules or the law. In addition, we replaced the parent and agency worker version wording with the wording of the youth form. The resulting scales for the Short Form of the Ohio Scales consist of the 20 item functioning scale (reworded for parent and agency worker forms), the 4 item hopefulness scale (unchanged), the 4 item satisfaction scale (unchanged), and the 20 item problem severity scale (reworded for parent and agency worker forms). This makes a very reasonable 48 total items. Administration and scoring of the short form is identical to the long form and is described above and in the User's Manual.

Because the Short Form of the Ohio Scales is the same style as the original form and the majority of the items are identical, psychometric properties were examined to assure the correlation of the short form with the original form. Re-examination of the validity coefficients, interrater reliabilities and sensitivity to change were not conducted after determining the substantial overlap of (or correlation between) the instruments. A brief set of studies is summarized here to provide evidence that the short form is a viable alternative to the long form. As research continues, it is likely that the short forms will become the instruments of choice.

To begin evaluating the psychometric properties of the Short Form, four new samples of data were collected.

- 1) Parents of 76 students (average age 13.02, SD 3.31) rated their child using the short form and the original form of the Ohio Scales. In addition to rating the two forms of the Ohio Scales, 43 of the parents also rated their child using the Connor's parent rating scale.
- 2) Another 37 parents of youth attending appointments with a psychiatrist at a local mental health center rated their child using the problem severity scale short form, original problem severity scale and the Connor's parent rating scales. The children were 27 boys and 10 girls with an average age of 10.14 years ($SD = 3.706$).
- 3) Another clinical sample was collected consisting of 35 case manager ratings of youth receiving behavioral health services using both the agency worker short form and original form. The 35 youth (27 boys, 8 girls) were an average 12.60 years old ($SD = 3.76$). An additional 22 parent ratings of children receiving services were collected with this sample.

- 4) Finally a sample of case manager, parent, and youth ratings using the short form were collected in another part of the state in order to get a more diverse sample and to investigate the possibility of any systematic rating differences based on race. Case managers ($n = 27$) from an agency in Cleveland rated 5 youth each using the Short Form of the Ohio Scales. In addition, 38 parents and 34 youth rated their respective short forms.

Procedures

Instruments and procedures for the samples were slightly different and are described separately here. The means and standard deviations on the Ohio Scales - Short Form for each sample are displayed in Table 21.

Sample 1. Research assistants distributed packets to grade school and high school students near the end of a school day. The packet included a brief letter explaining the study (including implied consent by returning the forms) and the scales. Two separate packets were distributed.⁶ The first packet included the short and original forms of the problem severity scale and the Connor's Parent Rating Scale. The second packet included the Parent Rated Ohio Scales both original and short forms, and several demographic questions.⁷ Students were instructed to ask their parents to complete the forms in the evening and return them in an envelope to the research assistants prior to school the next morning. All students who returned the forms received a gift certificate for their participation. Research assistants collected forms two consecutive mornings after the packets were distributed. Students could also return the forms to the school secretary thereafter. A total of 76 parents returned completed forms (some individual items were inadvertently left blank for some participants).

Sample 2. This clinical sample was collected at a community mental health center in southeastern Ohio. Parents (or primary caregivers) who were attending a consultation with the psychiatrist along with their child were asked to rate their child using the short form of the Problem Severity Scale, the long form of the Problem Severity Scale, and the Connor's Parent Rating Scale. A total of 37 parents completed the ratings. The 27 boys and 10 girls who were rated were on average 10.14 years old ($SD = 3.706$).

Sample 3. This clinical sample was collected at a second community mental health center in southeastern Ohio. Case managers rated the cases using the short and original forms (all scales) of the Ohio Scales. A total of 35 youth currently receiving services were rated. In addition, 22 parents participated by rating their child using the long and short forms of the Ohio Scales.

⁶ Two packets were distributed because half of the data were collected within the design of a student Thesis project.

⁷ The satisfaction scale was not included since most of the children were not participating in mental health services.

Sample 4. This clinical sample included youth and their parent or primary caregiver who were receiving community support services at an agency in Cleveland. Case managers (N = 27) rated 5 children each using the short form of the Ohio Scales. In addition, 38 parents (or primary caregivers) and 34 youth rated the respective short forms of the Ohio Scales.

Means and standard deviations for the 4 samples are displayed in Table 21.

Table 21. Means and Standard Deviations on the Short Form of the Ohio Scales for the different samples.

<u>Population: Sample Number</u>	<u>N</u>	<u>Problems</u> <u>M (SD)</u>	<u>Functioning</u> <u>M (SD)</u>
Community: Sample # 1			
• Parents (Packet A)	43	13.28 (10.01)	NA
• Parents (Packet B)	33	13.12 (12.24)	67.79 (10.20)
Clinical: Sample # 2			
• Parents	37	35.43 (19.72)	NA
Clinical: Sample # 3			
• Agency workers	35	19.48 (18.06)	63.38 (14.63)
• Parents	22	28.91 (14.71)	44.81 (13.93)
Clinical: Sample # 4			
• Youth	34	29.56 (13.78)	60.03 (11.30)
• Parents	38	40.47 (18.08)	39.60 (17.15)
• Agency workers	135	41.04 (14.40)	33.94 (12.91)

Reliability

No extensive evaluation of the reliability was conducted for the Ohio Scales - short form. Internal consistency estimates for the problem severity and functioning scales are presented in Table 22. No other forms of reliability were examined.

Table 22. Internal Consistency Estimates (Cronbach's Alpha) for each Scale on the Short Form for Community and Clinical Samples.

<u>Scale</u>	<u>Community</u>		<u>Clinical</u>	
	Parent (1a) (<u>n</u> = 43)	Parent (1b) (<u>n</u> = 33)	Parent (2) (<u>n</u> = 37)	Agency worker (4) (<u>n</u> = 124)
Problem Severity	.89	.90	.93	.86
Functioning	NA	.93	NA	.91

Validity

The primary evidence for validity of the reworded functioning scale and the reworded and shortened problem severity scale is a high correlation with the original Ohio Scales. Data were collected for parent and agency worker versions to demonstrate consistency of the measurement between the short and original forms and are presented by source of ratings.

Agency Worker. The agency worker original form and short form Ohio Scale ratings in sample 3 were highly correlated (see Table 23).

Table 23. Correlations Between the Agency Worker Rated Short Form and Original Ohio Scales.

<u>Original</u>	<u>Short Form</u>	
	<u>Problem Severity</u>	<u>Functioning</u>
Problem Severity	.80*	-
Functioning	-	.91*

* $p < .01$ (2-tailed); $n = 35$

Parent Ratings. In samples 1, 2, and 3, parent ratings of the youth's problem severity and functioning on the short and original forms of the Ohio Scales were correlated. As can be seen in Table 24, the correlations are significant for all samples.

Table 24. Correlation Coefficients for the Original and Short Forms of the Parent Rated Ohio Scales

<u>Original Ohio Scales</u>	<u>Short Form</u>	
	<u>Problem Severity</u>	<u>Functioning</u>
Sample 1A	<u>Problem Severity</u>	
Problem Severity	.95*	
Connor's	.84*	
Sample 1B	<u>Problem Severity</u>	<u>Functioning</u>
Problem Severity	.89*	-
Functioning	-	.96*
Sample 2	<u>Problem Severity</u>	
Problem Severity	.97	
Sample 3	<u>Problem Severity</u>	<u>Functioning</u>
Problem Severity	.91*	-
Functioning	-	.86*

* $p < .001$

Because the original validation of the Ohio Scales used samples from Southeast Ohio, no data for diverse groups were collected. As a result, a data were collected from an urban site (Cleveland) to investigate the possibility of any systematic differences in scores based on race. In this sample, 27 case managers rated 5 clients each. Total scores for problem severity and functioning for minority and majority youth were compared to see if differences existed. As can be seen in the Table 25, no significant differences existed between the case manager ratings of majority (n = 62) and minority (n = 73) youth.

Table 25. Comparison of Case Manager Ratings of Minority and Majority Youth

<u>Scale</u>	<u>Group</u>	<u>Mean</u>	<u>Std. Deviation</u>
Problem Severity	Majority	40.88*	12.52
	Minority	41.16	15.91
Functioning	Majority	33.80	12.84
	Minority	34.05	13.05

* $n = 135$; no significant differences between means evident.

Similarly, data collection from youth and parents from the urban location revealed no differences between majority and minority group ratings by parents or youth report. (See Tables 26 and 27). There were also no differences in hopefulness or satisfaction with services on parent or youth ratings.

Table 26. Comparison of Parent Ratings of Minority and Majority Youth

<u>Scale</u>	<u>Group</u>	<u>Mean</u>	<u>Std. Deviation</u>
Problem Severity	Majority	38.42*	19.97
	Minority	41.67	17.21
Functioning	Majority	42.07	14.08
	Minority	38.17	18.85

* $n = 38$; no significant differences between means evident.

Table 27. Comparison of Minority and Majority Youth Self-Report Ratings

<u>Scale</u>	<u>Group</u>	<u>Mean</u>	<u>Std. Deviation</u>
Problem Severity	Majority	27.44*	12.33
	Minority	31.45	15.05
Functioning	Majority	61.50	9.35
	Minority	58.72	12.91

* $n = 34$; no significant differences between means evident.

Summary

After using factor analysis and discrimination between clinical and non-clinical samples to shorten the problem severity scale, we replaced the wording of the parent and agency worker rated problem severity and functioning scales with the wording used on the youth self-report form. We then examined the revised scales to ascertain the overlap between the short and original versions of the scales. Correlation coefficients between the short and original scales for both problem severity and functioning are highly correlated. This suggests that the short form can be reasonably applied as an alternative to the original scales with some practical benefits while maintaining the integrity of the original conceptualization.

In addition, a more diverse sample from a metropolitan area was collected to investigate the possibility of any differences or sensitivities of ratings on the Ohio Scales to race. When comparing majority and minority ratings for parents, youth, and agency workers, no differences were evident on any of the four content areas (problem severity, functioning, satisfaction, and hopefulness).

CONCLUSION

After reviewing the current state of outcome measurement within children's behavioral health services, we developed three brief measures of outcome covering multiple content areas from multiple sources. Our intent was to develop measures that could be used to track the progress of youth with serious emotional disorders as they receive behavioral health services. We hoped to develop pragmatic yet empirically sound measures that are grounded in the theoretical and practical world of multi-need youth.

The inclusion of multiple content areas (problem severity, functioning, satisfaction, and hopefulness) rated by multiple sources using identical items (problem severity and functioning only) in a brief practical form offers a substantial advantage for the test administrator. In the typical circumstance, the test user would need to gather multiple tests from various test developers or distributors in order to cover several content areas. These tests are likely to have varying lengths, formats, costs, scoring procedures, practices for interpretation etc. For the Ohio Scales, however, the format, length, scoring, interpretation, and cost are all similar. This simple, practical format allows the user to collect meaningful global data regarding relevant content areas from the principal sources of information.

One convenient aspect of the Ohio Scales is its compartmentalization. Some users are taking sections/scales of the Ohio Scales rather than using the entire package. For example, one agency uses only the youth self-report of functioning combined with other measures for parent and therapist data collection. We attempted to develop brief measures that are also easy to administer, score, and interpret.

Notably, the Ohio Scales are not diagnostic instruments. The instruments do in fact provide useful pretreatment information (see the User's Manual for examples). However, the instruments were not developed to broadly assess or screen for the range of potential diagnostic issues and symptoms that might be relevant for more in depth evaluations. Other measures are available for collecting more in depth diagnostic information at intake (e.g., CBCL). The Ohio Scales were developed to be repeatedly administered over time as a way of evaluating and tracking the effectiveness of services using items that are endorsed by a large number of parents and youth who present for services. As a result, some tradeoffs were made to maintain the practical nature of the scales resulting in the sacrifice of the potential diagnostic utility of the instrument.

The results of the initial studies investigating the psychometric properties of the original Ohio Scales are quite positive. The Ohio Scales have adequate internal consistency and test-retest reliability. The inter-rater reliability of the agency worker functioning scale is adequate when using a standardized format for data collection. Preliminary evidence of concurrent and construct validity suggests the measures are assessing satisfaction, severity of problems, and youth levels of functioning. Finally, the instruments appear to be sensitive to change. Clearly additional data are needed to continue the validation of the Ohio Scales. For now, however, we feel confident that the

data collected to date suggest the instruments are sufficiently tested for use in applied settings.

Based on qualitative feedback from users of the test, we further enhanced the Ohio Scales by developing a Short Form of all three scales. The shorter (48 items) version includes 20 problem severity items, 4 satisfaction items, 4 hopefulness items, and 20 functioning items. In addition to making the problem severity scale shorter, the wording of the case worker and parent versions of the short forms were changed to match the youth form. This makes the wording identical for all three forms and reduces the reading level for the parent and case worker versions. Initial data were also collected to verify that the short forms are substantively equivalent to the long forms. Overall, the psychometric properties appear to remain satisfactory despite the brevity.

Ultimately, it is our hope that by conforming to the rather stringent conceptual and psychometric requirements, the final result is pragmatically useful yet methodologically rigorous outcome measures. The final usefulness of the Ohio Scales and this manual, however, will be determined by those who use the scales. We welcome your comments and hope that the delicate balance between research rigor and pragmatics does not diminish the quality of the work. Please send comments to ogles@ohio.edu or Ben Ogles, Ph. D., Porter Hall 241, Ohio University, Athens, OH 45701.

References

- Achenbach, T. M., & Edelbrock, C. (1983). Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington, VT: University of Vermont Department of Psychiatry.
- Attkisson, C. C. & Zwick, R. (1982). The client satisfaction questionnaire: Psychometric properties and correlations with service utilizations and psychotherapy outcome. Evaluation and Program Planning, 5, 233-237.
- Barth, R. P. (1986). Social and Cognitive Treatment of Children and Adolescents. San Francisco, CA: Jossey-Bass.
- Bickman, L., Guthrie, P. R., Foster, E. M., Lambert, W., Summerfelt, W. T., Breda, C. S., & Heflinger, C. A., (1995). Evaluating managed mental health services: The Fort Bragg experiment. New York: Plenum.
- Bickman, L., Lambert, E. W., Summerfelt, W. T., Karver, M. (1996). The Vanderbilt Functioning Index: Preliminary parent version. Unpublished manuscript.
- Bryk, A. S. & Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CS: Sage.
- Burchard, J. D. & Clarke, R. T. (1990). The role of individualized care in a service delivery system for children and adolescents with severely maladjusted behavior. The Journal of Mental Health Administration, 17, 48-60.
- Burchard, J. D. & Schaefer, M. (1992). Improving accountability in a service delivery system in children's mental health. Clinical Psychology Review, 12, 867-882.
- Burns, B. J. & Friedman R. M. (1988). The research base for child mental health services and policy: How solid is the foundation. Conference Proceedings: Children's Mental Health Services and Policy: Building a Research Base. Tampa, FL: Research and Training Center for Children's Mental Health.
- Cochran, M. (1987). Empowering families: An alternative to the deficit model. In Hurrelmann, K., Hurrelmann, F., & Lostel, F. (Eds.), Social Intervention: Potential and Constraints (pp. 105-119). Berlin: Walter de Gruyter.
- Connors, ** add reference
- Dohrenwend B. P. & Dohrenwend, B. S. (1981). Perspectives on the past and future of psychiatric epidemiology: The 1981 Rena Lapuse Lecture. American Journal of Public Health, 72(1), 1271-1279.
- Duchnowski, A. J. & Friedman, R. M. (1990). Children's mental health: Challenges for the nineties. Journal of Mental Health Administration, 17, 3-12.
- Duchnowski, A. J., Johnson, M. K., Hall, K. S., Kutash, K., & Friedman, R. M. (1993). The alternatives to residential treatment study: Initial findings. Journal of Emotional and Behavioral Disorders, 1(1), 17-26.
- Dunst, C. J., Trivette, C. M., & Deal, A. G. (1988). Enabling and Empowering Families: Principles and Guidelines for Practice. Cambridge, MA: Brookline Books.
- Evans, M. E., Dollard, N., Huz, S., & Rahn, D. S. (1990). Outcomes of Children and Youth Intensive Case Management in New York State. Paper presented at the American Public Health Association Meetings, Atlanta, GA.

Friesen, B. J. & Koroloff, N. M. (1990). Family-centered services: Implications for mental health administration and research. The Journal of Mental Health Administration, 17, 13-25.

Friesen, B. J., Koren, P. E., & Koroloff, N. M. (1992). How parents view professional behaviors: A cross-professional analysis. Journal of Child and Family Studies, 1, 209-231.

Gillespie, D. K. (1993). Enhancing the methodology of social validation: The application of psychometric measures to the Pennsylvania project social validation instrument. Unpublished masters's thesis, Ohio University, Athens.

Gold, N. (1983). Stakeholders and program evaluation: Characteristics and reflections. In Bryk (Ed.), Stakeholder-Based Evaluation (pp. 63-72). San Francisco, CA: Jossey-Bass.

Grundy, C. T., Lunnen, K. M., Lambert, M. J., Ashton, J. E., & Tovey, D. (1994). Hamilton Rating Scale for Depression: One scale or many? Clinical Psychology - Science & Practice, 1, 197-205

Grundy, C. T., Lambert, M. J., & Grundy, E. M. (1996). Assessing clinical significance: Application to the Hamilton Rating Scale for Depression. Journal of Mental Health, 5, 25-33.

Hawkins, R. P., Almeida, M. C., Fabry, B., Reitz, A. (1992). A scale to measure restrictiveness of living environments for troubled children and youths. Hospital and Community Psychiatry, 43, 54-58.

Hodges, K. & Wong, M. M. (1996). Psychometric characteristics of a multidimensional measure to assess impairment: The Child and Adolescent Functional Assessment Scale. Journal of Child and Family Studies, 5, 445-467.

Ihlevich, D. & Glessner, G. C. (1979). A Manual for the Progress Evaluation Scales. Shiawasse, MI: Community Mental Health Services Board.

Kotlowitz, A. (1991). There Are No Children Here: The Story of Two Boys Growing Up in the Other America. New York, NY: Doubleday.

Kutash, K., Duchnowski, A., Johnson, M. & Rugs, D. (1993). Multi-stage evaluation for a community mental health system for children. Administration and Policy in Mental Health, 20, 311-322.

Lambert, M. J. & Bergin, A. E. (1994). The effectiveness of psychotherapy. In Bergin, A. E. & Garfield, S. L. (Eds.), Handbook of Psychotherapy and Behavior Change (pp. 143-189) (4th ed.). New York, NY: John Wiley.

Lambert, M. J. & Hill, C. E. (1994). Assessing psychotherapy outcomes and processes. In Bergin, A. E. & Garfield, S. L. (Eds.), Handbook of Psychotherapy and Behavior Change (pp. 72-113) (4th ed.). New York, NY: John Wiley.

Lambert, M. J., Christensen, E. R., & DeJulio, S. S. (1983). The Assessment of Psychotherapy Outcome. New York, NY: John Wiley.

Lambert, M. J., Ogles, B. M., & Masters, K. S. (1992). Choosing outcome assessment devices: An organizational and conceptual scheme. Journal of Counseling and Development, 70, 538-539.

Ogles, B. M., Davis, D C., & Lunnen, K. M. (1998, March). The interrater reliability of four measures of functioning. Paper presented at the Research and Training Center for Children's Mental Health's 11th Annual Research Conference, Tampa.

Ogles, B. M., Lambert, M. J., & Masters, K. S. (1996). Assessing outcome in clinical practice. Boston: Allyn and Bacon.

Ogles, B. M., Lunnen, K. M., Gillespie, D. K., Trout, S. C. (1996). Conceptualization and initial development of the Ohio Scales. In C. Liberton, K. Kutash, & R. Friedman, (Eds.), The 8th Annual Research Conference Proceedings, A System of Care for Children's Mental Health: Expanding the Research Base. (pp. 33-37). Tampa FL: University of South Florida, Florida Mental Health Institute, Research and Training Center for Children's Mental Health.

Poertner, J. & Ronnau, J. (1992). A strengths approach to children with emotional disabilities. In Saleebey, D. (Ed.), The Strengths Perspective in Social Work Practice (pp. 111-121). New York, NY: Longman.

Rosenberg, M. (1979). Conceiving the Self. New York, NY: Basic Books.

Schriner, K. F. & Fawcett, S. B. (1988). Development and validation of a community concerns report method. Journal of Community Psychology, 16, 306-316.

Sederer, L. I. & Dickey, B. (Eds.). (1996). Outcomes assessment in clinical practice. Baltimore, MD: Williams & Wilkins.

Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahlia, S. (1983). A Children's Global Assessment Scale (CGAS). Archives of General Psychiatry, 40, 1228-1231.

Stroul, B. A. & Friedman, R. M. (1986). A System of Care for Severely Emotionally Disturbed Children and Youth (Revised edition). Washington, D. C.: Georgetown University Child Development Center.

Strupp, H. H. & Hadley, S. W. (1977). A tripartite model of mental health and therapeutic outcome: With special reference to negative effects in psychotherapy. American Psychologist, 32, 187-196.

VanDenBerg, J., Beck, S., & Pierce, J. (1992). The Pennsylvania Outcome Project for Children's Services. Paper presented at the 5th annual research meeting of the Research and Training Center for Children's Mental Health, Tampa, FL.

Weber, D. O. (1998). A field in its infancy: Measuring outcomes for children and adolescents. In K. J. Midgail (Ed.). The behavioral outcomes & guidelines sourcebook. Washington, DC: Faulkner and Gray's Healthcare Information Center.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. Journal of Applied Behavior Analysis, 11, 203-214.